Reweighted Eigenvalues: A New Approach to Spectral Theory beyond Undirected Graphs

by

Kam Chuen (Alex) Tung

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Santosh Vempala, Professor
School of Computer Science
Georgia Institute of Technology

Supervisor: Lap Chi Lau, Professor
Cheriton School of Computer Science
University of Waterloo

Internal Members: Eric Blais, Associate Professor
Cheriton School of Computer Science
University of Waterloo

Sepehr Assadi, Associate Professor
Cheriton School of Computer Science
University of Waterloo

Internal-External Member: Levent Tunçel, Professor
Department of Combinatorics and Optimization
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

The main results of this thesis are based on papers that I have coauthored, as well as unpublished work of mine. The details are as follows:

- Chapter 4 is based on the paper [KLT22]: Tsz Chiu Kwok, Lap Chi Lau, and Kam Chuen Tung. Cheeger inequalities for vertex expansion and reweighted eigenvalues. In *Proceedings of the 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 366-377. IEEE, 2022.

- Chapter 5 is based on [LTW23]: Lap Chi Lau, Kam Chuen Tung, and Robert Wang. Cheeger inequalities for directed graphs and hypergraphs using reweighted eigenvalues. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1834–1847, 2023.

- The first part of Chapter 8 is based on [LTW24]: Lap Chi Lau, Kam Chuen Tung, and Robert Wang. Fast algorithms for directed graph partitioning using flows and reweighted eigenvalues. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 591–624. SIAM, 2024.

- The second part of Chapter 8, as well as Chapters 6 and 7, are based on my unpublished work. Much of the work in Chapter 6 was done with Yuichi Yoshida while I was a visiting student at the National Institute of Informatics (NII), Tokyo, Japan.

Abstract

We develop a concept called reweighted eigenvalues, to extend spectral graph theory beyond undirected graphs. Our main motivation is to derive Cheeger inequalities and spectral rounding algorithms for a general class of graph expansion problems, including vertex expansion and edge conductance in directed graphs and hypergraphs. The goal is to have a unified approach to achieve the best known results in all these settings.

The first main result is an optimal Cheeger inequality for undirected vertex expansion. Our result connects (i) reweighted eigenvalues, (ii) vertex expansion, and (iii) fastest mixing time [BDX04] of graphs, similar to the way the classical theory connects (i) Laplacian eigenvalues, (ii) edge conductance, and (iii) mixing time of graphs. We also obtain close analogues of several interesting generalizations of Cheeger's inequality [Tre09, LOT12, LRTV12, KLL+13] using higher reweighted eigenvalues, many of which were previously unknown.

The second main result is Cheeger inequalities for directed graphs. The idea of *Eulerian reweighting* is used to effectively reduce these directed expansion problems to the basic setting of edge conductance in undirected graphs. Our result connects (i) Eulerian reweighted eigenvalues, (ii) directed vertex expansion, and (iii) fastest mixing time of directed graphs. This provides the first combinatorial characterization of fastest mixing time of general (non-reversible) Markov chains. Another application is to use Eulerian reweighted eigenvalues to certify that a directed graph is an expander graph.

Several additional results are developed to support this theory. One class of results is to show that adding ℓ_2^2 triangle inequalities [ARV09] to reweighted eigenvalues provides simpler semidefinite programming relaxations, that achieve or improve upon the previous best approximations for a general class of expansion problems. These include edge expansion and vertex expansion in directed graphs and hypergraphs, as well as multi-way variations of some undirected expansion problems. Another class of results is to prove upper bounds on reweighted eigenvalues for special classes of graphs, including planar, bounded genus, and minor free graphs. These provide the best known spectral partitioning algorithm for finding balanced separators, improving upon previous algorithms and analyses [ST96, BLR10, KLPT11] using ordinary Laplacian eigenvalues.

Acknowledgements

This thesis would not have been possible if not for the many people in my life, and this is a good place for me to express my gratitude to them. Any omission is solely my fault and I ask for your forgiveness.

First and foremost, I was incredibly fortunate to have Lap Chi Lau as my advisor. He is everything I could have hoped for from a PhD advisor. He has taught me so much both in domain knowledge and in how to research. He has an eye for important research problems and writes and presents well – traits that hopefully I have inherited if only slightly. He is demanding, yet understanding and supportive at all times. Lap Chi, this thesis would not have been possible without you.

Second, I would like to thank the professors that have taught or advised me through the years: Anastasios Sidiropoulos, Anna Lubiw, Ashwin Nayak, Eric Blais, Gautam Kamath, Levent Tunçel, Rafael Oliviera, Richard Cleve, Richard Peng, Santosh Vempala, Sepehr Assadi, Shai Ben-David, Siu On Chan, Tasuku Soma, Thatchaphol Saranurak, Yuichi Yoshida. Special thanks to Yuichi Yoshida who hosted my visit to NII Tokyo in summer 2023, to Siu On Chan who introduced me to theoretical computer science, and to everyone on the Thesis Committee for taking time to read the (long) thesis and to give valuable feedback.

Third, I am grateful to have met many wonderful peers both in Waterloo and in conferences and workshops. Your friendships have led to many good times, and the sense of solidarity has helped me through the difficult times. Special mention to everyone who is/was in Lap Chi's group: Tsz Chiu Kwok, Alan Chan, Vedat Alev, Hong Zhou, Akshay Ramachandran, Robert Wang, Yueheng Zhang, Dante Tjowasi.

Fourth, I would like to express my gratitude to my friends from Hong Kong. The meetups are sparse and often online, but I treasure every single one of them. Thank you for your continued friendship and support – and you know who you are.

Last but not least, I would like to thank my parents and grandparents, my brother, and my girlfriend. Thank you for your unconditional love and support.

Dedication

To my friends and family.

Table of Contents

Chapter 1

Introduction

Spectral graph theory is the study of graphs via the spectrum of associated matrices or operators. A fundamental result in spectral graph theory, Cheeger's inequality [Che70, AM85, Alo86, SJ89] connects the edge expansion property of an undirected graph $G = (V, E)$ to the second eigenvalue of its associated Laplacian matrix:

$$\frac{\lambda_2(G)}{2} \leq \phi(G) \leq \sqrt{2\lambda_2(G)} \tag{1.1}$$

where $\phi(G)$ is the edge conductance of G, which is a natural graph isoperimetric constant that informally measures how well the graph is connected, and $\lambda_2(G)$ is the second smallest eigenvalue of its normalized Laplacian matrix[1]. Crucially, the edge conductance is NP-hard to compute [ŠS06], while the second eigenvector can be computed in near-linear time. Plus, there is a fast and simple rounding algorithm called the sweep-cut algorithm, that extracts a set S with small edge conductance from the second eigenvector. Thus, spectral methods open up an avenue for fast graph algorithms.

There are two important applications of Cheeger's inequality. One is to use the second eigenvalue to study expander graphs (see [HLW06] for survey) and its eigenvector for graph partitioning (see [Lux07] for survey), notably image segmentation [SM00]. The other application is to use the edge conductance to bound the mixing time of random walks [AF02, LP17]. Together, Cheeger's inequality connects (i) edge conductance, (ii) the second eigenvalue, and (iii) mixing time. More recently, the spectral theory for undirected graphs is enriched by several interesting generalizations of Cheeger's inequality [Tre09, ABS10, LOT12, LRTV12, KLL+13], which establish further connections between edge

[1]See Chapter 2 for various definitions that are not stated in this introduction.

expansion properties of the graph to higher eigenvalues $\lambda_k(G)$ of its normalized Laplacian matrix.

The richness of the "basic" spectral theory for edge conductance in undirected graphs inspires many efforts to derive a spectral theory for more general settings such as directed graphs and hypergraphs. Despite these efforts, the current spectral theories for these more general settings fail to fully capture the rich applications of the basic spectral theory.

In this thesis, we develop the framework of "reweighted eigenvalues", which is a unifying framework that reduces the study of expansion quantities in more general settings to the basic setting of edge conductance in undirected graphs. Here is a resumé of our results:

1. The first main result is to develop a spectral theory to study vertex expansion in graphs, that connects (i) reweighted eigenvalues, (ii) vertex expansion, and (iii) fastest mixing time, and captures direct analogues of several generalizations of Cheeger's inequality [Tre09, LOT12, LRTV12, KLL+13] for vertex expansion (Chapter 4).

2. The second main result is to develop a spectral theory for directed graphs, with applications in characterizing fastest mixing time for general Markov chains and certifying directed expander graphs. Our spectral theory can be readily extended to hypergraphs as well (Chapter 5).

3. We then attempt to further generalize our framework to submodular transformations [Yos19], and obtain Cheeger inequalities for directed hypergraphs, which is a common generalization of vertex expansion and expansion problems on directed graphs and hypergraphs (Chapter 6).

4. We derive upper bounds on the reweighted eigenvalues for planar graphs and beyond, with applications to graph partitioning (Chapter 7).

5. We obtain tighter approximations to generalized expansion quantities by adding "ℓ_2^2 triangle inequalities" constraints as in [ARV09] to the reweighted eigenvalue formulations (Chapter 8).

1.1 A Spectral Theory for Vertex Expansion

There are different ways to measure the connectedness of a graph. Vertex expansion is one major alternative to edge conductance, which is based on vertex cuts rather than edge cuts.

In applications such as error-correcting codes, divide-and-conquer algorithms on graphs, and network design, vertex expansion is often the expansion quantity of interest.

Despite the importance of vertex expansion, there is no known satisfactory spectral theory for it. The earliest spectral theory using Laplacian eigenvalues [Tan84, AM85, Alo86] only works well for bounded-degree graphs. Other spectral formulations [BHT00, LRV13, Lou15, CLTZ18] do not yield an optimal Cheeger-type inequality that captures some other appealing aspects of the classical theory; for example, the connection to mixing time and the various generalizations. The first research question that we address in this thesis is to derive a "good" spectral theory for vertex expansion in undirected graphs.

The concept of reweighted eigenvalues, which is the main theme of this thesis, comes from a line of work [BDX04, Roc05, OZ22] that relates vertex expansion with the so-called fastest mixing time. The fastest mixing time problem is introduced by Boyd, Diaconis and Xiao [BDX04]. In the problem, we are given an undirected graph $G = (V, E)$ and a target probability distribution $\pi : V \to \mathbb{R}^+$. The task is to find a time-reversible transition matrix $P \in \mathbb{R}^{|V|\times|V|}$ supported on the edges of the graph G, so that the stationary distribution of random walks with transition matrix P is π. The objective is to find such a transition matrix that minimizes the mixing time to the stationary distribution π.[2] It is well-known that the mixing time to the stationary distribution is approximately inversely proportional to the spectral gap of the transition matrix P. The fastest mixing time problem is thus formulated as follows in [BDX04] by the maximum spectral gap achievable through such a "reweighting" P of the graph G.

Definition 1.1.1 (Maximum Reweighted Second Eigenvalue [BDX04])**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted second eigenvalue is defined as*

$$\begin{aligned} \lambda_2^*(G) \ := \ \max_{P \geq 0} \quad & 1 - \alpha_2(P) & \\ \text{subject to} \quad & P(u,v) = P(v,u) = 0 & \forall uv \notin E \\ & \sum_{v \in V} P(u,v) = 1 & \forall u \in V \\ & \pi(u)P(u,v) = \pi(v)P(v,u) & \forall uv \in E, \end{aligned}$$

where $\alpha_2(P)$ is the second largest eigenvalue of P.
(Discussion of the finer aspects of the definition are deferred to Section 3.2.2.)

[2]See Section 2.6 for relevant definitions.

Boyd, Diaconis and Xiao showed that this optimization problem can be written as a semidefinite program (SDP) and thus $\lambda_2^*(G)$ can be computed in polynomial time. Subsequently, the fastest mixing time problem has been studied in various works (see [Roc05, BDSX06, BDPX09, FK13, CA15] and more references in [OZ22]).

The surprising connection between the fastest mixing Markov chain and vertex expansion is discovered through the works of Roch [Roc05] and Olesker-Taylor and Zanetti [OZ22]. Roch showed that the fastest mixing time of a graph is slow if the graph has small vertex expansion. Olesker-Taylor and Zanetti then discovered an elegant Cheeger-type inequality for vertex expansion and the maximum reweighted second eigenvalue, showing that the fastest mixing time of a graph is slow *if and only if* the graph has small vertex expansion. Our first main result is to improve their result to an "optimal" Cheeger-type inequality.

Theorem 1.1.2 (Optimal Cheeger Inequality for Vertex Expansion [Roc05, OZ22, KLT22])**.** *For any undirected graph $G = (V, E)$ with maximum degree Δ and any probability distribution π on V,*

$$\frac{\psi(G)^2}{\log \Delta} \lesssim \lambda_2^*(G) \lesssim \psi(G),$$

where $\psi(G)$ is the vertex expansion[3] *of G. Furthermore, there are tight examples showing that the $\log \Delta$ factor is asymptotically optimal. In terms of the fastest mixing time $\tau^*_{\text{mix}}(G)$ to the distribution π, writing $\pi_{\min} := \min_{v \in V} \pi(v)$ we have*

$$\frac{1}{\psi(G)} \lesssim \tau^*_{\text{mix}}(G) \lesssim \frac{\log \Delta \cdot \log \pi_{\min}^{-1}}{\psi^2(G)}.$$

This is the starting point of our spectral theory for vertex expansion, one that relates (i) vertex expansion, (ii) reweighted eigenvalue, and (iii) fastest mixing time, similar to how Cheeger's inequality connects (i) edge conductance, (ii) the second eigenvalue, and (iii) mixing time. Building on this connection, we develop a new spectral theory for vertex expansion via reweighted eigenvalues. We define $\lambda_k^*(G)$ and discover that several interesting generalizations of Cheeger's inequality have close analogues for vertex expansion.

Theorem 1.1.3 (Informal)**.** *Several generalizations of Cheeger's inequality [Tre09, LOT12, LRTV12, KLL⁺13] relating higher eigenvalues and edge conductance-type quantities all have analogues relating higher reweighted eigenvalues and vertex expansion-type quantities.*

Finally, inspired by this connection, we present negative evidence to the 0/1-polytope edge expansion conjecture by Mihail and Vazirani (see [FM92]), by constructing 0/1-polytopes whose graphs have very poor vertex expansion and hence by Theorem 1.1.2 has slow fastest mixing time.

[3]The definition of (π-weighted) vertex expansion can be found in Definition 4.1.2.

1.2 Towards A Spectral Theory for Directed Graphs and Hypergraphs

For directed graphs, one may define directed analogues of edge conductance and vertex expansion. From the spectral theory of undirected graphs, one might expect a spectral theory for directed graphs that has applications in expander certification, directed graph partitioning, mixing time analysis, and so on.

In contrast to the spectral theory for undirected graphs, however, the spectral theory for directed graphs has not been nearly as well developed. One major issue is that the Laplacian matrix of a directed graph is not Hermitian, and so its eigenvalues are not necessarily real numbers. There are formulations [Fil91, Chu05, GM17, LL15] that associate certain Hermitian matrices to a directed graph, and use the second eigenvalue of these matrices to bound the mixing time of random walks [Fil91, Chu05]. But, to our knowledge, there are no known formulations that relate the expansion properties of a directed graph to the eigenvalues of an associated matrix[4].

The second research question that we address in this thesis is to develop a "good" spectral theory for directed graphs that relates to their expansion properties. We propose such a formulation using reweighted eigenvalues and develop a spectral theory for directed graphs whose utility is comparable to that for undirected graphs.

As before, we find an "optimal" reweighting of the directed graph, in the sense that the second smallest eigenvalue of an appropriately defined Laplacian of the reweighted graph is maximized. To deal with the inherent asymmetry of directed graphs, our idea is to require that the reweighted graph be *Eulerian*. In addition, we require that the reweighted subgraph satisfy edge capacity constraints (for directed edge conductance) or vertex capacity constraints (for directed vertex expansion). Below is the formal definition of the reweighted second eigenvalue with edge capacity constraints.

Definition 1.2.1 (Maximum Reweighted Spectral Gap with Edge Capacity Constraints)**.** *Given a weighted directed graph $G = (V, E, w)$ with edge weights $w : E \to \mathbb{R}^+$, the maximum*

[4]The only spectral formulation that we know about expansion properties of a directed graph is a *nonlinear* Laplacian operator in [Yos16, Yos19]. See Section 3.3.2 for details.

reweighted spectral gap with edge capacity constraints is defined as

$$\begin{aligned}
\vec{\lambda}_2^{e*}(G) \;:=\; \max_{A\geq 0} \quad & \lambda_2\left(D^{-\frac{1}{2}}\left(D_A - \frac{A+A^T}{2}\right)D^{-\frac{1}{2}}\right) & \\
\text{subject to} \quad & A(u,v) = 0 & \forall uv \notin E \\
& \sum_{v\in V} A(u,v) = \sum_{v\in V} A(v,u) & \forall u \in V \\
& A(u,v) \leq w(uv) & \forall uv \in E
\end{aligned}$$

where A is the adjacency matrix of the reweighted Eulerian subgraph, D_A is the diagonal degree matrix of $(A + A^T)/2$, and D is the diagonal degree matrix of G. We remark that $\vec{\lambda}_2^{e*}(G)$ is an SDP, and hence can be solved in polynomial time.

In comparison, the well-known result of Fill [Fil91] and Chung [Chu05] can be interpreted as looking at one fixed Eulerian reweighting using the stationary distribution of the directed graph. See Section 3.3.1 for more details.

Our first main result is a Cheeger inequality relating directed edge conductance and reweighted eigenvalues with edge capacity constraints. Hence, parallel to how Laplacian eigenvalues can be used to certify undirected expander graphs, reweighted eigenvalues can be used to certify directed expander graphs.

Theorem 1.2.2 (Cheeger Inequality for Directed Edge Conductance [LTW23])**.** *For any directed graph G, let $\vec{\phi}(G)$ be its directed edge conductance and $\vec{\lambda}_2^{e*}(G)$ be its edge-capacitated reweighted eigenvalue. Then,*

$$\vec{\lambda}_2^{e*}(G) \lesssim \vec{\phi}(G) \lesssim \sqrt{\vec{\lambda}_2^{e*}(G) \cdot \log \frac{1}{\vec{\lambda}_2^{e*}(G)}}.$$

Our second main result is a Cheeger inequality relating directed vertex expansion and reweighted eigenvalues with vertex capacity constraints. It turns out that the latter is closely related to fastest mixing time in general Markov chains, and so our result relates (i) directed vertex expansion, (ii) Eulerian reweighted eigenvalue (with vertex capacity constraints), and (iii) fastest mixing time in general Markov chains.

Theorem 1.2.3 (Informal)**.** *Similar to Theorem 1.1.2, there is a Cheeger-type inequality relating the directed vertex expansion and the vertex-capacitated reweighted eigenvalue of a directed graph. As a corollary, small directed vertex expansion is provably the only obstacle to fastest mixing on general Markov chains.*

This is the first known combinatorial characterization of fastest mixing on general Markov chains.

Finally, we show that reweighted eigenvalues can be used to derive Cheeger inequality and generalizations for hypergraph conductance, either recovering or improving on previous results by Louis [Lou15] and Chan, Louis, Tang, Zhang [CLTZ18], as well as being conceptually simpler.

1.3 Submodular Transformations

So far, the reweighted eigenvalues framework has been successful in building a spectral theory for graph-like settings. The third research question that we address in this thesis is a natural follow-up to the investigation so far: to derive a spectral theory for a general class of problems that encompasses these graph-like settings, using reweighted eigenvalues.

All the expansion quantities considered thus far may be considered as the ratio between the value of the "cut function" evaluated at a subset and the value of the "size function" evaluated at the subset. For example, for (unweighted) vertex expansion on undirected graphs, the cut function is the number of the vertices of the neighborhood, and the size function is just the number of vertices.

Yoshida [Yos19] and Li and Milenkovic [LM18] considered a general class of cut functions that enjoy a property called *submodularity*. There are two main reasons why submodularity is interesting. On the one hand, all graph-like cut functions that we have considered thus far can be shown to be submodular. On the other hand, submodularity can be understood as capturing a diminishing return property that permeates many application domains, such as graph theory, machine learning, economics, and game theory [LM18, Yos19]. They developed a spectral theory for submodular transformations, which included a non-algorithmic Cheeger inequality and an SDP relaxation to make the theory algorithmic. Their result is general, but fails to capture some key applications in graph settings, such as expander certification and mixing time analysis.

We thus define reweighted eigenvalues and attempt to build an alternative spectral theory for submodular transformations. On the positive side, we prove that for a class of "simple" submodular transformations, there is a Cheeger inequality relating reweighted eigenvalues and expansions. This class corresponds exactly to directed hypergraph conductance, which is a nice common generalization of all the graph expansion problems considered thus far.

Theorem 1.3.1 (Informal)**.** *There is a Cheeger-type inequality relating the conductance and the reweighted second smallest eigenvalue of a directed hypergraph.*

On the negative side, we show that the reweighted eigenvalue framework faces serious difficulty in tackling general submodular transformations. While we do not have counter-examples that definitively rule out a spectral theory for submodular transformations via reweighted eigenvalues, such a theory, if exists, would likely require a significant deviation from the current approach. Therefore, our investigation suggests that directed hypergraphs lies at the boundary of effectiveness of reweighted eigenvalues.

1.4 Upper Bounds on Graph Reweighted Eigenvalues

One other application of classical spectral theory is to find small, balanced separators in certain graphs. A balanced separator is a vertex subset $S \subseteq V$, whose removal breaks the remaining graph into connected components each of size at most (say) $2|V|/3$.

One classical result about balanced separators is the planar separator theorem proven by Lipton and Tarjan [LT79], which states that every planar graph $G = (V, E)$ admits a balanced separator of size $O(\sqrt{|V|})$. The proof is purely combinatorial and gives a linear-time algorithm for finding such a separator, given an explicit planar embedding of the graph. Moreover, by considering the grid graph, it can be shown that the separator size of $O(\sqrt{|V|})$ is optimal up to constants.

The objective of the work by Spielman and Teng [ST96], then, is to show that spectral partitioning algorithms attain a similar guarantee for planar graphs, without the prior knowledge of a planar embedding. They do so by proving the following eigenvalue bound for planar graphs.

Theorem 1.4.1 ([ST96, Theorem 3.3])**.** *For any planar graph $G = (V, E)$ with maximum degree Δ, $\lambda_2'(G) \lesssim \Delta/|V|$. Here, $\lambda_2'(G)$ is the second smallest eigenvalue of the unnormalized Laplacian (see Section 2.5 for definition).*

This implies that $\lambda_2(G) \leq \lambda_2'(G) \leq O(1/|V|)$ for bounded-degree graphs, and by Cheeger's inequality, the sweep-cut algorithm finds a set $S \subseteq V$ with edge conductance $O(\sqrt{1/|V|})$, and $|S| \leq |V|/2$. Therefore, the recursive spectral partitioning algorithm that repeatedly applies the sweep-cut algorithm, adding S to one of the components and the neighbors of S to the separator, matches the $O(\sqrt{|V|})$ separator size guarantee of the planar separator theorem, albeit only in the special case of bounded degree graphs. In general, their result gives an $O(\sqrt{\Delta \cdot |V|})$ upper bound on the separator size.

In [ST96], they conjectured that similar eigenvalue bounds hold for graphs with bounded genus g and graphs which are K_h-minor free (see Section 3.5.3 for relevant definitions). These conjectures were answered affirmatively by Kelner [Kel06] in the bounded genus case and Biswal, Lee, Rao [BLR10] in the K_h-minor free case. Later, these results were further extended by Kelner, Lee, Price, and Teng [KLPT11] to address higher eigenvalues $\lambda'_k(G)$. We will review some of these results in Section 3.5.

We remark that the non-spectral algorithms by Lipton and Tarjan [LT79] (for planar graphs) and by Gilbert, Hutchinson, and Tarjan [GHT84] (for bounded genus graphs) guarantee smaller separator sizes, and these algorithms run in linear time *given an explicit embedding of the graph in low-genus surfaces*. However, computing the genus of a graph is NP-hard [Tho89]; the current best polynomial-time algorithms [CS13, KS15, KS19] can only compute an embedding of the graph in genus $O(g \operatorname{polylog}(n))$ surface for genus g graphs. Thus, spectral methods have the crucial advantage that the computationally difficult task of computing a low-genus embedding of the given graph is bypassed.

Our fourth research question is well motivated in this context: can we upper-bound reweighted eigenvalues for these special graph classes? Such upper bounds will provide an alternative spectral approach to finding small separators. We focus on reweighted eigenvalues for undirected graphs, which relate to vertex expansions in undirected graphs and appear to be most relevant to vertex-based partitioning algorithms.

Theorem 1.4.2 (Reweighted Second Eigenvalue Upper Bound)**.** *Let $G = (V, E)$ be an undirected graph with n vertices, and $\pi = \vec{1}/n$ be the uniform distribution on V.*

- *If G is a planar graph, then $\lambda_2^*(G) \leq O(1/n)$.*
- *If G is a graph with genus $g \geq 0$, then $\lambda_2^*(G) \leq O((g+1)\log^2(g+1)/n)$.*
- *If G is a graph which is K_h-minor free for some $h \geq 3$, then $\lambda_2^*(G) \leq O(h^6 \log h/n)$.*

Notably, these upper bounds have no dependence on the maximum degree Δ of the graph. As a result, spectral partitioning using the second reweighted eigenvalue has the best-known performance guarantee of vertex-based graph partitioning algorithms on these classes of graphs, without the requirement that the input graph be of bounded degree. For instance, for planar graphs Theorem 1.4.2 imply $O(\sqrt{(\log \Delta) \cdot n})$-sized balanced separators using reweighted eigenvalues[5], which improves over the $O(\sqrt{\Delta \cdot n})$ bound in [ST96].

[5] This can be further improved to $O(\sqrt{n})$, which is optimal up to constants. See Chapter 7.

By adapting the proofs in [KLPT11], we obtain similar upper bounds on the higher reweighted eigenvalues $\lambda_k^*(G)$. Together, these results illustrate the power of reweighted eigenvalues in lifting results in basic settings to results in generalized settings.

1.5 ℓ_2^2 Triangle Inequalities

In their groundbreaking work, Arora, Rao and Vazirani [ARV09] designed a polynomial-time $O(\sqrt{\log n})$-approximation algorithm for edge conductance. Their algorithm is based on rounding the solution to an SDP, which can be considered as adding the so-called ℓ_2^2 triangle inequality constraints to the second eigenvalue $\lambda_2(G)$, viewed as a minimization problem. The core of their proof is a geometric structure theorem for graph embeddings respecting the ℓ_2^2 triangle inequalities. Subsequently, fast algorithms [AK07, AHK05, She09] have been developed for obtaining the ARV approximation guarantee.

The key geometric argument of [ARV09] was then used to design $O(\sqrt{\log n})$ approximation algorithms for other graph expansion problems [ACMM05, FHL08, CS18]. These SDPs, however, are more complicated and look different from each other. Thus, our fifth research question is to find a unifying approach for deriving $O(\sqrt{\log n})$ approximation algorithms for graph expansions.

We discover that, by just adding ℓ_2^2 triangle inequality constraints to reweighted eigenvalues, we can design $O(\sqrt{\log n})$-approximation algorithms in settings as general as directed hypergraph expansion. These formulations lead to faster and simpler algorithms for computing small-expansion cuts; see [LTW24].

We also study multi-way variants of expansion problems. The goal of these problems is to find k disjoint vertex subsets, such that each of the subsets have small expansion. Past work [LOT12, LRTV12] has shown that (informally) there exist k disjoint vertex subsets of small edge conductance if and only if the k-th smallest Laplacian eigenvalue is small. Using reweighted eigenvalues [KLT22], we have also shown that there exist k disjoint vertex subsets of small vertex expansion if and only if the k-th smallest reweighted eigenvalue is small. These results, however, also incur the square-root loss that is expected of spectral methods.

Again, we show that adding ℓ_2^2 triangle inequality constraints to the k-th reweighted eigenvalue provides a much tighter approximation of the relevant multi-way expansion quantity. We use the orthogonal separators technique [CMM06, BFK$^+$14, LM14a, LM14b], which plays the role of the ARV geometric argument.

To summarize, we have demonstrated that reweighted eigenvalues can be easily combined with the powerful technique of adding ℓ_2^2 triangle inequalities to yield the best-known approximation algorithms to a general class of expansion problems.

1.6 Organization of the Thesis

In Chapter 2, we introduce the preliminary concepts and notations that will be used in this thesis. These include graph notations, linear algebra, Markov chains and random walks, basic spectral graph theory, convex geometry, flows, and convex optimization.

In Chapter 3, we review and present important results from the literature on which the remainder of the thesis is based. Proofs and proof outlines are included where applicable. The proof structure for the classical Cheeger's inequality deviates slightly from the standard treatise and will serve as a prototype for several main results in the thesis.

In Chapter 4, we investigate vertex expansion in undirected graphs, improving the result in [OZ22] and deriving exact analogues to generalizations of Cheeger's inequality [Tre09, LOT12, LRTV12, KLL+13]. We discuss an application of the result to studying the 0/1-polytope conjecture [FM92].

In Chapter 5, we investigate several expansion quantities in directed graphs and hypergraphs. In particular, we introduce the concept of Eulerian reweighting to define reweighted eigenvalues for directed graphs. We prove Cheeger-type inequalities for these settings, but show that the generalizations of Cheeger's inequality extend only partially to these settings. We discuss applications of these results.

In Chapter 6, we try to unify and generalize the results in the previous two chapters by recognizing those settings as instances of what is known as "submodular transformations". We formulate reweighted eigenvalues and attempt to derive Cheeger-type inequalities for submodular transformations. It turns out that we cannot obtain a strong result in full generality, and we shall give the positive results and discuss the hard instances.

In Chapter 7, we upper bound the value of the reweighted second eigenvalues and higher reweighted eigenvalues, when the given graph has special structures. In increasing order of complexity/generality: planar, of bounded genus g, or H-minor free.

In Chapter 8, we show how adding ℓ_2^2 triangle inequalities to the reweighted eigenvalues formulation leads to tighter SDP relaxations resembling, but also generalizing, those obtained in [ARV09] and [LM14a, LM14b]. This also provides a simpler, unified approach to obtain results in [ACMM05] and [FHL08].

Finally, we conclude the thesis in Chapter 9 with a brief summary and discussion of open problems and future directions.

We strongly suggest that the reader read Section 2.2.1, Section 2.5 and Section 3.1.1 before proceeding to the main chapters. A quick read of the rest of Chapter 2 is also prudent. When reading each main chapter, it may be helpful to revisit the relevant sections in Chapter 3. It is recommended to read Chapter 4, Chapter 5, Chapter 6, and Chapter 8 in this order, while Chapter 7 is best read after Chapter 4.

Chapter 2

Preliminaries

2.1 Mathematical Notations

We use $A \subseteq B$ to denote set inclusion and $A \subset B$ to denote *proper* set inclusion, i.e. $A \subseteq B$ and $A \neq B$. $A \sqcup B$ denotes the disjoint union of A and B. For any set A, 2^A denotes the set of subsets of A, i.e. $2^A := \{B : B \subseteq A\}$. Given a set $A \subseteq B$, we use A^c to denote the complement of A (in B), i.e. $A^c := \{x \in B \mid x \notin A\}$.

$\mathbb{R}^+$ denotes the set of positive real numbers and $\mathbb{R}_{\geq 0}$ denotes the set of nonnegative real numbers. For positive integers k, we use $[k]$ to denote the set $\{1, 2, \ldots, k\}$.

We use $\vec{0}$ to denote the zero vector (of the appropriate dimension). The notation $\mathbb{1}$ when used in isolation denotes the all-one vector (of the appropriate dimension). The notation $\mathbb{1}[\text{event}]$ denotes the indicator variable for that event, i.e. $\mathbb{1}[\text{event}]$ equals 1 if event happens and equals 0 otherwise.

Let D be a finite set. Denote by $\mathbb{R}^D$ the vector space isomorphic to $\mathbb{R}^{|D|}$, and each vector $x \in \mathbb{R}^D$ is a $|D|$-tuple of real numbers with entries indexed by D. So, $x = (x(u))_{u \in D}$. Given $x, x' \in \mathbb{R}^D$, their inner product is $\langle x, x' \rangle := \sum_{u \in D} x(u)x'(u)$. Unless otherwise specified, for a vector $x \in \mathbb{R}^D$, $\|x\|$ denotes the Euclidean 2-norm, i.e. $\|x\| := \sqrt{\sum_{u \in D} x(u)^2}$. $\|x\|_1$ denotes the 1-norm of x, i.e. $\|x\|_1 := \sum_{u \in D} |x(u)|$, and $\|x\|_\infty$ denotes the ∞-norm of x, i.e. $\|x\|_\infty := \max_{u \in D} |x(u)|$.

Given a function $f : D \to \mathbb{R}$ and any subset $S \subseteq D$, denote by $f(S)$ the sum of values of f at $u \in S$, i.e. $f(S) := \sum_{u \in S} f(u)$, and denote by $\text{supp}(f)$ the domain subset on which f is nonzero. Similarly, given a vector $x \in \mathbb{R}^D$ and any $S \subseteq D$, let $x(S) := \sum_{u \in S} x(u)$. $\mathbb{1}_S \in \mathbb{R}^D$ denotes the indicator vector of S in D, i.e. $\mathbb{1}_S(u) = 1$ if $u \in S$ and $\mathbb{1}_S(u) = 0$

if $u \notin S$. We abuse notation slightly and use $\mathbb{1}_u$ in lieu of $\mathbb{1}_{\{u\}}$ for $u \in D$. The functions $f^+, f^- : D \to \mathbb{R}$ are known respectively as the positive part and the negative part of f, defined as $f^+(u) := \max(f(u), 0)$ and $f^-(u) := \max(-f(u), 0)$.

Given a function $f : D \to \mathbb{R}^k$ for some k, and another map $w : D \to \mathbb{R}_{\geq 0}$, define the w-mass, or simply mass when the context is clear, of f to be $\|f\|_w^2 := \sum_{u \in D} w(u) \|f(u)\|^2$.

Assuming familiarity with the standard "big-O" notations, we use $f \lesssim g$ to denote $f = O(g)$, $f \gtrsim g$ to denote $f = \Omega(g)$, and $f \asymp g$ to denote $f = \Theta(g)$. We use $f \lesssim_\varepsilon g$ and $f = O_\varepsilon(g)$ to hide the dependence on ε, meaning that $f \lesssim g$ and $f = O(g)$ when ε is fixed. Somewhat unconventionally, we use $f \approx_k g$ to denote $f/g = 1 + o_k(1)$.

2.2 Graphs, Directed Graphs, and Hypergraphs

Undirected graphs: Let $G = (V, E, w)$ be a (weighted) undirected graph, or simply (weighted) graph, with edge weights $w : E \to \mathbb{R}_{\geq 0}$. If uv is an edge in G, we either write $uv \in E$ or use the notation $u \sim v$. The weighted degree of a vertex $v \in V$, denoted by $\deg_w(v)$, is defined as $\deg_w(v) := \sum_{u:uv \in E} w(uv)$. The maximum weighted degree of a graph, denoted $\Delta_w(G)$ or simply Δ_w, is defined as $\Delta_w(G) := \max_{v \in V} \deg_w(v)$.

We use the notation $\deg(v) := \deg_{\mathbb{1}}(v)$ and $\Delta(G) := \Delta_{\mathbb{1}}(G)$ to denote the combinatorial degree of a vertex $v \in V$ and the maximum (combinatorial) degree of G, respectively. When specialized to $w = \mathbb{1}$, the graph G is unweighted; we use the notation $G = (V, E)$ for the graph, and the degree notations $\deg(v)$ and $\Delta(G)$ as above.

Given $V' \subseteq V$, the induced edge set of V' is defined as $E[V'] := \{uv \in E \mid u \in V' \text{ and } v \in V'\}$, and the induced graph of V' is the graph $G[V'] := (V', E[V'])$. Definitions relevant to expansion quantities are deferred to the next section.

Directed graphs: Let $G = (V, E, w)$ be a (weighted) directed graph, with arc weights $w : E \to \mathbb{R}_{\geq 0}$. We sometimes write "edges" instead of "arcs". If uv is an arc in G, we either write $uv \in E$ or use the notation $u \to v$. The weighted indegree of a vertex v is defined as $\deg_w^-(v) := \sum_{u:u \to v} w(uv)$. The weighted outdegree of a vertex v is similarly defined as $\deg_w^+(v) := \sum_{u:v \to u} w(vu)$. The total weighted degree, or simply weighted degree, of a vertex v is the sum of its indegree and outdegree, i.e. $\deg_w(v) := \deg_w^-(v) + \deg_w^+(v)$. The maximum weighted degree of G, again denoted $\Delta_w(G)$ or Δ_w, is defined as $\Delta_w(G) := \max_{v \in V} \deg_w(v)$.

Similar to the undirected case, we use $\deg^-(v)$, $\deg^+(v)$, $\deg(v)$, and $\Delta(G)$ to denote the combinatorial/unweighted counterparts of these definitions, and when specialized to unweighted directed graphs we simplify the notation to $G = (V, E)$.

A directed graph is called Eulerian if $\deg^-_w(v) = \deg^+_w(v)$ for all $v \in V$. Definitions relevant to expansion quantities are deferred to the next section.

An undirected graph $G = (V, E, w)$ can be regarded as a directed graph by the following bidirection $G' = (V, E', w')$ where $E' := \{u \to v : uv \in E\} \cup \{v \to u : uv \in E\}$ and the arc weights inheriting the edge weights.

Hypergraphs: A (weighted) hypergraph is a 3-tuple $H = (V, E, w)$ where each $e \in E$, called a hyperedge, is a vertex subset $e \subseteq V$, and $w : E \to \mathbb{R}_{\geq 0}$ is the (hyper)edge weight function. The size of a hyperedge $e \in E$ is $|e|$ and the rank $r(H)$ of a hypergraph is defined as the maximum hyperedge size, i.e. $r(H) := \max_{e \in E} |e|$. The weighted degree of a vertex $v \in V$, denoted by $\deg_w(v)$, is defined as the sum of weights of hyperedges that contain v, i.e. $\deg_w(v) := \sum_{e : v \in e} w(e)$.

We again use $\deg(v)$ to denote the combinatorial/unweighted counterpart of $\deg_w(v)$, and when specialized to unweighted hypergraphs we simplify the notation to $H = (V, E)$.

An undirected graph $G = (V, E, w)$ can be regarded as a hypergraph $H = (V, E', w')$ where $E' := \{\{u, v\} \mid uv \in E\}$ is such that each hyperedge is of size two, and the hyperedge weights inheriting the edge weights.

Directed Hypergraphs: A (weighted) hypergraph is a 3-tuple $H = (V, E, w)$, where each $e \in E$, called a hyperedge (of H), is of the form $e = (e^-, e^+)$, where e^-, e^+ are called the source set and the sink set of e respectively and are nonempty subsets of V, and $w : E \to \mathbb{R}_{\geq 0}$ is the (hyper)edge weight function. Note that e^- and e^+ need not be disjoint. The size of a hyperedge $e \in E$ is $|e^- \cup e^+|$. The rank $r(H)$ of a directed hypergraph is defined as the maximum hyperedge size, i.e. $r(H) := \max_{e \in E} |e^- \cup e^+|$. The weighted indegree of a vertex $v \in V$, denoted by $\deg^-_w(v)$, is defined as $\sum_{e : v \in e^+} w(e)$. The weighted outdegree of a vertex $v \in V$, denoted by $\deg^+_w(v)$, is defined as $\sum_{e : v \in e^-} w(e)$. The weighted total degree of a vertex $v \in V$ is defined as $\deg_w(v) := \deg^-_w(v) + \deg^+_w(v)$.

We again use $\deg^-(v), \deg^+(v)$, and $\deg(v)$ to denote the combinatorial/unweighted counterparts of these definitions, and when specialized to unweighted directed hypergraphs we simplify the notation to $H = (V, E)$.

A hypergraph $H = (V, E, w)$ can be regarded as a directed hypergraph $H' = (V, E', w')$ where $E' := \{(e, e) \mid e \in E\}$, and $w'((e, e)) = w(e)$ for all $e \in E$. Therefore, a graph can be regarded as a directed hypergraph as well.

A directed graph $G = (V, E, w)$ can also be regarded as a directed hypergraph $H'' = (V, E'', w'')$ where $E'' := \{(\{u\}, \{v\}) \mid uv \in E\}$, and $w''((\{u\}, \{v\})) = w(uv)$ for all $uv \in E$.

Hence, directed hypergraph captures all the graph models that we consider in this thesis.

2.2.1 Some Conventions

We use the term "generalized graph" to refer to any of the graph models defined in this section. Given a generalized graph G on vertex set V, a vertex measure is a function $\pi : V \to \mathbb{R}^+$, and a probability distribution is a vertex measure π satisfying $\pi(V) = 1$.

Throughout the thesis, we use $n := |V|$ to denote the number of vertices and $m := |E|$ to denote the number of edges in a generalized graph. We also assume that the given generalized graphs have no isolated vertices, i.e. that each vertex has nonzero degree.

2.3 Expansion Quantities

2.3.1 Expansion Quantities for Undirected Graphs

Edge Expansion: Let $G = (V, E, w)$ be a graph and $S \subseteq V$ be a subset of vertices. The edge boundary of S is defined as $\delta(S) := \{uv \in E \mid u \in S, v \notin S\}$. The volume of S is defined as $\mathrm{vol}_w(S) := \sum_{v \in S} \deg_w(v)$, and denoted simply as $\mathrm{vol}(S)$ for unweighted graphs.

Let $\pi : V \to \mathbb{R}^+$ be a vertex measure. The π-weighted edge expansion of a subset $S \subseteq V$ and of G are defined as

$$\phi_\pi(S) := \frac{w(\delta(S))}{\pi(S)} \quad \text{and} \quad \phi_\pi(G) := \min_{S \subseteq V : 0 < \pi(S) \leq \pi(V)/2} \phi_\pi(S).$$

When $\pi = \deg_w$ is the degree measure, ϕ_π is the *edge conductance* of G, and we use the notation ϕ so that

$$\phi(S) := \frac{w(\delta(S))}{\mathrm{vol}_w(S)} \quad \text{and} \quad \phi(G) := \min_{S \subseteq V : 0 < \mathrm{vol}_w(S) \leq \mathrm{vol}_w(V)/2} \phi(S).$$

Note that $\phi(S) \leq 1$ always.

When $\pi = \mathbb{1}$ is the counting measure, ϕ_π is the *edge expansion* or *sparsest cut* of G, and we use the notation φ so that

$$\varphi(S) := \frac{w(\delta(S))}{|S|} \quad \text{and} \quad \varphi(G) := \min_{S \subseteq V: 0 < |S| \leq n/2} \varphi(S).$$

Vertex Expansion: Let $G = (V, E, w)$ be a graph[1] and $S \subseteq V$ be a subset of vertices. The vertex boundary of S is defined as $\partial S := \{v \in V \setminus S \mid \exists u \in S \text{ with } uv \in E\}$.

Let $\pi : V \to \mathbb{R}^+$ be a vertex measure. The π-weighted vertex expansion of a subset $S \subseteq V$ and of G are defined as

$$\psi_\pi(S) := \frac{\pi(\partial S)}{\pi(S)} \quad \text{and} \quad \psi_\pi(G) := \min\left\{1, \min_{S \subseteq V: 0 < \pi(S) \leq \pi(V)/2} \psi_\pi(S)\right\}^{2}.$$

When π is the counting measure or the uniform distribution, $\psi = \psi_\pi$ is the usual vertex expansion with $\psi(S) = |\partial S|/|S|$.

2.3.2 Expansion Quantities for Directed Graphs

Edge Expansion: Let $G = (V, E, w)$ be a directed graph. The following definition of edge expansion for directed graphs is due to Yoshida [Yos16, Yos19]. Perhaps the more widely used expansion quantity is the Cheeger constant $h(G)$ by Fill [Fil91] and Chung [Chu05], which we defer to Section 3.3.1 for review and Chapter 5 for further discussion and comparison.

Let $S \subseteq V$ be a subset of vertices. The set of outgoing edges of S is defined as $\delta^+(S) := \{uv \in E \mid u \in S \text{ and } v \notin S\}$, and the set of incoming edges to S is defined as $\delta^-(S) := \delta^+(S^c)$. The volume of S is defined as $\mathrm{vol}_w(S) := \sum_{v \in S} \deg_w(v)$, so it is the sum of total weighted degrees of vertices in S.

Let $\pi : V \to \mathbb{R}^+$ be a vertex measure. The π-weighted directed edge expansion of a subset $S \subseteq V$ and of G are defined as

$$\vec{\phi}_\pi(S) := \frac{\min\left\{w\big(\delta^+(S)\big), w\big(\delta^+(S^c)\big)\right\}}{\min\left\{\pi(S), \pi(S^c)\right\}} \quad \textit{and} \quad \vec{\phi}_\pi(G) := \min_{\emptyset \neq S \subset V} \vec{\phi}_\pi(S).$$

[1]Note, however, that the definitions below do not involve the edge weights w.

[2]We take the minimum with 1 in the definition of $\psi_\pi(G)$ to avoid dealing with the edge case where there is a vertex v with $\pi(v)/\pi(V)$ close to 1.

So, $\vec{\phi}_\pi(S)$ is big if and only if there are both a lot of incoming edge weight to S and outgoing edge weight from S, relative to the size of the smaller of S and S^c. Note that this definition is compatible with the definition of ϕ_π for undirected graphs.

When $\pi = \deg_w$ is the (total) degree measure, $\vec{\phi}_\pi$ is the *edge conductance* of G:

Definition 2.3.1 (Directed Edge Conductance [Yos16, Yos19])**.** *Let $G = (V, E, w)$ be a directed graph. The directed edge conductance of a set $S \subseteq V$ and of the graph G are defined as*

$$\vec{\phi}(S) := \frac{\min\left\{w\big(\delta^+(S)\big), w\big(\delta^+(S^c)\big)\right\}}{\min\left\{\mathrm{vol}_w(S), \mathrm{vol}_w(S^c)\right\}} \quad \textit{and} \quad \vec{\phi}(G) := \min_{\emptyset \neq S \subset V} \vec{\phi}(S).$$

When $\pi = \mathbb{1}$ is the counting measure, $\vec{\phi}$ is the *edge expansion* or *sparsest cut* of G, and we use the notation $\vec{\varphi}$ so that

$$\vec{\varphi}(S) := \frac{\min\left\{w\big(\delta^+(S)\big), w\big(\delta^+(S^c)\big)\right\}}{\min\left\{|S|, |S^c|\right\}} \quad \text{and} \quad \vec{\varphi}(G) := \min_{\emptyset \neq S \subset V} \vec{\varphi}(S).$$

Vertex Expansion: Let $G = (V, E, w)$ be a directed graph[3] and $S \subseteq V$ be a subset of vertices. $\partial^+(S) := \{v \notin S \mid \exists u \in S \text{ with } uv \in E\}$ be the set of out-neighbors of S.

Definition 2.3.2 (Directed Vertex Expansion)**.** *Let $G = (V, E, w)$ be a directed graph and $\pi : V \to \mathbb{R}^+$ be a vertex measure. The directed vertex expansion of a set $S \subseteq V$ and of the graph G are defined as*[4]

$$\vec{\psi}_\pi(S) := \frac{\min\left\{\pi\big(\partial^+(S)\big), \pi\big(\partial^+(S^c)\big)\right\}}{\min\left\{\pi(S), \pi(S^c)\right\}} \quad \textit{and} \quad \vec{\psi}_\pi(G) := \min_{\emptyset \neq S \subset V} \vec{\psi}_\pi(S).$$

Note that $\vec{\psi}_\pi(S) \leq 1$ for all $S \subseteq V$ as $\partial^+(S^c) \subseteq S$. Note also that this is compatible with the definition of ψ_π for undirected graphs. We write $\vec{\psi}$ instead of $\vec{\psi}_\pi$ when it is clear from context.

[3]Here again, the definitions below do not involve the arc weights w.

[4]When specialized to undirected graphs (by considering the bidirected graph), the current definitions are slightly different from the undirected definitions presented in Section 2.3.1. We remark that the two definitions of $\psi(G)$ are within a factor of 2 of each other. The current definitions have the advantages that $\vec{\psi}_\pi(S) \leq 1$ and are more convenient in the proofs.

2.3.3 Expansion Quantities for Hypergraphs

Undirected Hypergraphs: Let $H = (V, E, w)$ be a hypergraph and $S \subseteq V$ be a subset of vertices. Let $S \subseteq V$ be a subset of vertices. The edge boundary of S is defined as $\delta(S) := \{e \in E \mid e \cap S \neq \emptyset \text{ and } e \cap \overline{S} \neq \emptyset\}$. The volume of S is defined as $\mathrm{vol}_w(S) := \sum_{v \in S} \deg_w(v)$.

Let $\pi : V \to \mathbb{R}^+$ be a vertex measure. The π-weighted hypergraph edge expansion of a subset $S \subseteq V$ and of H are defined as

$$\phi_\pi(S) := \frac{w(\delta(S))}{\min(\pi(S), \pi(S^c))} \qquad \text{and} \qquad \phi_\pi(H) := \min_{\emptyset \neq S \subset V} \phi_\pi(S).$$

This is the same notation as π-weighted edge expansion for ordinary graphs. We do not disambiguate the two when the context is clear.

When $\pi = \deg_w$ is the degree measure, ϕ_π is the *hypergraph edge conductance* of H:

Definition 2.3.3 (Hypergraph Edge Conductance [Lou15, CLTZ18])**.** *Let $H = (V, E, w)$ be an undirected hypergraph. The hypergraph edge conductance of a set $S \subseteq V$ and of the hypergraph H are defined as*

$$\phi(S) := \frac{w\big(\delta(S)\big)}{\min\big\{\mathrm{vol}_w(S), \mathrm{vol}_w(S^c)\big\}} \quad \textit{and} \quad \phi(H) := \min_{\emptyset \neq S \subset V} \phi(S).$$

We note that there is a significant drawback to this definition of hypergraph edge conductance — namely, that any hyperedge e that contributes $w(e)$ to the cut weight in the numerator may contribute $\Theta(|e| \cdot w(e))$ to the volume in the denominator. Indeed, one can show that $\phi(H) \leq O(1/r)$ if H is r-uniform, i.e. all hyperedges in H are of size r.

Directed Hypergraphs: Let $H = (V, E, w)$ be a directed hypergraph and $S \subseteq V$ be a subset of vertices. The set of outgoing hyperedges of S is defined as $\delta^+(S) := \{e \in E \mid e^- \cap S \neq \emptyset \text{ and } e^+ \cap S^c \neq \emptyset\}$, and the set of incoming hyperedges to S is defined similarly as $\delta^-(S) := \{e \in E \mid e^- \cap S^c \neq \emptyset \text{ and } e^+ \cap S \neq \emptyset\}$. The volume of S is defined as $\mathrm{vol}_w(S) := \sum_{v \in S} \deg_w(v)$, so it is the sum of total weighted degrees of vertices in S.

Let $\pi : V \to \mathbb{R}^+$ be a vertex measure. The π-weighted directed hypergraph edge expansion of a subset $S \subseteq V$ and of H are defined as

$$\vec{\phi}_\pi(S) := \frac{\min\big\{w\big(\delta^+(S)\big), w\big(\delta^+(S^c)\big)\big\}}{\min\big\{\pi(S), \pi(S^c)\big\}} \quad \text{and} \quad \vec{\phi}_\pi(H) := \min_{\emptyset \neq S \subset V} \vec{\phi}_\pi(S).$$

When $\pi = \deg_w$ is the degree measure, $\vec{\phi}_\pi$ is the directed hypergraph edge conductance, which we denote simply by $\vec{\phi}$. Again, this notation coincides with that for directed graphs, and we do not disambiguate when the context is clear.

Remark 2.3.4 (Hypergraph Vertex Expansion)**.** *It is possible to define vertex expansion for hypergraphs as the minimum ratio between $\pi(\partial S)$ and $\pi(S)$, where*

$$\partial S := \{v \in V \setminus S : \exists e \in \delta(S) \text{ s.t. } v \in e\}.$$

However, this definition is not more powerful than vertex expansion for ordinary graphs, as we may reduce the former to the latter by considering the clique-graph of the hypergraph, which is formed by drawing an edge between every pair $u, v \in V$ for which a hyperedge e exists such that $u, v \in e$. Then, a vertex is in ∂S in the hypergraph if and only if a vertex is in ∂S in the clique-graph. A similar reduction from vertex expansion for directed hypergraphs to vertex expansion for directed graphs is possible.

2.3.4 Some Reductions between Expansion Quantities

In this subsection, we collect reductions between expansion quantities that will be useful later.

Undirected Vertex Expansion to Hypergraph Edge Expansion

The reduction result presented below has two steps. The first step is to reduce from vertex expansion to "symmetric vertex expansion" defined below, and the second step is to reduce from symmetric vertex expansion to hypergraph edge expansion.

Definition 2.3.5 (Symmetric Vertex Expansion)**.** *Let $G = (V, E)$ be a graph with vertex distribution $\pi : V \to \mathbb{R}^+$. The symmetric vertex boundary of a set $S \subseteq V$ is defined as $\partial_{\text{sym}}(S) := \partial(S) \cup \partial(V \setminus S)$, the symmetric vertex expansion of S is defined as $\psi_{\text{sym}}(S) := \pi(\partial_{\text{sym}}(S))/\pi(S)$, and the symmetric vertex expansion of G is defined as*

$$\psi_{\text{sym}}(G) := \min\left\{1, \min_{S: 0 < \pi(S) \le 1/2} \psi_{\text{sym}}(S)\right\}.$$

Proposition 2.3.6 (Vertex Expansion to Symmetric Vertex Expansion [LRV13, Theorem A.2])**.** *Let $G = (V, E)$ be a graph with a distribution $\pi : V \to \mathbb{R}^+$ on the vertices. There exists a graph $G' = (V', E')$ having $O(|V| + |E|)$ vertices with a distribution $\pi' : V' \to \mathbb{R}^+$, computable in polynomial time, such that $\psi(G) = \Theta(\psi_{\text{sym}}(G'))$.*

Moreover, there is a map ι that takes subsets $S' \subseteq V'$ to subsets $S \subseteq V$ such that:

- $\psi(S) \leq \psi_{\mathrm{sym}}(S')/(1-\psi_{\mathrm{sym}}(S'))$;
- $\pi(S) \leq 2\pi'(S')$; and
- If S'_1 and S'_2 are disjoint subsets of V', then $\iota(S'_1)$ and $\iota(S'_2)$ are disjoint.

Proposition 2.3.7 (Symmetric Vertex Expansion to Hypergraph Edge Expansion [LM14b, Theorem 1.6])**.** *Let $G' = (V', E')$ be a graph with a distribution $\pi' : V \to \mathbb{R}^+$ on the vertices. Then, there exists a hypergraph $H = (V'', E'', w'')$, with a distribution $\pi'' : V'' \to \mathbb{R}^+$, such that $\psi_{\mathrm{sym}}(G') = \phi(H)$. Moreover, the construction satisfies:*

- $V'' = V'$;
- *For any subset $S' \subseteq V'$ we have $\phi(S') = \psi_{\mathrm{sym}}(S')$; and*
- *The rank $r(H)$ of H equals $\Delta(G') + 1$ where $\Delta(G')$ is the maximum degree of G'.*

From Edge Conductance in Graphs to Edge Expansion in Graphs

The following reduction is standard; for reference see e.g. [ARV09].

Proposition 2.3.8 (Reduction from Edge Conductance to Edge Expansion [ARV09])**.** *Let $G = (V, E, w)$ be a graph with n vertices and m edges. Then, one can construct in polynomial time a graph $G' = (V', E', w')$ with $|V'| = O(mW)$ vertices where $W := (\max_{e\in E} w(e))/(\min_{e\in E} w(e))$, such that $\varphi(G') = \Theta(\phi(G))$. In particular, if G is unweighted, then G' has $O(m)$ vertices.*

2.4 Linear Algebra

2.4.1 Basic Facts

Let $M \in \mathbb{R}^{d\times d}$ be a matrix. When M is symmetric, the spectral theorem states that M admits an orthonormal eigendecomposition $M = UDU^{-1}$, where $D \in \mathbb{R}^{d\times d}$ is a diagonal matrix and U is a unitary matrix such that $U^{-1}U = I_d$ where I_d is the $d \times d$ identity matrix.

Two matrices $M, N \in \mathbb{R}^{d\times d}$ are said to be cospectral if they are both diagonalizable, and their eigenvalues are the same. There are two well-known cases of cospectral matrices that we will use; see e.g. [Str16, Chapter 6.2].

Fact 2.4.1. *Let $M, N \in \mathbb{R}^{d\times d}$. Suppose that M is diagonalizable and that M and N are similar (i.e. $M = X^{-1}NX$ for some invertible matrix $X \in \mathbb{R}^{d\times d}$). Then, N is also diagonalizable, and M and N are cospectral.*

Fact 2.4.2. *Let $M, N \in \mathbb{R}^{d\times d}$. Suppose that there exist $A, B \in \mathbb{R}^{d\times d}$ such that $M = AB$ and $N = BA$. If M is diagonalizable, then N is also diagonalizable, and M and N are cospectral.*

Given that M is symmetric, we say that M is positive semidefinite (PSD) if $x^T Mx \geq 0$ for all $x \in \mathbb{R}^d$, and we write $M \succeq 0$. Equivalently, M is PSD if all its eigenvalues are nonnegative. Also equivalently, M is PSD if there exists X such that $M = X^T X$. Let $x_i \in \mathbb{R}^d$ be the i-th column of X. Then M is called the Gram matrix of $x_1, \ldots, x_d \in \mathbb{R}^d$ as $M(i, j) = \langle x_i, x_j\rangle$ for all $i, j \in [d]$.

The trace of a matrix $M \in \mathbb{R}^{d\times d}$ is defined as $\operatorname{tr}(M) := \sum_{i=1}^{d} M(i, i)$. We will often use the fact that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for two matrices of compatible dimensions. Given two matrices $M, N \in \mathbb{R}^{d\times d}$, their inner product is defined as $\langle M, N\rangle := \operatorname{tr}(M^T N)$. They are orthogonal iff $\langle M, N\rangle = 0$, in which case we write $M \perp N$.

2.4.2 Computational Aspects

Let $M \in \mathbb{R}^{d\times d}$ be a PSD matrix, and let $\alpha_1 \geq 0$ be its largest eigenvalue. The power method is useful for computing α_1 and an associated eigenvector. It is an iterative algorithm that starts with a random vector and take many steps of update to bring it close to the desired eigenvector, in the sense that the value of the normalized quadratic form (i.e. the Rayleigh quotient) satisfies

$$\frac{x^T Mx}{x^T x} \approx \alpha.$$

We summarize the guarantees of the algorithm in Proposition 2.4.3 below. The interested reader can consult e.g. [Tre16, Chapter 4] for details.

Proposition 2.4.3 (Power Method Guarantee)**.** *Let $M \in \mathbb{R}^{d\times d}$ be a PSD matrix and α_1 be its largest eigenvalue. There is a randomized algorithm that, given $\varepsilon > 0$, runs in time $O(\varepsilon^{-1}\operatorname{nnz}(M) \cdot \log d)$ time and finds a vector $x \in \mathbb{R}^d$ such that*

$$\frac{x^T Mx}{x^T x} \geq (1 - \varepsilon)\alpha_1.$$

Here, $\operatorname{nnz}(M)$ denotes the number of nonzero entries of M.

Once the power method returns an approximate largest eigenvalue $\tilde{\alpha}_1$ with corresponding normalized eigenvector $\tilde{f}_1$ is found, we may apply the power method to the matrix $M_1 := M - \alpha \cdot \tilde{f}_1 \tilde{f}_1^T$ to find an approximate second largest eigenvalue of M, but the number of nonzero entries of M_1 might be large. A workaround is to eliminate the $\tilde{f}_1$ component of the starting random vector, by projecting it to the orthogonal complement of $\tilde{f}_1$ before the update steps.

2.4.3 Courant-Fischer Theorem

Let $M \in \mathbb{R}^{d\times d}$ be a symmetric matrix with real eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_d$. The following theorem characterizes the eigenvalues of M as the value of a min/max problem.

Proposition 2.4.4 (Courant-Fischer Theorem; c.f. [Spi19, Chapter 2])**.** *Let $M \in \mathbb{R}^{d\times d}$ be a symmetric matrix, with $\lambda_k \in \mathbb{R}$ being the k-th smallest eigenvalue for $1 \le k \le d$. Then,*

$$\lambda_k = \min_{\substack{Q:\,\text{subspace of }\mathbb{R}^d\\ \dim(Q)=k}} \max_{f\in Q} \frac{f^T M f}{f^T f}.$$

Proof. Let $f_1, \ldots, f_d \in \mathbb{R}^d$ be an orthonormal eigenbasis of M, so that $Mf_i = \lambda_i f_i$ and $\langle f_i, f_j\rangle = \mathbb{1}[i = j]$. To prove that

$$\lambda_k \ge \min_{\substack{Q:\,\text{subspace of }\mathbb{R}^d\\ \dim(Q)=k}} \max_{f\in Q} \frac{f^T M f}{f^T f},$$

take Q to be the subspace spanned by the first k eigenvectors $f_1, \ldots, f_k$, so that each $f \in Q$ can be written as $f = \sum_{i\in[k]} \alpha_i f_i$. Then,

$$\frac{f^T M f}{f^T f} = \frac{\sum_{i\in[k]} \lambda_i \alpha_i^2}{\sum_{i\in[k]} \alpha_i^2} \le \lambda_k.$$

To prove that

$$\lambda_k \le \min_{\substack{Q:\,\text{subspace of }\mathbb{R}^d\\ \dim(Q)=k}} \max_{f\in Q} \frac{f^T M f}{f^T f},$$

take any dimension-k subspace Q. It must have a nontrivial intersection with the subspace R generated by last last $(d-k+1)$ eigenvectors, i.e. $f_k, f_{k+1}, \ldots, f_d$. Let $f \in Q \cap R$ be nonzero. Then, $f = \sum_{i=k}^d \alpha_i f_i$ for some $\alpha_i \in \mathbb{R}$, and so

$$\frac{f^T M f}{f^T f} = \frac{\sum_{i=k}^d \lambda_i \alpha_i^2}{\sum_{i=k}^d \alpha_i^2} \ge \lambda_k.$$

Taking minimum over all Q yields the inequality. $\square$

2.5 Basic Facts in Spectral Graph Theory

2.5.1 Matrices and Normalization

Given a graph $G = (V, E, w)$, its adjacency matrix $A' = A'(G)$ is an $n \times n$ matrix where the (u, v)-entry is $w(uv)$ (or 0 if $uv \notin E$). Its Laplacian matrix is defined as $L' = L'(G) := D - A'$, where $D = D(G) := \text{diag}(\{\deg_w(v)\}_{v \in V})$ is the diagonal degree matrix. For a vector $x \in \mathbb{R}^n$, the Laplacian matrix has a useful quadratic form

$$x^T L' x = \sum_{uv \in E} w(uv)(x(u) - x(v))^2. \tag{2.1}$$

It turns out to be more useful to consider normalized versions of these matrices. The normalized adjacency matrix is defined as $A = A(G) := D^{-1/2} A' D^{-1/2}$, and the normalized Laplacian matrix is defined as $L = L(G) := D^{-1/2} L' D^{-1/2}$. Here, $D^{-1/2}$ is the inverse square root of the diagonal degree matrix and has entries $\deg_w(v)^{-1/2}$ for $v \in V$. Note that $L = I - A$.

2.5.2 Eigenvalues

Note that $A'(G), L'(G), A(G), L(G)$ defined above are all symmetric matrices, and so are diagonalizable with all real eigenvalues. We denote the eigenvalues of $A'(G)$ (resp. $A(G)$) by $\alpha'_1(G) \geq \alpha'_2(G) \geq \cdots \geq \alpha'_n(G)$ (resp. $\alpha_1(G) \geq \cdots \geq \alpha_n(G)$). We denote the eigenvalues of $L'(G)$ (resp. $L(G)$) by $\lambda'_1(G) \leq \lambda'_2(G) \leq \cdots \leq \lambda'_n(G)$ (resp. $\lambda_1(G) \leq \cdots \leq \lambda_n(G)$).

By the quadratic form (2.1), $L'(G)$ is PSD, and since $\mathbb{1}$ is in the nullspace of $L'(G)$ we have $\lambda'_1(G) = 0$. It follows from the definition of $L(G)$ that $L(G)$ is PSD and has a nontrivial nullspace, and so $\lambda_1(G) = 0$ as well. We will show in the next subsection that:

Fact 2.5.1 (Bound on $\lambda_n(G)$)**.** *For any graph $G = (V, E, w)$, $\lambda_n(G) \leq 2$.*

Since $A(G) = I_n - L(G)$, the spectra of $A(G)$ and $L(G)$ are related by $\alpha_i(G) = 1 - \lambda_i(G)$. Therefore, we can bound the eigenvalues of $A(G)$ by $1 = \alpha_1(G) \geq \cdots \geq \alpha_n(G) \geq -1$.

2.5.3 Eigenvalues as Minimization Problems

The following variational characterization of $\lambda_2(G)$ will prove to be very useful.

Proposition 2.5.2. *For any graph $G = (V, E, w)$,*

$$\lambda_2(G) = \min_{\substack{x:V\to\mathbb{R}\\ x\perp\sqrt{\deg_w}}} \frac{x^T L(G) x}{x^T x} = \min_{\substack{f:V\to\mathbb{R}\\ f\perp\deg_w}} \frac{\sum_{uv\in E} w(uv)(f(u)-f(v))^2}{\sum_{v\in V}\deg_w(v) f(v)^2}. \tag{2.2}$$

Proof. The first equality is by checking that the vector $\mu(v) := \sqrt{\deg_w(v)}$ is the first eigenvector of $L(G)$ (with eigenvalue 0), and then using the Courant-Fischer theorem in Proposition 2.4.4. The second equality is by the change of variables $f := D^{-1/2}x$. □

Given a graph $G = (V, E, w)$ and $f : V \to \mathbb{R}$, we define the Rayleigh quotient of the normalized Laplacian as

$$R(f) := \frac{\sum_{uv\in E} w(uv)(f(u)-f(v))^2}{\sum_{v\in V}\deg_w(v) f(v)^2}.$$

Then, $\lambda_2(G)$ is the optimization problem of minimizing $R(f)$ over vectors $f : V \to \mathbb{R}$ with f nonzero and $\sum_{v\in V}\deg_w(v) f(v) = 0$.

Generally, for a symmetric matrix M, we define the Rayleigh quotient of M as either

$$R(f) := \frac{f^T M f}{f^T f} \quad \text{or} \quad R(f) := \frac{f^T M' f}{f^T Q f} \qquad \text{where } M = Q^{-1/2} M' Q^{-1/2},$$

depending on which one is more convenient.

For higher eigenvalues, the same argument gives the following:

Proposition 2.5.3. *For any graph $G = (V, E, w)$,*

$$\lambda_k(G) = \min_{\substack{Q:\,\text{subspace of } \mathbb{R}^V\\ \dim(Q)=k}} \max_{f\in Q} R(f).$$

Fact 2.5.1 now follows from Proposition 2.5.3 since

$$R(f) = \frac{\sum_{uv\in E} w(uv)(f(u)-f(v))^2}{\sum_{v\in V}\deg_w(v) f(v)^2} \le \frac{2\sum_{uv\in E} w(uv)(f(u)^2+f(v)^2)}{\sum_{v\in V}\deg_w(v) f(v)^2} = 2$$

for all $f : V \to \mathbb{R}$.

The following allows one to upper-bound $\lambda_k(G)$ using functions with disjoint support.

Proposition 2.5.4. *Let $G = (V, E, w)$ be a graph and $f_1, f_2, \ldots, f_k : V \to \mathbb{R}$ be functions with disjoint support. Then,*

$$\lambda_k(G) \leq 2 \max_{i \in [k]} R(f_i).$$

Proof. Let Q be the subspace spanned by $f_1, \ldots, f_k$, which has dimension k and so by Proposition 2.5.3 we have

$$\lambda_k(G) \leq \max_{f \in Q} R(f).$$

Now our goal is to bound $R(f)$ for any $f \in Q$, which can be written as $f = \sum_{i=1}^{k} \alpha_i f_i$. For the denominator,

$$\sum_{v \in V} \deg_w(v) f(v)^2 = \sum_{i=1}^{k} \alpha_i^2 \sum_{v \in V} \deg_w(v) f_i(v)^2,$$

since at most one $f_i(v)$ can be nonzero for each $v \in V$. Before bounding the numerator, we define some new notations. Let $S_i := \operatorname{supp}(f_i)$ and for each $v \in V$ let

$$\iota(v) := \begin{cases} i, & \text{if } v \in S_i; \\ 0, & \text{if } v \notin S_1 \sqcup \cdots \sqcup S_k. \end{cases}$$

The map $\iota : V \to \{0\} \cup [k]$ is well-defined because the f_i's have disjoint support. Write $\alpha_0 = 0$ and $f_0 = \vec{0} \in \mathbb{R}^V$. For each edge $uv \in E$, if $\iota(u) = \iota(v)$ then

$$\begin{aligned} (f(u) - f(v))^2 &= (\alpha_{\iota(u)} f_{\iota(u)}(u) - \alpha_{\iota(v)} f_{\iota(v)}(v))^2 \\ &= \alpha_{\iota(u)}^2 (f_{\iota(u)}(u) - f_{\iota(u)}(v))^2 \leq \sum_{i \in [k]} \alpha_i^2 (f_i(u) - f_i(v))^2, \end{aligned}$$

whereas if $\iota(u) \neq \iota(v)$ then

$$\begin{aligned} (f(u) - f(v))^2 &= (\alpha_{\iota(u)} f_{\iota(u)}(u) - \alpha_{\iota(v)} f_{\iota(v)}(v))^2 \\ &\leq 2 \left[\alpha_{\iota(u)}^2 (f_{\iota(u)}(u) - f_{\iota(u)}(v))^2 + \alpha_{\iota(v)}^2 (f_{\iota(v)}(u) - f_{\iota(v)}(v))^2 \right] \\ &\leq 2 \sum_{i \in [k]} \alpha_i^2 (f_i(u) - f_i(v))^2 \end{aligned}$$

where we used $f_{\iota(u)}(v) = 0$ and $f_{\iota(v)}(u) = 0$. Applying this to all edges and using also the denominator result, we finally obtain

$$\begin{aligned} R(f) &= \frac{\sum_{uv\in E} w(uv)(f(u)-f(v))^2}{\sum_{v\in V} \deg_w(v) f(v)^2} \\ &\le \frac{2\sum_{i\in[k]} \left\{\alpha_i^2 \left[\sum_{uv\in E} w(uv)(f_i(u)-f_i(v))^2\right]\right\}}{\sum_{i\in[k]} \left\{\alpha_i^2 \left[\sum_{v\in V} \deg_w(v) f_i(v)^2\right]\right\}} \\ &\le 2\max_{i\in[k]} R(f_i). \end{aligned}$$

The proof is complete after taking maximum over all $f \in Q$. □

Finally, we present the variational characterization for $2-\lambda_n$.

Proposition 2.5.5. *For any graph $G=(V,E,w)$,*

$$2-\lambda_n(G) = \min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E} w(uv)(f(u)+f(v))^2}{\sum_{u\in V} \deg_w(u) f(u)^2}.$$

Proof. Observe that $2-\lambda_n(G)$ is the *smallest* eigenvalue of the matrix $I+A$, which has the useful quadratic form

$$x^T(I+A)x = f^T(D+A')f = \sum_{uv\in E} w(uv)(f(u)+f(v))^2$$

using the change of variables $f := D^{-1/2}x$. Therefore, using the Courant-Fischer theorem in Proposition 2.4.4 we have

$$2-\lambda_n(G) = \min_{x:V\to\mathbb{R}} \frac{x^T(I+A)x}{x^Tx} = \min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E} w(uv)(f(u)+f(v))^2}{\sum_{u\in V} \deg_w(u) f(u)^2},$$

as desired. □

2.5.4 Eigenvalues and Graph Properties

Using the variational characterization of $\lambda_k(G)$, we discover the following connection between eigenvalues and combinatorial properties of a graph.

Proposition 2.5.6. *For any graph $G=(V,E,w)$ with positive edge weights, $\lambda_k(G)=0$ iff G has at least k connected components. In particular, $\lambda_2(G)=0$ iff the graph G is disconnected.*

Proof. We first prove the result for $k = 2$. By Proposition 2.5.2, $\lambda_2(G) = 0$ iff there exists a nonzero $f \perp \deg_w$ such that $f(u) = f(v)$ for all $uv \in E$. If the graph is connected, it then follows that f is constant, and must be identically zero, so $\lambda_2(G) > 0$. If the graph is disconnected, say there exists a nonempty nontrival vertex subset $S \subset V$, then we can choose $a, b \in \mathbb{R}$ such that the vector $f(u) = a$ for $u \in S$ and $f(u) = b$ for $u \notin S$ satisfies $f \perp \deg_w$.

Now we generalize to arbitrary k. Let $S_1, \ldots, S_\ell$ be the connected components of G. We can write the Laplacian matrix $L(G)$ into the block diagonal form

$$L(G) = \begin{pmatrix} L(G[S_1]) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & L(G[S_\ell]) \end{pmatrix},$$

from which it follows that the nullity of $L(G)$ is ℓ, and $\lambda_k(G) = 0$ if and only if $\ell \geq k$. □

Let $\phi(G) := \min_{S \subseteq V : 0 < \pi(S) \leq 1/2} \phi(S)$ be the edge conductance of the graph G. Cheeger's inequality [Che70, AM85, Alo86] is a robust generalization of Proposition 2.5.6 that

$$\frac{\lambda_2(G)}{2} \leq \phi(G) \leq \sqrt{2\lambda_2(G)}.$$

This theorem is important because it connects (i) the spectral gap of the normalized Laplacian matrix, (ii) the edge conductance of the graph and (iii) the mixing time of random walks. We will defer the proof and further discussion to Section 3.1.1.

Another basic result in spectral graph theory relates $2 - \lambda_n(G)$ with the existence of bipartite components in G.

Proposition 2.5.7. *For any graph $G = (V, E, w)$ with positive edge weights, $\lambda_n(G) = 2$ if and only if G has a bipartite component S, or equivalently G has a set S of conductance zero with the induced subgraph $G[S]$ being bipartite.*

Proof. Suppose G has a bipartite component S. By Proposition 2.5.5 it suffices to find a nonzero $f : V \to \mathbb{R}$ such that

$$\sum_{uv \in E} w(uv)(f(u) + f(v))^2 = 0. \qquad (*)$$

If $S = A \sqcup B$ is the bipartitioning, we can set $f(u) = \mathbb{1}[u \in A] - \mathbb{1}[u \in B]$.

Conversely, if $\lambda_n(G) = 2$, that means there is a nonzero $f : V \to \mathbb{R}$ satisfying $(*)$ above. That means for any $uv \in E$ we have $f(u) + f(v) = 0$. Choose any $u_0 \in V$ with $f(u_0) \neq 0$ and let S be the connected component containing u_0. Then, S is necessarily bipartite with partitions $A := \{u \in S : f(u) = f(u_0)\}$ and $B := \{u \in S : f(u) = -f(u_0)\}$, and by definition S is of conductance zero. □

2.5.5 Isotropy

Given a graph $G = (V, E, w)$ and let $f_1, f_2, \ldots, f_k$ be the first k orthonormal eigenvectors of the normalized Laplacian $L(G)$. The spectral embedding of dimension k is defined as

$$f(v) := (f_1(v), f_2(v), \ldots, f_k(v)) \in \mathbb{R}^k$$

for each $v \in V$. The following proposition asserts that the sum of outer products of $f(v)$ satisfies a nice property called isotropy.

Proposition 2.5.8 (Isotropy of Spectral Embedding)**.** *Let $G = (V, E, w)$ be a graph and $f : V \to \mathbb{R}^k$ be the spectral embedding defined above. Then,*

$$\sum_{v \in V} f(v) f(v)^T = I_k.$$

Proof. For any vector $h \in \mathbb{R}^k$, we have by $\langle f_i, f_j \rangle = \mathbb{1}[i = j]$ that

$$h^T \left[\sum_{v \in V} f(v) f(v)^T \right] h = \sum_{v \in V} \sum_{i=1}^{k} \sum_{j=1}^{k} h_i h_j f_i(v) f_j(v) = \sum_{i=1}^{k} h_i^2 = h^T h.$$

This means the matrix in the bracket is the identity. □

2.5.6 Computing the Spectrum

Let $G = (V, E, w)$ be a graph and $L(G)$ be its Laplacian. We describe how to apply the power method to compute the second eigenvalue and an associated eigenvector. The following is based on [Tre16, Chapter 4], which we point the readers to for details.

One possible approach is to consider the matrix $M := 2I - L(G)$, which is PSD by Fact 2.5.1. We know that $\mathbb{1}_V$ is the largest eigenvector of M and therefore we can find (approximate) second largest eigenvalue of M and associated eigenvector using the strategy outlined in Section 2.4.2: namely, by projecting the starting random vecotr to the orthogonal complement of $\mathbb{1}_V$. The second largest eigenvalue of M equals $2 - \lambda_2(G)$, and so this gives an algorithm for computing the second smallest eigenvalue and eigenvector. As $\text{nnz}(M) = O(n+m)$, the runtime of this algorithm is $O(\varepsilon^{-1}(n+m)\log n)$, where $\varepsilon > 0$ is the desired multiplicative accuracy. This is near-linear when ε is a constant. However, due to the transformation $\lambda \mapsto 2 - \lambda$, if we wish the computed second smallest eigenvalue $\tilde{\lambda}_2(G)$ to satisfy e.g. $\tilde{\lambda}_2(G) \leq 2\lambda_2(G)$, we would need to take $\varepsilon = O(\lambda_2(G))$, which blows up the runtime when $\lambda_2(G) = o(1)$.

To remedy this, we consider instead the pseudo-inverse $L^\dagger$. The iterative steps of the power method requires computing matrix-vector products of the form $L^\dagger x$, which is the same as solving Laplacian systems. Using the fast Laplacian solvers of Spielman and Teng [ST14], the running time of the algorithm can be shown to be $\tilde{O}(\varepsilon^{-1}(n+m))$, but if we wish the computed second smallest eigenvalue $\tilde{\lambda}_2(G)$ to satisfy $\tilde{\lambda}_2(G) \leq 2\lambda_2(G)$ we only need to take $\varepsilon = O(1)$, since

$$\frac{1}{\tilde{\lambda}_2(G)} = \Theta\left(\frac{1}{\lambda_2(G)}\right) \implies \tilde{\lambda}_2(G) = \Theta(\lambda_2(G)).$$

2.6 Markov Chains and Random Walks

Given a finite set X called the state space, a Markov chain on X is represented by a matrix $P \in \mathbb{R}^{X \times X}$, where $P(u,v)$ is the probability of traversing from state u to state v in one step. Thus, P has nonnegative entries and satisfies $\sum_{v \in X} P(u,v) = 1$ for all $u \in X$. The matrix P^t then corresponds to taking t steps of the Markov chain.

A probability distribution $\pi : X \to \mathbb{R}_{\geq 0}$ is said to be a stationary distribution of P if $\pi^T P = \pi^T$.[5] A Markov chain with transition matrix P is said to be irreducible if for any state pair $u, v \in X$, $P^t(u,v) > 0$ for some $t \in \mathbb{N}$, and said to be aperiodic if, for all $u \in V$, $\gcd\{t \in \mathbb{N} : P^t(u,u) > 0\} = 1$. This is equivalent to the underlying directed graph (whose vertex set is X and arc set is $\{uv : P(u,v) > 0\}$) being strongly connected and aperiodic. It is well-known that (see e.g. Chapter 4 of [LP17]):

Proposition 2.6.1 ([LP17])**.** *If P is the transition matrix of an irreducible, aperiodic Markov chain, then it has a unique stationary distribution π, and furthermore $p_0^T P^t \to \pi^T$ as $t \to \infty$ for any initial distribution p_0.*

[5] We treat distributions as column vectors in $\mathbb{R}^X$, which is different from the standard convention.

For $\varepsilon \in (0,1)$, we define the ε-mixing time $\tau_{\text{mix}}(P,\varepsilon)$ of P to be the smallest $t \in \mathbb{N}$ such that $d_{TV}(\pi, (P^t)^T \rho) \leq \varepsilon$ for any initial distribution ρ. Here, $d_{TV}(\cdot,\cdot)$ is the total variation distance, defined as $d_{TV}(\rho_1,\rho_2) := \max_{S \subseteq X} |\rho_1(S) - \rho_2(S)|$ for any two distributions $\rho_1, \rho_2 : X \to \mathbb{R}_{\geq 0}$. This describes how fast the Markov chain converges to its stationary distribution.

A transition matrix P is said to be time-reversible with respect to π if $\pi(u)P(u,v) = \pi(v)P(v,u)$ for any $u, v \in X$. Note that this implies that π is a stationary distribution of P. The time reversibility condition can be written as $\Pi P = P^T \Pi$, where $\Pi := \text{diag}(\pi)$. Thus, $\Pi^{1/2} P \Pi^{-1/2}$ is symmetric, hence diagonalizable with eigenvalues $1 = \alpha_1(P) \geq \alpha_2(P) \geq \cdots \geq \alpha_n(P) \geq -1$. As P is similar to $\Pi^{1/2} P \Pi^{-1/2}$, they have the same eigenvalues by Fact 2.4.1. The spectral gap of P is defined as $1 - \alpha_2(P)$. The relaxation time $\tau_{\text{rel}}(P)$ of P is defined as the reciprocal of the spectral gap, so $\tau_{\text{rel}}(P) := \frac{1}{1-\alpha_2(P)}$.

Suppose P' is the transition matrix of an irreducible Markov chain. We consider the lazy version defined as $P = (I + P')/2$. One motivation is to get rid of periodicity in P', so that P always has a unique stationary distribution π. Let $\pi_{\min} := \min_{u \in V} \pi(u)$. The following proposition relates the relaxation time with the mixing time of an irreducible, reversible chain (see e.g. [LP17, Chapter 12]). We provide a proof for reference, as we will use similar ideas to derive some other mixing time bounds in the later chapters.

Proposition 2.6.2 (Mixing Time Bound using Relaxation Time for Reversible Chains [LP17]). *Let $P = (I + P')/2$ be the lazy version of an irreducible Markov chain with transition matrix P', and π be the stationary distribution of P. Suppose further that P is time-reversible with respect to π. Then,*

$$\big(\tau_{\text{rel}}(P) - 1\big) \cdot \log \frac{1}{2\varepsilon} \leq \tau_{\text{mix}}(P, \varepsilon) \leq \tau_{\text{rel}}(P) \cdot \log \frac{1}{\varepsilon \cdot \pi_{\min}}. \tag{2.3}$$

Proof. For the mixing time upper bound, notice that the eigenvalues of P satisfy $1 = \alpha_1(P) > \cdots \geq \alpha_n(P) \geq 0$. Since $P \sim \Pi^{1/2} P \Pi^{-1/2} =: \tilde{P}$ and the latter is symmetric by reversibility, $f_i^T := \tilde{f}_i^T \Pi^{1/2}$ is a left eigenbasis of P, where $\tilde{f}_i$ is an orthonormal eigenbasis for $\tilde{P}$, with $\tilde{f}_i^T \tilde{P} = \alpha_i \tilde{f}_i^T$ (we write α_i for $\alpha_i(P)$). Since $\tilde{f}_i$ are orthonormal, we have

$$\langle f_i, f_j \rangle_{1/\pi} = \sum_{x \in X} \frac{1}{\pi(x)} f_i(x) f_j(x) = \langle \tilde{f}_i, \tilde{f}_j \rangle = \mathbb{1}[i = j].$$

For any starting distribution p_0, there exist $a_i \in \mathbb{R}$ such that $p_0 = \sum_{i=1}^{|X|} a_i f_i$, and $f_1 \propto \pi$ since $\pi^T P = \pi^T$ is a left eigenvector of P with eigenvalue 1. Then,

$$p_0^T P^t = \sum_{i=1}^{|X|} a_i f_i^T P^t = a_1 f_1^T + \sum_{i=2}^{|X|} a_i \alpha_i^t f_i^T.$$

We know that $a_1 f_1 = \pi$ since

$$1 = p_0^T P^t \mathbb{1} = a_1 f_1^T \mathbb{1} + \sum_{i=2}^{|X|} a_i \alpha_i^t f_i^T \mathbb{1} \to a_1 f_1^T \mathbb{1}$$

as $t \to \infty$, and $\pi^T \mathbb{1} = 1$. Therefore,

$$\begin{aligned} d_{TV}(\pi, (P^t)^T p_0) &= \frac{1}{2} \left\| p_0^T P^t - \pi^T \right\|_1 \\ &= \frac{1}{2} \left\| \sum_{i=2}^{|X|} a_i \alpha_i^t f_i^T \right\|_1 \\ &\overset{(*)}{\leq} \frac{1}{2} \left\| \sum_{i=2}^{|X|} a_i \alpha_i^t f_i^T \right\|_{1/\pi} \\ &= \frac{1}{2} \sqrt{\sum_{i=2}^{|X|} a_i^2 \alpha_i^{2t}} \\ &\leq \frac{1}{2} (1 - (1 - \alpha_2))^t \cdot \sqrt{\sum_{i=2}^{|X|} a_i^2} \end{aligned}$$

where the step $(*)$ is by Cauchy-Schwarz inequality (see Section 2.10) that

$$\|b\|_1 = \sum_{x \in X} |b(x)| \leq \sqrt{\sum_{x \in X} \frac{1}{\pi(x)} b(x)^2 \cdot \sum_{x \in X} \pi(x)} = \|b\|_{1/\pi}.$$

We can upper-bound $\sum_{i=2}^{|X|} a_i^2$ by

$$\sum_{i=2}^{|X|} a_i^2 \leq \sum_{i=1}^{|X|} a_i^2 = \langle p_0, p_0 \rangle_{1/\pi} \leq \frac{1}{\pi_{\min}} \langle p_0, p_0 \rangle \leq \frac{1}{\pi_{\min}} \|p_0\|_1^2 = \frac{1}{\pi_{\min}}.$$

Plugging this back to the bound on d_{TV}, we have

$$\begin{aligned} d_{TV}(\pi, (P^t)^T p_0) &\leq \frac{1}{2} (1 - (1 - \alpha_2))^t \cdot \sqrt{\frac{1}{\pi_{\min}}} \\ &\leq \frac{1}{2} e^{-(1-\alpha_2)t} \cdot \sqrt{\frac{1}{\pi_{\min}}}, \end{aligned}$$

which is $\le \varepsilon$ if we take $t \ge (1-\alpha_2)^{-1} \log(1/(2\varepsilon\sqrt{\pi_{\min}}))$. This gives the upper bound.

For the mixing time lower bound, we consider right eigenvectors of P. We know that $g_i := \Pi^{-1/2} \tilde{f}_i^T$ is a right eigenbasis of P with $Pg_i = \alpha_i g_i$. We also know that $g_1 \propto \mathbb{1}$ from $P\mathbb{1} = \mathbb{1}$. Consider the second eigenvector g_2, and let $x \in X$ be the entry such that $|g_2(x)|$ is maximized. Note that $g_2 \perp \pi$ since

$$\langle g_2, \pi \rangle = \langle g_2, g_1 \rangle_\pi = \langle \tilde{f}_2, \tilde{f}_1 \rangle = 0.$$

Then,

$$(1-\alpha_2)^t |g_2(x)| = |P^t g_2(x)| = \left| \sum_{y \in X} \left[P^t(x,y) g_2(y) - \pi(y) g_2(y) \right] \right| \le 2 \left\| g_2 \right\|_\infty d_{TV}(\mathbb{1}_x^T P^t, \pi^T),$$

where $\mathbb{1}_x$ is the distribution concentrated at x. Therefore,

$$(1-(1-\alpha_2))^t \le 2 \sup_{p_0} d_{TV}(p_0^T P^t, \pi^T).$$

When $t \le (\tau_{\text{rel}} - 1)\log(1/2\varepsilon)$, using $(1-1/x)^{x-1} \ge 1/e$ for $x > 0$, we have

$$(1-(1-\alpha_2))^t \ge e^{-\log(1/2\varepsilon)} = 2\varepsilon$$

and so $\tau_{\text{mix}}(P) \ge (\tau_{\text{rel}} - 1)\log(1/2\varepsilon)$, as required. □

2.6.1 Random Walks on Graphs

Given an undirected or a directed graph $G = (V, E, w)$, with positive edge weights, the transition matrix $P' \in \mathbb{R}^{V \times V}$ of the canonical random walk on G is defined as $P'(u,v) = w(uv) / \sum_{x \in V} w(ux)$, where $w(ux) = 0$ if $ux \notin E$. If G is strongly connected (but not necessarily aperiodic), the lazy random walk $P := (I + P')/2$ nevertheless admits a unique stationary distribution π, and $p_0^T P^t \to \pi^T$ for any initial distribution p_0 as $t \to \infty$. The mixing time measures how fast $p_0^T P^t$ converges to π^T. We use $\tau_{\text{mix}}(P)$ to denote the $(1/e)$-mixing time of P (to π), and we write "mixing time" instead of "$(1/e)$-mixing time".

When the graph G is undirected, we have $P' = D^{-1}A'$ where A' is the adjacency matrix and D is the diagonal degree matrix defined in Section 2.5. We see that P' is similar to $A = D^{-1/2}A'D^{-1/2}$, and so they have the same eigenvalues. Moreover,

$$1 - \alpha_2(P') = 1 - \alpha_2(G) = \lambda_2(G) \quad \text{and} \quad 1 - \alpha_2(P) = 1 - \left(\frac{1 + \alpha_2(P')}{2} \right) = \frac{\lambda_2(G)}{2}.$$

Finally, when G is connected, $\pi \propto \deg_w$ is the unique stationary distribution of P. Therefore, specializing (2.3) to the undirected graph setting we have

$$\frac{1}{\lambda_2(G)} \lesssim \tau_{\text{mix}}(P) \lesssim \frac{1}{\lambda_2(G)} \cdot \log \frac{1}{\pi_{\min}}. \tag{2.4}$$

If the graph G is unweighted, then $\pi(v) = \deg(v)/\text{vol}(V)$, so $\log(1/\pi_{\min}) \leq \log(2m) \lesssim \log n$, and the above yields

$$\frac{1}{\lambda_2(G)} \lesssim \tau_{\text{mix}}(P) \lesssim \frac{1}{\lambda_2(G)} \cdot \log n.$$

2.7 Convex Geometry

(In this section, n does not carry the meaning of the number of vertices of a graph.)

A polytope $\emptyset \neq Q \subseteq \mathbb{R}^n$ is a solution set to affine inequalities, i.e. $Q = \{x \in \mathbb{R}^n : Ax \leq b\}$ for some $A \in \mathbb{R}^{d\times n}$ and $b \in \mathbb{R}^d$. Given a point set $X \subseteq \mathbb{R}^n$, its convex hull $\text{conv}(X) \subset \mathbb{R}^n$ is the set of all convex combinations of points in X. Equivalently, $\text{conv}(X)$ is the smallest convex set containing X. A basic result is that the convex hull of a finite point set in $\mathbb{R}^n$ is a polytope.

Given a polytope $Q \subseteq \mathbb{R}^n$. A subset $F \subseteq Q$ is a face of Q if, for any $x, y \in Q$, if $tx + (1-t)y \in F$ for some $t \in (0,1)$ then $x, y \in F$. If F is a face of Q, then it is the intersection of Q and an affine subspace $Y \subseteq \mathbb{R}^n$. The dimension of a face F is the dimension of the smallest (by inclusion) affine subspace $Y \subseteq \mathbb{R}^n$ such that $F = Q \cap Y$.

Dimension-0 faces are called extreme points or vertices, whereas dimension-1 faces are called edges. The graph $G_Q = (V, E)$ of Q has the dimension-0 faces as the vertices and the dimension-1 faces as the edges. This is also called the 1-skeleton of the polytope. The following fact about the non-existence of edges between two vertices of a polytope will be useful.

Proposition 2.7.1 ([KR03]). *Let $Q \subseteq \mathbb{R}^n$ be a bounded polytope with vertex set $F_0 = F_0(Q)$. Let $x \neq y \in F_0$, and let $L(x,y)$ be the line segment with endpoints x and y. Suppose that $\text{conv}\big(F_0 \setminus \{x,y\}\big) \cap L(x,y)$ is nonempty. Then, $L(x,y)$ is not an edge of Q.*

The following version of hyperplane separation theorem (c.f. [BV04, Section 2.5]) will be useful.

Proposition 2.7.2 ([BV04]). *Let $M \subseteq \mathbb{R}^n$ be compact and convex, and let $x \in \mathbb{R}^n$ be such that $x \notin M$. Then there exists an affine function $l : \mathbb{R}^n \to \mathbb{R}$ such that $l(x) = 0$ and $l(y) < 0$ for all $y \in M$.*

2.8 Convex Optimization

2.8.1 Optimization Programs

Consider optimization programs in the standard form

$$\begin{array}{rll} \mathcal{P} := & \min_{x\in\Omega} & f(x) \\ & \text{subject to} & g_i(x) \leq 0 \quad \forall i \in [l] \\ & & h_j(x) = 0 \quad \forall i \in [p] \end{array}$$

where $\Omega \subseteq \mathbb{R}^d$ and $f, g_i, h_j : \mathbb{R}^d \to \mathbb{R} \cup \{\pm\infty\}$. To define its Lagrangian dual, consider

$$\Lambda(x, \lambda, \mu) := f(x) + \sum_{i\in[l]} \lambda_i g_i(x) + \sum_{j\in[p]} \mu_j h_j(x)$$

defined on $\Omega \times \mathbb{R}^l \times \mathbb{R}^p$. The dual program is

$$\mathcal{D} := \max_{\lambda\geq 0,\mu} \inf_{x\in\Omega} \Lambda(x, \lambda, \mu).$$

Weak duality always holds, that is, $\mathcal{D} \leq \mathcal{P}$.

Linear programs (LP) are optimization programs where $\Omega = \mathbb{R}^d$ and f, g_i, h_j are all affine functions. Semidefinite programs (SDP) are optimization programs where the ambient space is $\mathbb{R}^{d\times d}$, $\Omega := \{X \in \mathbb{R}^{d\times d} : X \succeq 0\}$, and f, g_i, h_j are all affine functions. These are the two main classes of convex programs that we use in this thesis.

2.8.2 Computational Aspects

Some remarks about the time complexity of solving an SDP are in order. See e.g. [GM12, WSV12, LR05] for reference.

Let R be an upper bound on the maximum Frobenius norm of any feasible solution X to the SDP. We make the additional assumption that R is at most polynomial in the size of the program, which is satisfied by all SDP's we will encounter in this thesis.

The two most common choices for SDP solvers are the ellipsoid method and the interior-point method. The ellipsoid method produces a solution with additive error ε in time polynomial in the input size and $\log(R/\varepsilon)$, under the bit number model of computation, although in practice the runtime can be slow. The interior-point method tends to be significantly more efficient in practice, but a rigorous runtime bound requires further assumptions on the structure of the SDP, e.g. the existence of both primal and dual Slater points, and only under the real number model of computation.

We simply note that all the SDP's introduced in this thesis can be (approximately) solved in polynomial time, in bit complexity.

2.8.3 Strong Duality

We say that strong duality holds if $\mathcal{P} = \mathcal{D}$. Note that it is not required that the common value is attained by a primal-dual solution pair. It is well known that strong duality always holds for linear programs. Unlike linear programs, there are SDP's where strong duality does not hold. In a few places in the thesis, we shall use von Neumann's minimax theorem to establish strong duality.

Theorem 2.8.1 (Von Neumann's Minimax Theorem (see [Sim95])). *Let X, Y be nonempty compact convex sets. If f is a real-valued continuous function on $X \times Y$ with $f(x, \cdot)$ concave on Y for all $x \in X$ and $f(\cdot, y)$ convex on X for all $y \in Y$, then*

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

Another sufficient condition for strong duality to hold is the existence of a Slater point, i.e. a "strictly feasible" solution. Formally:

Theorem 2.8.2 (Slater's Condition (see [BV04, Section 5.2.3])). *Suppose that the convex program $\mathcal{P}$ admits a feasible solution $x \in \mathrm{relint}(\Omega)$, $g_i(x) < 0$ for all nonlinear inequality constraints. Then, strong duality holds that $\mathcal{P} = \mathcal{D}$, and if $\mathcal{D}$ is feasible then a dual optimum solution exists.*

When applied to semidefinite programs, Slater's condition amounts to establishing a *positive definite* feasible solution $X \succ 0$, since all the constraints are affine.

2.8.4 Sum of Eigenvalues

Several eigenvalue optimization problems can be formulated as semidefinite programs; see Section 3.2.2. In Chapter 4 and Chapter 5, we will use the following proposition in writing the maximum reweighted sum of k smallest eigenvalue problem as a semidefinite program, which follows from e.g. [HJ13, Corollary 4.3.39].

Proposition 2.8.3 (Sum of k Smallest Eigenvalues [HJ13]). *Let $X \in \mathbb{R}^{d\times d}$ be a symmetric matrix and let $1 \leq k \leq d$. Suppose the eigenvalues of X are $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$. Then, $\lambda_1 + \lambda_2 + \cdots + \lambda_k$ is the value of the following semidefinite program:*

$$\begin{aligned} \min_{Y\in\mathbb{R}^{d\times d}} \quad & \operatorname{tr}(XY) \\ \text{subject to} \quad & 0 \preceq Y \preceq I_d \\ & \operatorname{tr}(Y) = k. \end{aligned}$$

2.8.5 Vector Program for λ_2

We will use the following exact SDP formulation of the second eigenvalue in (2.2).

Proposition 2.8.4 (Vector Program for λ_2). *For any graph $G = (V, E, w)$,*

$$\lambda_2(G) = \min_{\substack{f:V\to\mathbb{R}^n \\ \sum_{v\in V} \deg_w(v) f(v) = \vec{0}}} \frac{\sum_{uv\in E} w(uv)\cdot \|f(u) - f(v)\|^2}{\sum_{v\in V} \deg_w(v)\cdot \|f(v)\|^2}. \tag{2.5}$$

Proof. Let $f_i : V \to \mathbb{R}$ be the i-th coordinate function of f. Then $\deg_w \perp f_i$ for each $i \in [n]$. The minimization objective can be rewritten as

$$\frac{\sum_{i\in[n]}\sum_{uv\in E} w(uv)\cdot (f_i(u) - f_i(v))^2}{\sum_{i\in[n]}\sum_{v\in V} \deg_w(v)\cdot f_i(v)^2}.$$

By Proposition 2.5.2, $\sum_{uv\in E} w(uv)\cdot(f_i(u) - f_i(v))^2 \geq \lambda_2(G)\cdot \sum_{v\in V}\deg_w(v)\cdot f_i(v)^2$ for every $i \in [n]$, and so RHS is at least $\lambda_2(G)$. Equality can be established by taking f_i to be the second eigenvector for all $i \in [n]$. □

To see why (2.5) is an SDP, consider the Gram matrix $X \in \mathbb{R}^{V\times V}$ with $X(u, v) = \langle f(u), f(v)\rangle$ and rewrite the program as follows:

$$\begin{aligned} \min_{X\succeq 0} \quad & \sum_{uv\in E} w(uv)(X(u,u) + X(v,v) - X(u,v) - X(v,u)) \\ \text{subject to} \quad & \langle D, X\rangle = 1 \\ & \langle \deg_w \deg_w^T, X\rangle = 0 \end{aligned}$$

where $D = \operatorname{diag}(\deg_w)$. Then, the objective and the constraints are all linear in X. This also explains why it is needed to lift $f : V \to \mathbb{R}$ in the original Rayleigh quotient definition of $\lambda_2(G)$ to $f : V \to \mathbb{R}^n$, as the one-dimensional embedding will impose a rank-1 constraint on X, which makes the program non-convex.

2.9 Flows

Given a directed or undirected graph $G = (V, E)$. If the graph is undirected, consider its bidirection, which is the directed graph with arcs (u, v) and (v, u) for every edge $uv \in E$.

Single-Commodity Flows: A single-commodity flow problem is the problem of sending flows from s to t subject to certain constraints. The typical setting is that there is a capacity constraint on each arc restricting the amount of flow that may pass through it, and the goal is to maximize the total amount of flow sent from s to t. Formally,

Definition 2.9.1 (Maximum Flow)**.** *Given a directed graph $G = (V, E)$ with capacity $C(e) \geq 0$ on arc $e \in E$. Given vertices $s \neq t$. The maximum s-t flow problem is the following program:*

$$\begin{aligned} \max_{f: E \to \mathbb{R}_{\geq 0}} \quad & M \\ \text{subject to} \quad & \sum_{v: vu \in E} f(vu) - \sum_{v: uv \in E} f(uv) = M(\mathbb{1}[u = s] - \mathbb{1}[u = t]) \qquad \forall u \in V \\ & f(e) \leq C(e) \qquad \forall e \in E. \end{aligned}$$

We also refer to G as a flow network.

The first constraint says that the net amount of flow out of vertex s should be M, the net amount of flow into vertex t should be M, and there should be a net zero flow in or out of every other vertex. This means that M is the amount of flow sent from s and t and so the program seeks to maximize M. The second constraint says that the amount of flow through an arc should be at most its capacity.

Given a feasible s-t flow f, it is possible to decompose f into a sum of path flows f_p, so that f_p is an s-t flow along a path p from s to t.

The minimum s-t cut is the problem of removing arcs so that there is no directed path from s to t in the resulting graph, and the total capacity of the removed arcs is minimized. The max-flow min-cut duality states that the maximum s-t flow is equal to the minimum s-t cut.

Multicommodity Flows: A multi-commodity flow is the problem of sending flows from multiple source vertices to their corresponding target vertices. These demands may be represented using a matrix $D \in \mathbb{R}_{\geq 0}^{V \times V}$, so that $D(u, v) \geq 0$ is the required amount of flow to send from u to v. Instead of specifying the amount of flow on each arc, we need to use one variable per path. Let $\mathcal{P}$ be the set of paths on G and $\mathcal{P}(u,v)$ be the set of u-v paths on G. A multicommodity flow $F : \mathcal{P} \to \mathbb{R}_{\geq 0}$ satisfies the demand $D \in \mathbb{R}_{\geq 0}^{V \times V}$ if

$$\sum_{p \in \mathcal{P}(u,v)} F(p) = D(u,v)$$

for all $u, v \in V$. The goal is usually to find a feasible multicommodity flow that minimizes some form of congestion, which detects if a large amount of flow passes through particular vertices or arcs. This is equivalent to imposing capacity constraints on the vertices or the arcs and asking for the largest scalar $\alpha \geq 0$, such that it is possible to satisfy the demand αD while respecting the capacity constraints.

2.10 Inequalities

2.10.1 Elementary Inequalities

The following is the most frequently used inequality in this thesis.

Fact 2.10.1 (Cauchy-Schwarz Inequality)**.** *For real numbers $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$,*

$$\left(\sum_{i=1}^{k} a_i^2\right)\left(\sum_{i=1}^{k} b_i^2\right) \geq \left(\sum_{i=1}^{k} a_i b_i\right)^2.$$

Corollary 2.10.2 (Quadratic Mean Inequality)**.** *For real numbers $a_1, \ldots, a_k$,*

$$\frac{a_1^2 + \cdots + a_k^2}{k} \geq \left(\frac{a_1 + \cdots + a_k}{k}\right)^2.$$

Proof. Apply Cauchy-Schwarz inequality with $b_i = 1/k$. □

We also collect some useful inequalities here.

Fact 2.10.3. *Let $f : D \to \mathbb{R}$ and $\pi : D \to \mathbb{R}_{\geq 0}$ be two functions. If $\sum_{x \in D} \pi(x) f(x) = 0$, then for any $c \in \mathbb{R}$,*

$$\sum_{x \in D} \pi(x)(f(x) - c)^2 \geq \sum_{x \in D} \pi(x) f(x)^2.$$

Proof. Differentiate LHS with respect to c to see that LHS is minimized when $c = 0$. □

Fact 2.10.4. *Let $f : D \to \mathbb{R}^k$ be such that $\sum_{u \in D} \pi(u) f(u) = \vec{0}$ for some positively-valued function $\pi : D \to \mathbb{R}^+$. Then,*

$$\sum_{u \in D} \pi(u) \|f(u)\|^2 = \frac{1}{2\pi(D)} \sum_{u,v \in D} \pi(u)\pi(v) \|f(u) - f(v)\|^2 .$$

Proof. Expanding RHS we have

$$\begin{aligned}
& \frac{1}{2|D|} \sum_{u,v \in D} \pi(u)\pi(v) \|f(u) - f(v)\|^2 \\
= \; & \frac{1}{2\pi(D)} \sum_{u,v \in D} \pi(u)\pi(v) \left[\|f(u)\|^2 + \|f(v)\|^2 - 2\langle f(u), f(v) \rangle \right] \\
= \; & \frac{1}{2\pi(D)} \left[\pi(D) \cdot \sum_{u \in D} \pi(u) \|f(u)\|^2 + \pi(D) \cdot \sum_{v \in D} \pi(v) \|f(v)\|^2 - 2 \left\| \sum_{u \in D} \pi(u) f(u) \right\|^2 \right] \\
= \; & \sum_{u \in D} \pi(u) \|f(u)\|^2 ,
\end{aligned}$$

which is exactly LHS. □

2.10.2 Gaussians

We collect some useful inequalities about Gaussians here.

Proposition 2.10.5 (Expected Maximum of χ-Squared Distribution)**.** *Let (Γ_{ij}) for $1 \leq i \leq d$ and $1 \leq j \leq T$ be Gaussian random variables with mean 0 and variance at most 1, and such that $\Gamma_{i1}, \Gamma_{i2}, \ldots, \Gamma_{iT}$ are mutually independent for each $i \in [d]$. Let $Y_i := \frac{1}{T} \sum_{1 \leq j \leq T} \Gamma_{ij}^2$ and let $Y := \max_{1 \leq i \leq d} Y_i$. Then,*

$$\mathbb{E}[Y] \leq 4 \left(1 + \frac{1 + \log d}{T} \right).$$

Proof. By the Laurent-Massart bound of χ-squared distribution [LM00, Lemma 1], for any $\delta > 0$ and $i \leq d$,

$$\Pr\left[Y_i - 1 \geq 2\sqrt{\delta/T} + 2(\delta/T) \right] \leq e^{-\delta}.$$

Since

$$1 + 2\sqrt{\delta/T} + 2(\delta/T) \leq 4 + 4(\delta/T)$$

it follows that

$$\Pr[Y_i \geq 4 + 4(\delta/T)] \leq e^{-\delta}.$$

Recall that $Y = \max_{i \leq d} Y_i$. Taking union bound over the Y_i's,

$$\Pr[Y \geq 4 + 4(\delta/T)] \leq d \cdot e^{-\delta}.$$

With this probability tail bound, the expectation of Y can be bounded as follows:

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^\infty \Pr[Y \geq t]\, dt \;\leq\; 4 + \int_0^\infty \Pr[Y \geq 4 + t]\, dt \;\leq\; 4 + \int_0^\infty \min\left\{1, d \cdot e^{-T \cdot t/4}\right\} dt \\
&= 4 + \frac{4 \log d}{T} + \int_{\frac{4\log d}{T}}^\infty d \cdot e^{-T \cdot t/4}\, dt \\
&= 4 + \frac{4 \log d}{T} + \frac{4d}{T} \cdot e^{-T \cdot t/4}\Big|_{\frac{4\log d}{T}}^\infty \;=\; 4 + \frac{4 \log d}{T} + \frac{4}{T}.
\end{aligned}$$

□

Fact 2.10.6 ([LRV13, Fact B.3])**.** *Let $Y_1, Y_2, \ldots, Y_d$ be d Gaussian random variables with mean 0 and variance at most σ^2. Let Y be the random variable defined as $Y := \max\{Y_i \mid i \in [d]\}$. Then*

$$\mathbb{E}[Y^2] \leq 4\sigma^2 \log d.$$

Fact 2.10.7 ([LRV13, Lemma 9.8])**.** *Suppose $Y_1, \ldots, Y_d$ are Gaussian random variables such that $\mathbb{E}[\sum_{i=1}^d Y_i^2] = 1$. Then*

$$\Pr\left[\sum_{i=1}^d Y_i^2 \geq \frac{1}{2}\right] \geq \frac{1}{12}.$$

Lemma 2.10.8 (Johnson-Lindenstrauss Lemma [JLS86])**.** *Let $f : D \to \mathbb{R}^d$ be a finite collection of points, and let $\varepsilon \in (0,1)$. Then, for $\ell = \Omega(\varepsilon^{-2} \log |D|)$ there exists a linear map $A : \mathbb{R}^d \to \mathbb{R}^\ell$ such that for all $u, v \in D$,*

$$(1-\varepsilon)\left\|f(u) - f(v)\right\|^2 \leq \left\|A \cdot f(u) - A \cdot f(v)\right\|^2 \leq (1+\varepsilon)\left\|f(u) - f(v)\right\|^2.$$

Such map may be computed in randomized polynomial time.

Chapter 3

Literature Review

In this chapter, we review important results in the literature that laid the foundation for our new work. The chapter is structured as follows:

- In Section 3.1, we review the classical Cheeger's inequality as well as several important generalizations to higher eigenvalues. This is relevant to both Chapter 4 and Chapter 5.
- In Section 3.2, we review past work towards a spectral graph theory for vertex expansion in graphs. This is most relevant to Chapter 4, except the review of [JPV22] in Section 3.2.3 which is relevant to Chapter 5.
- In Section 3.3, we review past attempts at developing a spectral graph theory for directed graphs. This is most relevant to Chapter 5.
- In Section 3.4, we review past attempts at developing a spectral graph theory for hypergraphs. This is relevant to both Chapter 4 (via reduction) and Chapter 5.
- In Section 3.5, we review past work on upper bounding $\lambda_2(G)$ and $\lambda_k(G)$ of a graph G enjoying special structural properties. This is most relevant to Chapter 7.
- In Section 3.6, we review past work on $O(\sqrt{\log n})$ approximation of expansion quantities using the techniques of [ARV09], as well as approximation algorithms of small-set and multi-way expansion quantities using orthogonal separators [CMM06, BFK+14, LM14a, LM14b]. This is most relevant to Chapter 8.

3.1 Cheeger's Inequality and Generalizations

For simplicity, we state and prove the results in this section for unweighted graphs. All of them can be shown to hold for weighted graphs as well.

3.1.1 Classical Cheeger's Inequality

The classical Cheeger's inequality is named after Cheeger [Che70], who devised the continuous version of the inequality. The discrete version for graphs was proved by Alon and Milman [AM85], Alon [Alo86], and Sinclair and Jerrum [SJ89].

Theorem 3.1.1 (Cheeger's Inequality [Che70, AM85, Alo86, SJ89]). *Let $G = (V, E)$ be a graph. Then,*

$$\frac{\lambda_2(G)}{2} \leq \phi(G) \leq \sqrt{2\lambda_2(G)}.$$

The proof of Cheeger's Inequality below will serve as a template for the proof of all other Cheeger-type inequalities to follow. We call the inequality which uses eigenvalues to lower-bound expansion quantities the "easy" direction, as the proof is typically easier, by showing that the minimization problem from the eigenvalues is a relaxation of the minimization problem from the expansion quantities. We call the inequality which uses eigenvalues to upper-bound expansion quantities the "hard" direction, as the proof is typically harder: given a vector solution to the minimization problem from the eigenvalues, round it to a set with small expansion.

Proof. For the "easy direction", let $S \subseteq V$ such that $0 < \text{vol}(S) \leq \text{vol}(V)/2$ and $\phi(S) = \phi(G)$. Construct the vector $f : V \to \mathbb{R}$ defined as

$$f(u) := \begin{cases} \frac{-1}{\text{vol}(S)}, & \text{if } u \in S \\ \frac{1}{\text{vol}(S^c)}, & \text{if } u \in S^c. \end{cases}$$

Check that $\sum_u \deg(u) f(u) = 0$, and so

$$\begin{aligned} \lambda_2(G) &\leq \frac{\sum_{uv \in E}(f(u) - f(v))^2}{\sum_{v \in V} \deg(v) f(v)^2} \qquad \text{(by Proposition 2.5.2)} \\ &= \frac{|E(S, S^c)| \cdot \left(\frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(S^c)}\right)^2}{\frac{1}{\text{vol}(S)} + \frac{1}{\text{vol}(S^c)}} \\ &\leq 2\frac{|E(S, S^c)|}{\text{vol}(S)} = 2\phi(S). \end{aligned}$$

For the "hard direction", the proof is algorithmic and rounds any vector $f : V \to \mathbb{R}$ satisfying $\sum_u \deg(u) f(u) = 0$ to a vertex subset $S \subseteq V$ satisfying $0 < \mathrm{vol}(S) \le \mathrm{vol}(V)/2$ and

$$\phi(S) \le \sqrt{2 \cdot \frac{\sum_{uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} \deg(v) f(v)^2}}.$$

We will present the proof slightly differently than the most common treatise, to decouple the two key steps: the first step is to go from ℓ_2^2 to ℓ_1, and the second step is to go from ℓ_1 to cuts. The idea of considering ℓ_1 quantities appeared in the work of Trevisan [Tre09] (see Section 3.1.2 to follow) as well as later in [CLTZ18]. We will use the same two-step approach for the proofs of the new Cheeger-type inequalities, to give cleaner proofs and, in some situations, out of necessity.

Step 1 (ℓ_2^2 to ℓ_1). Given $f : V \to \mathbb{R}$, our goal is to construct $h : V \to \mathbb{R}$ such that

$$\frac{\sum_{uv \in E} |h(u) - h(v)|}{\sum_{v \in V} \deg(v) |h(v)|} \le \sqrt{2 \cdot \frac{\sum_{uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} \deg(v) f(v)^2}}.$$

This explains the name for this step – $f : V \to \mathbb{R}$ is ℓ_2^2 because the objective in the numerator is the sum of squared ℓ_2 distances across the edges in the embedding $u \mapsto f(u) \in \mathbb{R}$, and $h : V \to \mathbb{R}$ is ℓ_1 because the objective in the numerator is the sum of ℓ_1 distances across the edges in the embedding $u \mapsto h(u) \in \mathbb{R}$. The same applies to the denominators. h needs to be balanced but not in the sense that $\sum_{v \in V} \deg(v) h(v) = 0$; rather, our construction ensures that 0 is a degree-weighted median of h, so that the second step yields a cut with volume at most $\mathrm{vol}(V)/2$.

Let $c \in \mathbb{R}$ be a degree-weighted median of f; that is, $\mathrm{vol}(\{v \in V : f(v) > c\}) \le \mathrm{vol}(V)/2$ and $\mathrm{vol}(\{v \in V : f(v) < c\}) \le \mathrm{vol}(V)/2$. Define $h : V \to \mathbb{R}$ so that

$$h(u) := \begin{cases} (f(u) - c)^2 & \text{if } f(u) \ge c \\ -(f(u) - c)^2 & \text{if } f(u) < c. \end{cases} \tag{3.1}$$

In the denominator,

$$\sum_{v \in V} \deg(v) |h(v)| = \sum_{v \in V} \deg(v) (f(v) - c)^2 \ge \sum_{v \in V} \deg(v) f(v)^2.$$

where the inequality is due to $\sum_{v \in V} \deg(v) f(v) = 0$ and using Fact 2.10.3. In the numerator, we show that for any $uv \in E$ (indeed for any $u, v \in V$),

$$|h(u) - h(v)| \le |f(u) - f(v)| \left(|f(u) - c| + |f(v) - c| \right). \tag{3.2}$$

If $h(u)$ and $h(v)$ are both positive or both negative, then

$$\begin{aligned} |h(u) - h(v)| &= |(f(u) - c)^2 - (f(v) - c)^2| \\ &= |f(u) - f(v)| \left(|f(u) - c| + |f(v) - c|\right). \end{aligned}$$

Otherwise, $(f(u) - c)(f(v) - c) \leq 0$, and so

$$\begin{aligned} |h(u) - h(v)| &= |(f(u) - c)^2 + (f(v) - c)^2| \\ &\leq (f(u) - c)^2 - 2(f(u) - c)(f(v) - c) + (f(v) - c)^2 \\ &= |f(u) - f(v)| \cdot |(f(u) - c) - (f(v) - c)| \\ &\leq |f(u) - f(v)| \left(|f(u) - c| + |f(v) - c|\right). \end{aligned}$$

Therefore, we can upper bound the numerator as follows:

$$\begin{aligned} \sum_{uv \in E} |h(u) - h(v)| &\leq \sum_{uv \in E} |f(u) - f(v)| \left(|f(u) - c| + |f(v) - c|\right) \\ &\leq \sqrt{\sum_{uv \in E} (f(u) - f(v))^2 \cdot 2 \sum_{uv \in E} \left(|f(u) - c|^2 + |f(v) - c|^2\right)} \\ &= \sqrt{\sum_{uv \in E} (f(u) - f(v))^2 \cdot 2 \sum_{v \in V} \deg(v) |h(v)|} \end{aligned}$$

In conclusion,

$$\frac{\sum_{uv \in E} |h(u) - h(v)|}{\sum_{v \in V} \deg(v) |h(v)|} \leq \sqrt{2 \cdot \frac{\sum_{uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} \deg(v) |h(v)|}} \leq \sqrt{2 \cdot \frac{\sum_{uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} \deg(v) f(v)^2}},$$

as claimed.

Step 2 (threshold rounding). We will exhibit a cut $S \subseteq V$ such that $0 < \mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ and

$$\phi(S) \leq \frac{\sum_{uv \in E} |h(u) - h(v)|}{\sum_{v \in V} \deg(v) |h(v)|}.$$

Let $t \in \mathbb{R}$ be a parameter, and define $S_t \subseteq V$ as follows:

$$S_t := \begin{cases} \{v \in V : h(v) > t\} & \text{if } t \geq 0 \\ \{v \in V : h(v) < t\} & \text{if } t < 0. \end{cases} \tag{3.3}$$

Note that, since 0 is a degree-weighted median of h, $\mathrm{vol}(S_t)$ is at most $\mathrm{vol}(V)/2$ for any $t \in \mathbb{R}$. The "average" volume of S_t is

$$\int_{-\infty}^{\infty} \mathrm{vol}(S_t)\, dt = \sum_{v \in V} \deg(v) \int_{-\infty}^{\infty} \mathbb{1}[v \in S_t]\, dt = \sum_{v \in V} \deg(v)|h(v)|,$$

and the "average" cut size induced by S_t is

$$\begin{aligned}\int_{-\infty}^{\infty} |E(S_t, S_t^c)|\, dt &= \sum_{uv \in E} \int_{-\infty}^{\infty} \mathbb{1}[\min(h(u), h(v)) < t < \max(h(u), h(v))]\, dt \\ &= \sum_{uv \in E} |h(u) - h(v)|.\end{aligned}$$

As $|E(S_t, S_t^c)| = 0$ when $S_t = \emptyset$, there exists S_t such that $0 < \mathrm{vol}(S_t) \leq \mathrm{vol}(V)/2$ and

$$\phi(S_t) = \frac{|E(S_t, S_t^c)|}{\mathrm{vol}(S_t)} \leq \frac{\int_{-\infty}^{\infty} |E(S_t, S_t^c)|\, dt}{\int_{-\infty}^{\infty} \mathrm{vol}(S_t)\, dt} = \frac{\sum_{uv \in E} |h(u) - h(v)|}{\sum_{v \in V} \deg(v)|h(v)|}.$$

This completes the proof of the hard direction of Cheeger's inequality. □

The Sweep-Cut Algorithm

The proof of Cheeger's inequality above starts with an eigenvector $f : V \to \mathbb{R}$ corresponding to the second smallest eigenvalue, transforms it to an "ℓ_1" vector $h : V \to \mathbb{R}$, and takes a threshold cut

$$S_t = \{u \in V : h(u) > t\} \qquad \text{or} \qquad S_t = \{u \in V : h(u) < t\}.$$

Note that the transformation from f to h preserves relative order, i.e. $f(u) \geq f(v)$ iff $h(u) \geq h(v)$. Therefore, h exists only for the purpose of proving Cheeger's inequality, and the following "sweep-cut" algorithm (see Algorithm 1) using simply the second eigenvector f suffices to produce a cut S with $\phi(S) \leq \sqrt{2\lambda_2(G)}$ and $0 < \mathrm{vol}(S) \leq \mathrm{vol}(V)/2$. Moreover, as we shall see, the algorithm runs in near-linear time.

Theorem 3.1.2 (Runtime of Sweep-Cut Algorithm)**.** *Algorithm 1 runs in time $\tilde{O}(n+m)$ and returns a cut $S \subseteq V$ such that $\phi(S) \leq \sqrt{2\lambda_2(G)}$ and $0 < \mathrm{vol}(S) \leq \mathrm{vol}(V)/2$.*

Proof. The conductance guarantee follows from the fact that, in Algorithm 1, we considered all possible sets S_t in the threshold rounding step of the proof of Theorem 3.1.1 and take

Algorithm 1 The Sweep-Cut Algorithm

Input: Graph $G = (V, E)$

Output: A cut $S \subseteq V$ with $0 < \mathrm{vol}(S) \le \mathrm{vol}(V)/2$

1: Compute normalized Laplacian $L := I - D^{-1/2} A D^{-1/2}$
2: Find the second smallest eigenvector f of L
3: Sort the vertices by f, so that $f(v_1) \le f(v_2) \le \cdots \le f(v_n)$
4: $S \leftarrow \emptyset$
5: **for** $i \leftarrow 1$ to $n-1$ **do**
6: $\quad S_i \leftarrow \{v_1, \ldots, v_i\}$
7: $\quad$ **if** $\mathrm{vol}(S_i) > \mathrm{vol}(V)/2$ **then**
8: $\quad\quad S_i \leftarrow V \setminus S_i$
9: $\quad$ **end if**
10: $\quad$ **if** $\phi(S_i) < \phi(S)$ **or** $S = \emptyset$ **then**
11: $\quad\quad S \leftarrow S_i$
12: $\quad$ **end if**
13: **end for**
14: **return** S

the one with minimum conductance. Plus, S_i is never empty and we always set S_i to be the smaller side, so $0 < \mathrm{vol}(S) \le \mathrm{vol}(V)/2$.

It remains to prove the runtime bound. First, by Section 2.5.6 it takes $\tilde{O}(n+m)$ time to find a second smallest eigenvector f of L. Sorting the vertices takes $O(n \log n)$ time, and we can update $|E(S_i, S_i^c)|$ and $\mathrm{vol}(S_i)$ as i increases in $O(\deg(v_{i+1}))$ time, so lines 5-13 take $O(n+m)$ time in total. Therefore, the overall runtime is $\tilde{O}(n+m)$. □

Mixing Time

Another application of Cheeger's inequality is to use conductance to analyze the mixing time of random walks on graphs. The following corollary follows directly from Cheeger's inequality and (2.4) relating $\lambda_2(G)$ and mixing time.

Corollary 3.1.3 (Conductance and Mixing Time)**.** *Let $G = (V, E)$ be a graph and P be the lazy random walk defined in Section 2.6.1. Then,*

$$\frac{1}{\phi(G)} \lesssim \tau_{\mathrm{mix}}(P) \lesssim \frac{\log n}{\phi(G)^2}.$$

Therefore, Cheeger's inequality connects (i) spectral gap of the normalized Laplacian matrix, (ii) edge conductance of the graph and (iii) mixing time of random walks.

3.1.2 Bipartite Cheeger's Inequality

Motivated to design a "spectral" algorithm for finding approximate max cut, Trevisan [Tre09] proved the following "bipartite" Cheeger's inequality, that relates the largest eigenvalue λ_n to the so-called bipartiteness ratio of a graph. Observe that $2-\lambda_n(G)$ is the smallest eigenvalue of $I+A$ where A is the normalized adjacency matrix of G, and so

$$2-\lambda_n(G) = \min_{f:V\to\mathbb{R}} \frac{\langle f,(I+A)f\rangle}{\langle f,f\rangle} = \min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E}(f(u)+f(v))^2}{\sum_{v\in V}\deg(v)f(v)^2}.$$

The bipartiteness ratio β is then defined as the ℓ_1 version of $2-\lambda_n$:

$$\beta(x) := \frac{\sum_{uv\in E}|x(u)+x(v)|}{\sum_{v\in V}\deg(v)|x(v)|}$$

for $x: V\to\{-1,0,1\}$ and

$$\beta(G) := \min_{x:V\to\{-1,0,1\}} \beta(x).$$

The name for β comes from the following combinatorial viewpoint: the "bipartite edge boundary" $\delta(A,B)$ is defined as $E(A)\cup E(B)\cup\delta(A\cup B)$ where $E(A)$ (resp. $E(B)$) is the set of induced edges in A (resp. in B), and the "bipartite edge conductance" $\phi(A,B)$ as $|\delta(A,B)|/\text{vol}(A\cup B)$. Then, $\beta(G)$ is (up to constant) the problem of finding two disjoint subsets $A,B\subseteq V$ such that $\phi(A,B)$ is minimized. So, $\beta(G)$ is small if and only if, for some $S=A\sqcup B\subseteq V$, both $\phi(S)$ is small and $G[S]$ is close to bipartite with bipartitioning (A,B).

Trevisan's result uncovers a Cheeger inequality relating $\beta(G)$ and $2-\lambda_n(G)$.

Theorem 3.1.4 ([Tre09])**.** *Let $G=(V,E)$ be a graph. Then,*

$$\frac{2-\lambda_n(G)}{2} \leq \beta(G) \leq \sqrt{2(2-\lambda_n(G))}.$$

Proof. For the easy direction, given any solution $x:V\to\{-1,0,1\}$ to $\beta(G)$, take $f=x$, so that $f(u)^2=|f(u)|$ for all $u\in V$ and $(f(u)+f(v))^2\leq 2|f(u)+f(v)|$ for all $uv\in E$. Thus,

$$\frac{2-\lambda_n(G)}{2} \leq \frac{1}{2}\frac{\sum_{uv\in E}(f(u)+f(v))^2}{\sum_{v\in V}\deg(v)f(v)^2} \leq \frac{2\sum_{uv\in E}|f(u)+f(v)|}{2\sum_{v\in V}\deg(v)|f(v)|} = \beta(x).$$

For the hard direction, we would like to round any given $f : V \to \mathbb{R}$ to an $x : V \to \{-1, 0, 1\}$ such that

$$\beta(x) \leq \sqrt{2 \cdot \frac{\sum_{uv \in E}(f(u)+f(v))^2}{\sum_{v \in V} \deg(v) f(v)^2}}.$$

The proof here follows the same two-step approach: ℓ_2^2 to ℓ_1, then ℓ_1 to bipartite cuts.

Step 1 (ℓ_2^2 to ℓ_1). The first step is to construct an $h : V \to \mathbb{R}$ such that

$$\frac{\sum_{uv \in E} |h(u)+h(v)|}{\sum_{v \in V} \deg(v)|h(v)|} \leq \sqrt{2 \cdot \frac{\sum_{uv \in E}(f(u)+f(v))^2}{\sum_{v \in V} \deg(v) f(v)^2}}.$$

We define $h : V \to \mathbb{R}$ by

$$h(u) := \begin{cases} f(u)^2 & \text{if } f(u) \geq 0 \\ -f(u)^2 & \text{if } f(u) < 0. \end{cases}$$

Clearly, $\sum_{v \in V} \deg(v)|h(v)| = \sum_{v \in V} \deg(v) f(v)^2$. For the numerator, if $h(u)$ and $h(v)$ are of the same sign, then

$$|h(u)+h(v)| = f(u)^2 + f(v)^2 \leq (f(u)+f(v))^2 = |f(u)+f(v)|(|f(u)|+|f(v)|),$$

and if $h(u)$ and $h(v)$ are of different signs, then

$$|h(u)+h(v)| = |f(u)^2 - f(v)^2| \leq |f(u)+f(v)|(|f(u)|+|f(v)|).$$

Therefore, for any $uv \in E$ (indeed for any $u, v \in V$),

$$|h(u)+h(v)| \leq |f(u)+f(v)|(|f(u)|+|f(v)|), \tag{3.4}$$

and so

$$\begin{aligned} \sum_{uv \in E} |h(u)+h(v)| &\leq \sum_{uv \in E} |f(u)+f(v)|(|f(u)|+|f(v)|) \\ &\leq \sqrt{2 \sum_{uv \in E} (f(u)+f(v))^2 \sum_{uv \in E} (f(u)^2+f(v)^2)} \\ &= \sqrt{2 \sum_{uv \in E} (f(u)+f(v))^2 \sum_{v \in V} \deg(v) f(v)^2}. \end{aligned}$$

Thus, h satisfies the required condition.

Step 2 (threshold rounding). In the second step, we will exhibit a vector $x : V \to \{-1, 0, 1\}$ such that

$$\beta(x) \le \frac{\sum_{uv \in E} |h(u) + h(v)|}{\sum_{v \in V} \deg(v)|h(v)|}.$$

Let $t \in \mathbb{R}_{\ge 0}$ be a parameter, and define $x_t : V \to \{-1, 0, 1\}$ as follows:

$$x_t(u) := \mathbb{1}[h(u) > t] - \mathbb{1}[h(u) < -t].$$

Then, the "average" denominator value is

$$\int_0^\infty \sum_{v \in V} \deg(v)|x_t(v)| \, dt = \sum_{v \in V} d(v) \int_0^\infty \mathbb{1}[h(v) \notin [-t, t]] \, dt = \sum_{v \in V} \deg(v)|h(v)|$$

and the "average" numerator value is

$$\int_0^\infty \sum_{uv \in E} |x_t(u) + x_t(v)| \, dt = \sum_{uv \in E} \int_0^\infty |x_t(u) + x_t(v)| \, dt \overset{(*)}{=} \sum_{uv \in E} |h(u) + h(v)|,$$

where the (*) step may be seen to be true by verifying that

$$\int_0^\infty |x_t(u) + x_t(v)| \, dt = |h(u) + h(v)|$$

both when $h(u)$ and $h(v)$ have the same sign and when they have different signs. This completes the proof of the bipartite Cheeger's inequality. □

3.1.3 Higher-order Cheeger's Inequality

In 2012, [LOT12] and [LRTV12] independently proved the higher-order Cheeger's inequality, which relates higher eigenvalues λ_k to higher-order conductances ϕ_k. The latter is a combinatorial quantity that generalizes the conductance ϕ.

Definition 3.1.5 (k-Way Conductance)**.** *Let $G = (V, E)$ be a graph and $2 \le k \le n$. Then, the k-way conductance for disjoint nonempty sets $S_1 \sqcup S_2 \sqcup \cdots \sqcup S_k \subseteq V$ and for the graph G are defined as*

$$\phi_k(S_1, S_2, \ldots, S_k) := \max_{i \in [k]} \phi(S_i) \quad \text{and} \quad \phi_k(G) := \min_{S_1 \sqcup S_2 \sqcup \cdots \sqcup S_k \subseteq V} \phi_k(S_1, S_2, \ldots, S_k).$$

Note that $\phi_2(S, S^c) = \phi(S)$ for $S \subseteq V$ with $\text{vol}(S) \leq \text{vol}(V)/2$, and $\phi_2(G) = \phi(G)$. Interpreting ϕ as measuring whether a given graph can be broken into two partitions, each having a small fraction of outgoing edges (i.e. small conductance), ϕ_k essentially measures whether a given graph can be broken into k partitions, each having a small fraction of outgoing edges.

Below we outline the approach of [LOT12]. The following theorem is a robust generalization of Proposition 2.5.6 that $\lambda_k(G) = 0$ if and only if G has at least k connected components.

Theorem 3.1.6 (Higher-Order Cheeger Inequality [LOT12])**.** *Let $G = (V, E)$ be a graph. Then, for any $2 \leq k \leq n$,*

$$\frac{\lambda_k(G)}{2} \leq \phi_k(G) \lesssim k^2\sqrt{\lambda_k(G)} \quad \textit{and} \quad \phi_{\lfloor k/2 \rfloor}(G) \lesssim \sqrt{\lambda_k(G)\log k}.$$

To prove the easy direction, recall from Proposition 2.5.3 that $\lambda_k(G)$ has the following variational characterization:

$$\lambda_k(G) = \min_{\substack{Q:\text{ subspace of } \mathbb{R}^V \\ \dim(Q)=k}} \max_{f \in Q} \frac{\langle f, L'f \rangle}{\langle f, Df \rangle}.$$

Then, given disjoint nonempty sets $S_1 \sqcup S_2 \sqcup \cdots \sqcup S_k$ such that $\phi_k(G) = \phi_k(S_1, S_2, \ldots, S_k)$, construct $w_1, w_2, \ldots, w_k \in \mathbb{R}^V$ where $w_i(v) = \mathbb{1}[v \in S_i]$, and set $Q := \mathit{span}(\{w_1, \ldots, w_k\})$. Since the w_i's are nonzero and have disjoint support, Q has dimension k.
Let $\iota : V \to \{0\} \cup [k]$ be the map that takes vertices in S_i to i and vertices outside $S_1 \sqcup S_2 \sqcup \cdots \sqcup S_k$ to 0. Any vector $f \in Q$ is characterized by $a_i := f|_{S_i}$, which is the common value of f at any vertex in S_i. (We further choose $a_0 := 0$.) We can then upper bound its Rayleigh quotient by $\phi_k(G)$ as follows:

$$\begin{aligned}
\frac{\langle f, L'f \rangle}{\langle f, Df \rangle} &= \frac{\sum_{uv \in E}(a_{\iota(u)} - a_{\iota(v)})^2}{\sum_{v \in V} \deg(v) a_{\iota(v)}^2} \\
&\leq \frac{2\sum_{\substack{uv \in E \\ \iota(u) \neq \iota(v)}} \left(a_{\iota(u)}^2 + a_{\iota(v)}^2\right)}{\sum_{v \in V} \deg(v) a_{\iota(v)}^2} \\
&= \frac{2\sum_{i \in [k]} a_i^2 \, |E(S_i, S_i^c)|}{\sum_{i \in [k]} a_i^2 \, \text{vol}(S_i)} \\
&\leq 2 \max_{i \in [k]} \phi(S_i) = 2\phi_k(G).
\end{aligned}$$

This proves the easy direction.

The proof of the hard direction is somewhat more involved. The idea is that the first k eigenvectors $f_1, f_2, \ldots, f_k$ induces an embedding $f : V \to \mathbb{R}^k$ defined for each vertex $v \in V$ as $f(v) = \big(f_1(v), \ldots, f_k(v)\big)$, which is called the *spectral embedding* and enjoys certain nice properties. Below we will give an account of the proof ideas; we will see more technical details when we do a similar proof in Section 4.5.

In [LOT12], the goal is to extract, from the spectral embedding $f : V \to \mathbb{R}^k$, $\ell \leq k$ disjoint subsets $S_1, S_2, \ldots, S_\ell$, such that $\phi_l(S_1, S_2, \ldots, S_\ell)$ is small. The first step is to project $f : V \to \mathbb{R}^k$ into a low-dimensional solution $\bar{f} : V \to \mathbb{R}^h$, such that several properties are approximately preserved. (This step is not strictly necessary, but the reduction in dimension will give better partitions in the next steps.) The second step is to perform some sort of random partitioning of the low-dimensional embedding, to decompose $\bar{f}$ into ℓ embeddings with pairwise disjoint support. Finally, threshold rounding is applied to each of these embeddings to extract a set with small conductance, and putting them together we have our desired ℓ disjoint subsets with small ℓ-way conductance.

Step 1 (Gaussian Projection). In order for prepare for the partitioning step, [LOT12] first projects the spectral embedding into a low-dimensional solution $\bar{f} : V \to \mathbb{R}^h$. The projection used is random Gaussian projection, which we formally define here.

Definition 3.1.7 (Gaussian Projection)**.** *Let $f : V \to \mathbb{R}^p$ be an embedding where each vertex $v \in V$ is mapped to a vector $f(v) \in \mathbb{R}^p$. Given an integer $1 \leq h \leq p$, let Γ be an $h \times p$ matrix where each entry $\Gamma_{i,j}$ for $1 \leq i \leq h$ and $1 \leq j \leq p$ is an independent standard Gaussian random variable $N(0,1)$. The Gaussian projection $\bar{f} : V \to \mathbb{R}^h$ of f is an embedding of each vertex $v \in V$ to an h-dimensional vector defined as*

$$\bar{f}(v) = \frac{1}{\sqrt{h}} \cdot \Gamma\big(f(v)\big).$$

The properties of f and of $\bar{f}$ that are of interest are the following:

- The *energy* of f, which is defined as $\mathcal{E}(f) := \sum_{uv \in E} \|f(u) - f(v)\|^2$;
- The *(π-)mass* of f, which is defined as $\|f\|_\pi^2 := \sum_{v \in V} \pi(v) \|f(v)\|^2$ for any given $\pi : V \to \mathbb{R}_{\geq 0}$;
- The *spreading property*, which informally asserts that the vectors $f(v)$ cannot be too concentrated in any one direction and is defined below.

Definition 3.1.8 (Spreading Property [LOT12])**.** *Let $\pi : V \to \mathbb{R}_{\geq 0}$ be a weight function on the vertices. For two parameters $\delta \in [0,1]$ and $\eta \in [0,1]$, an embedding $f : V \to \mathbb{R}^h$ is called (δ, η)-spreading if for every subset $S \subseteq V$,*

$$\operatorname{diam}_{d_f}(S) \leq \delta \quad \implies \quad \sum_{v \in S} \pi(v) \|f(v)\|^2 \leq \eta \cdot \sum_{v \in V} \pi(v) \|f(v)\|^2 ,$$

where $\operatorname{diam}_{d_f}(S) := \max_{u,v \in S} d_f(u,v)$ is the diameter of the set S under the radial projection distance function d_f to be defined below in Definition 3.1.10.

In this section, in the definitions of mass and spreading property, we will take $\pi = \deg$.

The following main lemma describes how these three properties are approximately preserved under random Gaussian projection. The mass and energy guarantees together give a bound on $\mathcal{R}(\bar{f})$, whereas the spreading property guarantee allows us to proceed with the partitioning of $\bar{f}$ in the next step.

Lemma 3.1.9 ([LOT12, Lemma 4.3])**.** *Let $f : V \to \mathbb{R}^p$ be an embedding that is (δ, η)-spreading. Let $\bar{f} : V \to \mathbb{R}^h$ be a Gaussian projection of f as defined in Definition 3.1.7. Let $\pi : V \to \mathbb{R}_{\geq 0}$. For some value* [1]

$$h \lesssim \frac{1}{\delta^2}\Big(\log\Big(\frac{1}{\eta\delta}\Big)\Big),$$

with probability at least $1/2$, the following three properties hold simultaneously:

$$\mathcal{E}(\bar{f}) \leq 4\mathcal{E}(f) \quad \text{and} \quad \|\bar{f}\|_\pi^2 \geq \frac{1}{2}\|f\|_\pi^2 \quad \text{and} \quad \bar{f} \text{ is } \Big(\frac{\delta}{4}, (1+\delta)\eta\Big)\text{–spreading.}$$

The proof uses standard concentration inequalities and the interested reader is directed to [LOT12, Section 4.1] for details.

Step 2 (Spectral Partitioning). Since the eigenvectors are orthonormal, the spectral embedding satisfies the isotropy condition $\sum_{v \in V} f(v) f(v)^T = I_k$ by Proposition 2.5.8. Lee, Oveis Gharan and Trevisan observed that the isotropy condition implies that not many points can be close in the radial projection distance defined below.

[1] Lemma 4.3 in [LOT12] was stated slightly differently. Their assumptions are that $f : V \to \mathbb{R}^k$ and $\eta \geq 1/k$, and their conclusion is that $h \lesssim \frac{1}{\delta^2}\log(\frac{k}{\delta})$. We note that the dependency on k in their conclusion is based on the substitution $\eta = 1/k$ in the bound on h we stated, which has no dependency on the ambient dimension p. Their proof, without the substitution $\eta = 1/k$, gives the bound we stated.

Definition 3.1.10 (Radial Projection Distance [LOT12]). *Let $G = (V, E)$ be a graph and $f : V \to \mathbb{R}^h$ be an embedding of the vertices. For each pair of vertices $u, v \in V$, the radial projection distance between u and v is defined as*

$$d_f(u,v) := \left\| \frac{f(u)}{\|f(u)\|} - \frac{f(v)}{\|f(v)\|} \right\|$$

if $\|f(u)\| > 0$ and $\|f(v)\| > 0$. Otherwise, if $f(u) = f(v) = \vec{0}$ then $d_f(u,v) := 0$, else $d_f(u,v) = \infty$.

More precisely, they proved the following bound on the spreading property in Definition 3.1.8 of the embedding f. Their result is originally stated for an embedding f satisfying the isotropy condition, but the same proof works for an embedding f satisfying the sub-isotropy condition that we will encounter in Proposition 4.5.2.

Lemma 3.1.11 (Sub-Isotropy Implies Spreading [LOT12, Lemma 3.2]). *Let $G = (V, E)$ be an undirected graph and $\pi : V \to \mathbb{R}_{\geq 0}$. Suppose $f : V \to \mathbb{R}^h$ is an embedding with mass $\|f\|_\pi^2$. Then, for any $\delta \in [0, 1)$,*

$$\sum_{u \in V} \pi(u) f(u) f(u)^T \preceq I_h \quad \implies \quad f \text{ is } \left(\delta, \frac{1}{\|f\|_\pi^2 \cdot (1 - \delta^2)}\right)\text{-spreading.}$$

As the embedding is spreading, any subset of points with small diameter in radial projection distance cannot have too much mass. In order to construct many disjointly supported embeddings $f_1, \ldots, f_l : V \to \mathbb{R}^h$, the points in $\mathbb{R}^h$ are partitioned into many groups of small diameter using the following definition and theorem from metric geometry.

Definition 3.1.12 (Padded Decomposition [LOT12]). *Let (X, d_X) be a finite metric space. For $\rho, \alpha, \beta > 0$, a random partitioning $\mathcal{P}$ of X is called (ρ, α, β)-padded if*

- *each partition in $\mathcal{P}$ has diameter at most ρ with respect to the distance function d_X;*
- $\Pr\left[B\big(x, \frac{\rho}{\alpha}\big) \subseteq \mathcal{P}(x)\right] \geq \beta$ *for every $x \in X$, where $B\big(x, \frac{\rho}{\alpha}\big)$ is the open ball of radius $\frac{\rho}{\alpha}$ centered at x and $\mathcal{P}(x)$ is the partition in $\mathcal{P}$ that contains x.*

Theorem 3.1.13 (Existence of Padded Partition [GKL03], [LOT12, Theorem 2.3]). *Let (X, d_X) be a finite metric space. If $X \subseteq \mathbb{R}^h$ and d_X is the Euclidean distance, then for every $\rho > 0$ and $\delta \in (0, 1)$, X admits a $\big(\rho, O(\frac{h}{\delta}), 1 - \delta\big)$-padded random partitioning.*

Let $\mathcal{P} = P_1 \sqcup P_2 \sqcup \cdots \sqcup P_T$ be a random partitioning sampled from Theorem 3.1.13. By the second property in Definition 3.1.12, there is only a small fraction of points close to the boundaries of the partitions. The points close to the boundaries are removed to form $P_1' \sqcup P_2' \sqcup \cdots \sqcup P_T'$, so that the distance between each pair P_i' and P_j' is lower bounded by say 2ε.

Define the (π-)mass of a subset $S \subseteq V$ to be $\mu(S) := \sum_{v \in S} \pi(v) \left\| \bar{f}(v) \right\|^2$. Each P_i does not have too much mass by the spreading property of $\bar{f}$, since each partition has diameter bounded by ρ. So, each P_i' does not have too much mass either, and by a greedy procedure they can be grouped into disjoint sets $S_1, \ldots, S_k$ where each has mass at least $\frac{1}{2k}$ of the total. Then the disjointly supported functions $f_1, \ldots, f_k$ are constructed on $S_1, \ldots, S_k$ by the following smooth localization procedure.

Lemma 3.1.14 (Smooth Localization [LOT12, Lemma 3.3])**.** *Let $G = (V, E)$ be an undirected graph and $f : V \to \mathbb{R}^h$ be an embedding. For any $S \subseteq V$ and any $\varepsilon > 0$, there is a mapping $f' : V \to \mathbb{R}^h$ which satsifies the following three properties:*

1. *$f'(v) = f(v)$ for all $v \in S$,*
2. *$\operatorname{supp}(f') \subseteq N_\varepsilon(S)$ where $N_\varepsilon(S) := \{v \in V \mid \exists u \in S \text{ with } d_f(u, v) \leq \varepsilon\}$ denotes the set of vertices with radial projection distance at most ε from S,*
3. *for $uv \in E$, $\|f'(u) - f'(v)\| \leq \left(1 + \frac{2}{\varepsilon}\right) \|f(u) - f(v)\|$.*

To summarize, the energy of each f_i is upper bounded by the third item in Lemma 3.1.14, and the mass of each f_i is lower bounded by $\frac{1}{2k}$ fraction of the total mass by the spreading property in Lemma 3.1.11. Property 2 of Lemma 3.1.14 and the condition that $d_f(S_i, S_j) \geq 2\varepsilon$ for all $i \neq j$ ensure that the functions f_i are indeed disjointly supported.

Step 3 (Cheeger Rounding). From the previous step, we constructed k disjointly supported functions $f_1, \ldots, f_k : V \to \mathbb{R}$ such that $R(f_i) = \mathcal{E}(f_i) / \|f_i\|_{\text{deg}}^2$ are uniformly bounded above. We can apply, to each f_i, the two-step approach to find small-conductance sets in Section 3.1.1, with the slight modification of taking $c = 0$ instead of the degree-weighted median of f in the ℓ_2^2 to ℓ_1 step. This ensures that the sets U_i produced by the rounding satisfy:

- $U_i \subseteq \operatorname{supp}(f_i) \subseteq S_i$ for each $i \in [k]$. In particular they are pairwise disjoint;
- $\phi(U_i) \lesssim \sqrt{R(f_i)}$ for each $i \in [k]$.

Then, $\phi_k(G) \leq \phi_k(U_1, \ldots, U_k) \lesssim \max_{i\in[k]} \sqrt{R(f_i)}$. By choosing suitable parameters, this is how they derive the hard direction of Theorem 3.1.6.

Remark 3.1.15 (Optimal Result)**.** *In [LOT12], in order to obtain the strongest version of higher-order Cheeger inequality as in Theorem 3.1.6, they abandoned the approach of localizing functions produced by padded decomposition. Instead, they used the so-called Lipschitz random partitioning, which asserts that the probability that two points in a metric space are separated by the partitioning is upper bounded by a factor times their distance. The Lipschitz property suits their setting because the energy is a simple sum of edge energy terms over all the edges (see [LOT12, Section 4.2] for details), but a significant modification of the partitioning tool is likely needed to apply that to our setting in Section 4.5.*

3.1.4 Improved Cheeger's Inequality

[KLL+13] derived the "improved" Cheeger's Inequality to describe situations where there is a tighter inequality relating conductance ϕ and second eigenvalue λ_2. They discovered that, if the graph does not have many (i.e. $\omega(1)$) sparse cuts, then the sweep-cut algorithm finds a cut whose conductance is of order $O(\phi)$ instead of $O(\sqrt{\phi})$. In many applications, notably image segmentation [SM00], the input graph indeed satisfies this prior condition, and so this result gives a theoretical justification of the performance of the sweep-cut algorithm for spectral partitioning. The following is their main result, stated in terms of λ_k. Qualitatively this is the same as considering k-way conductance ϕ_k, in light of the higher-order Cheeger's inequality relating the two parameters (see [KLL16] for a tight result that lower bounds ϕ in terms of λ_2 and ϕ_k).

Theorem 3.1.16 (Improved Cheeger's Inequality [KLL+13])**.** *Let $G = (V, E)$ be a graph. Then, for any $2 \leq k \leq n$,*

$$\phi(G) \lesssim \frac{k\lambda_2(G)}{\sqrt{\lambda_k(G)}}.$$

First we begin with the definition of a k-step function.

Definition 3.1.17 (k-Step Function and Approximation)**.** *Let $G = (V, E)$ be an undirected graph. Given $y : V \to \mathbb{R}$ and $k \in \mathbb{N}$, we call y a k-step function if the number of distinct values in $\{y(v)\}_{v\in V}$ is at most k.*

The proof consists of two main components. First of all, given a graph $G = (V, E)$, for any function $f : V \to \mathbb{R}$ recall from Section 2.5.3 that its Rayleigh quotient is defined as

$$R(f) := \frac{\sum_{uv\in E}(f(u) - f(v))^2}{\sum_{v\in V} \deg(v) f(v)^2}.$$

We have seen from Section 3.1.1 that if $\sum_{v \in V} \deg(v) f(v) = 0$ then the sweep-cut algorithm on f produces a cut $S \subseteq V$ such that $0 < \mathrm{vol}(S) \le \mathrm{vol}(V)/2$ and

$$\phi(S) \lesssim \sqrt{R(f)}.$$

The first main component is to show that, if f is closely approximated by a step function with few steps, then the sweep-cut algorithm produces a cut whose conductance bound depends on both $R(f)$ and the quality of approximation.

Before stating the result, we recall the following definitions from Section 3.1.3:

- Let $\|f\|_{\deg}^2 := \sum_{v \in V} \deg(v) f(v)^2$ be the mass of f;
- Let $\mathcal{E}(f) := \sum_{uv \in E} (f(u) - f(v))^2$ be the energy of f.

Using this terminology, $R(f) = \mathcal{E}(f) / \|f\|_{\deg}^2$.

Proposition 3.1.18. *Given a graph $G = (V, E)$. Let $f : V \to \mathbb{R}$ be a function with $\sum_{v \in V} \deg(v) f(v) = 0$ and $f^* : V \to \mathbb{R}$ be a k-step function for some $k \in \mathbb{N}$. Then,*

$$\frac{\phi(G) \|f\|_{\deg}^2}{k} \lesssim \mathcal{E}(f) + \|f - f^*\|_{\deg} \cdot \sqrt{\mathcal{E}(f)}.$$

Moreover, the sweep-cut algorithm in Algorithm 1 produces from f a cut $S \subseteq V$ such that $0 < \mathrm{vol}(S) \le \mathrm{vol}(V)/2$ and

$$\phi(S) \lesssim k \left(\frac{\mathcal{E}(f) + \|f - f^*\|_{\deg} \cdot \sqrt{\mathcal{E}(f)}}{\|f\|_{\deg}^2} \right)$$

As a special case, if f is itself a k-step function, then taking $f^* = f$ above, the sweep-cut algorithm can guarantee that the produced cut S satisfies

$$\phi(S) \lesssim k \cdot R(f).$$

The second main component is to show that, if $\lambda_k(G)$ is large, then any $f : V \to \mathbb{R}$ is well approximated by a $(2k)$-step function.

Proposition 3.1.19. *Given a graph $G = (V, E)$ and an integer k where $2 \le k \le n$. Then, for any $f : V \to \mathbb{R}$ there is a $(2k)$-step function $f^* : V \to \mathbb{R}$ such that*

$$\|f - f^*\|_{\deg}^2 \le \frac{\mathcal{E}(f)}{\lambda_k(G)}.$$

To finish the proof, choose f such that $R(f) = \lambda_2(G)$. Combining Proposition 3.1.18 and Proposition 3.1.19, and using the fact that $\lambda_k(G) \le O(1)$, the desired result holds.

Proof of Proposition 3.1.18. Given $f : V \to \mathbb{R}$ with $\sum_{v \in V} \deg(v) f(v) = 0$. Let $f^* : V \to \mathbb{R}$ be a k-step function, and $t_1 < t_2 < \cdots < t_k$ be the values that $f^*(v)$ takes. To analyze the sweep-cut algorithm, we follow the two-step plan for proving Theorem 3.1.1. The second step of the proof stays the same, and it suffices to exhibit an $h : V \to \mathbb{R}$ such that

$$\frac{\sum_{uv \in E} |h(u) - h(v)|}{\sum_{v \in V} \deg(v) |h(v)|} \lesssim k \left(\frac{\mathcal{E}(f) + \|f - f^*\|_{\deg} \cdot \sqrt{\mathcal{E}(f)}}{\|f\|_{\deg}^2} \right)$$

and 0 is a degree-weighted median of h.

Choose $c \in \mathbb{R}$ to be a degree-weighted median of f, and define $h : V \to \mathbb{R}$ so that

$$h(u) := \int_c^{f(u)} \nu(t)\, dt,$$

where $\nu(t) := \min_{i \in [k]} |t - t_i|$ is the distance from t to the closest value of the k-step function f^*. Note that the integral is negative when $f(u) < c$ and positive when $f(u) > c$, and so 0 is a degree-weighted median of $h(u)$.

One can think of $\nu(t)$ as a modified probability density function for sampling the sets S_t in the threshold rounding step, by using the information of f^*. This definition of h prioritizes choosing S_t with t far away from the points t_i in f^*, i.e. with $\nu(t)$ large.

Comparatively, the definition of h in (3.1) in Theorem 3.1.1 is equivalent to setting $f^* \equiv c$, so that $\nu(t) = |t - c|$. The Cauchy-Schwarz loss in the ℓ_2^2 to ℓ_1 step is then due to the possibility of $\nu(f(u))$ and $\nu(f(v))$ both being large, but $f(u)$ and $f(v)$ close. This more refined way of defining ν uses the information provided by f^* to avoid the Cauchy-Schwarz loss. In the ideal case where $f = f^*$, $\nu(f(v))$ is always equal to 0. Generally, however, one pays a price for f^* being potentially not the ideal approximation for f.

It remains to upper bound the numerator and lower bound the denominator. For the denominator term, we have

$$\sum_{v \in V} \deg(v) |h(v)| = \sum_{v \in V} \deg(v) \left| \int_c^{f(v)} \nu(t)\, dt \right|.$$

For a fixed $v \in V$, suppose $c < t_i < t_{i+1} < \cdots < t_j < f(v)$ are the points in the support of f^* between c and $f(v)$. Then, using the definition of $\nu(t)$ we may verify that

$$
\begin{aligned}
\int_c^{f(v)} \nu(t)\,dt &= \left[\int_c^{t_i} + \int_{t_i}^{t_{i+1}} + \cdots + \int_{t_j}^{f(v)}\right] \nu(t)\,dt \\
&\overset{(*)}{\geq} \frac{1}{4}\left[(t_i - c)^2 + (t_{i+1} - t_i)^2 + \cdots + (f(v) - t_j)^2\right] \\
&\overset{(**)}{\geq} \frac{(f(v)-c)^2}{4(j-i+2)} \geq \frac{(f(v)-c)^2}{4(k+1)},
\end{aligned}
$$

where in $(*)$ we used the fact that the points t_i lie outside of the open interval (a, b) for each of the integral $\int_a^b$, and so

$$
\int_a^b \nu(t)\,dt \geq \int_a^b \min(t-a, b-t)\,dt = \frac{(b-a)^2}{4},
$$

and $(**)$ is by Corollary 2.10.2. The case where $f(v) \leq c$ is similarly handled, and if there are no points between c and $f(v)$ we nevertheless have

$$
\int_c^{f(v)} \nu(t)\,dt \geq \frac{1}{4}(f(v)-c)^2 \geq \frac{(f(v)-c)^2}{4(k+1)}
$$

by the same reasoning. Therefore,

$$
\sum_{v \in V} \deg(v)|h(v)| \gtrsim \frac{1}{k} \sum_{v \in V} \deg(v)(f(v)-c)^2.
$$

For the numerator term, we have

$$
\begin{aligned}
& \sum_{uv \in E} |h(u) - h(v)| \\
=\; & \sum_{uv \in E} \int_{\min(f(u),f(v))}^{\max(f(u),f(v))} \nu(t)\,dt \\
\overset{(***)}{\leq}\; & \sum_{uv \in E} |f(u) - f(v)| \min\left(\nu(f(u)) + \frac{|f(u)-f(v)|}{2}, \nu(f(v)) + \frac{|f(u)-f(v)|}{2}\right) \\
\leq\; & \sum_{uv \in E} |f(u) - f(v)| \left(\frac{|f(u) - f^*(u)| + |f(v) - f^*(v)| + |f(u) - f(v)|}{2}\right) \\
=\; & \frac{1}{2}\left(\mathcal{E}(f) + \sum_{uv \in E} |f(u) - f(v)|(|f(u) - f^*(u)| + |f(v) - f^*(v)|)\right) \\
\lesssim\; & \mathcal{E}(f) + \|f - f^*\|_{\deg} \cdot \sqrt{\mathcal{E}(f)} \quad \text{(Cauchy-Schwarz)}.
\end{aligned}
$$

The step $(***)$ uses the fact that $\nu(t)$ is 1-Lipschitz, and so both $\nu(f(u)) + |f(u) - f(v)|/2$ and $\nu(f(v)) + |f(u) - f(v)|/2$ are upper bounds on the average value

$$\frac{1}{|f(u) - f(v)|} \int_{\min(f(u),f(v))}^{\max(f(u),f(v))} \nu(t)\, dt.$$

Combining the two bounds yields the desired guarantee on h. □

Proof of Proposition 3.1.19. If $\lambda_k(G) = 0$ there is nothing to prove, so consider only the case where $\lambda_k(G) > 0$.

Let $M > 0$ be a parameter to be determined later. Let $t_0 = -\infty$ and successively choose $t_1, t_2, \ldots$ such that $t_i > t_{i-1}$ is the smallest real number such that the following function

$$\bar{f}_i(u) := \begin{cases} \min\left(f(u) - t_{i-1}, t_i - f(u)\right), & \text{if } t_{i-1} < f(u) \leq t_i \\ 0, & \text{otherwise} \end{cases}$$

satisfies $\left\|\bar{f}_i\right\|_{\deg}^2 \geq M$. The role of the functions $\bar{f}_i$ is to measure how well the two threshold values t_{i-1} and t_i approximate the values of the function at between them. If such a t_i does not exist, we set $t_i = \infty$ and terminate the process. The process always terminates within n steps, and if it terminates with $t_{k+1} = \infty$ then the following function (which is determined once f and the t_i's are fixed)

$$f^*(u) := \arg\min_{t_i : i \in [k]} |f(u) - t_i|$$

is a k-step function. Observe also that the $\bar{f}_i$'s have disjoint support, and in fact

$$\sum_{i=1}^{k+1} \left\|\bar{f}_i\right\|_{\deg}^2 = \|f - f^*\|_{\deg}^2 .$$

Consider the scenario that the process does not terminate after $2k$ steps. That means $\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_{2k}$ are all well-defined and each having mass $\left\|\bar{f}_i\right\|_{\deg}^2 \geq M$. Moreover, the sum of their energies is

$$\begin{aligned} \sum_{i=1}^{2k} \mathcal{E}(\bar{f}_i) &= \sum_{uv \in E} \sum_{i=1}^{2k} (\bar{f}_i(u) - \bar{f}_i(v))^2 \\ &\leq \sum_{uv \in E} (f(u) - f(v))^2 = \mathcal{E}(f). \end{aligned}$$

Denoting $i(u)$ to be the unique index (possibly $> 2k$) that $t_{i(u)-1} < f(u) \le t_{i(u)}$, the inequality above is verified by considering separately the two cases where $i(u) = i(v)$ and $i(u) \ne i(v)$, and using the definition of $\bar{f}_i$. That means at least k of the $\bar{f}_i$'s for $i \in [2k]$ must each satisfy $R(\bar{f}_j) \le \mathcal{E}(f)/kM$, and by Proposition 2.5.4 we have

$$\lambda_k(G) \le \frac{2\mathcal{E}(f)}{kM}.$$

Choose $M := \frac{4\mathcal{E}(f)}{k\lambda_k(G)}$ so that the above inequality *fails*. This means the process terminates after at most $2k$ steps, and

$$\|f - f^*\|_{\deg}^2 \le 2kM \lesssim \frac{\mathcal{E}(f)}{\lambda_k(G)},$$

proving the proposition. □

3.1.5 Cheeger's Inequality for Small-Set Expansion

Given a graph $G = (V, E)$. Its δ-small set expansion is defined as

$$\varphi_\delta(G) := \min_{0<|S|\le\delta n} \varphi(S),$$

where $\varphi(S)$ is the edge expansion of S in Section 2.3.1. Algorithms for small-set expansion would be useful in applications such as community detection, but it is conjectured that, for any $\varepsilon \in (0, 1/2)$ there exists a $\delta < 1$ such that it is NP-hard to distinguish between $\varphi_\delta(G) \le \varepsilon$ and $\varphi_\delta(G) \ge 1 - \varepsilon$ [RS10]. This conjecture is called the small-set expansion hypothesis.

On the positive side, Arora, Barak, and Steurer [ABS10] derived a Cheeger-type inequality for small-set expansion. They applied the result to design subexponential time approximation algorithms for unique games and small set expansion.

Theorem 3.1.20 (Cheeger Inequality for Small-Set Expansion [ABS10])**.** *Let $G = (V, E)$ be a graph and $\beta, \eta \in (0, 1)$ be two parameters. Then, for $k \ge n^{O(\beta)}$, there exists $S \subseteq V$ such that $|S| \le n^{1-\beta}$ and $\varphi(S) \lesssim \sqrt{\lambda_k(G)/\beta}$.*

This result inspired subsequent research on spectral theory using higher eigenvalues (e.g. [LOT12, LRTV12, KLL+13]). We do not study small-set expansion in this thesis. While our results for k-way expansion can say something about δ-small set expansion for $\delta = \Theta(1/k)$, it is only a one-way bound that if the spectral quantity is small, then a small sparse cut can be extracted.

3.2 Vertex Expansion

3.2.1 Classical Result

Tanner [Tan84], Alon and Milman [AM85], and Alon [Alo86] studied the relation between the spectral gap and the vertex expansion of a graph, proving the following two-way bound.

Theorem 3.2.1 ([Alo86, Lemmas 2.2, 2.4], [Tan84, AM85])**.** *Let $G = (V, E)$ be an undirected graph with maximum degree Δ. Let $\lambda_2'(G)$ be the second smallest eigenvalue of the unnormalized Laplacian $L'(G)$, and $\psi(G)$ be the vertex expansion of G. Then,*

$$\psi(G) \geq \frac{2\lambda_2'(G)}{\Delta + 2\lambda_2'(G)} \qquad \textit{and} \qquad \lambda_2'(G) \geq \frac{\psi(G)^2}{4 + 2\psi(G)^2}.$$

Proof. We prove a slightly different "easy direction" that

$$\psi(G) \geq \frac{\lambda_2'(G)}{\Delta}.$$

The idea is the same as the proof of the original statement as found in [Tan84, AM85]. First, $\lambda_2'(G)$ admits a similar Rayleigh quotient characterization to $\lambda_2(G)$ in Proposition 2.5.2:

$$\lambda_2'(G) = \min_{f: f \perp \mathbb{1}} \frac{\sum_{uv \in E}(f(u) - f(v))^2}{\sum_{v \in V} f(v)^2},$$

the strategy is to construct appropriate test functions f. Let $S \subseteq V$ such that $0 < |S| \leq n/2$ and $\psi(S) = \psi(G)$. Write $T := V \setminus (S \cup \partial(S))$ and consider $f : V \to \mathbb{R}$ as follows:

$$f(u) = \begin{cases} \frac{-1}{|S|}, & \text{if } u \in S \\ 0, & \text{if } u \in \partial(S) \\ \frac{1}{|T|}, & \text{if } u \in T. \end{cases}$$

Note that $\sum_u f(u) = 0$. Assuming without loss of generality that $|S| \leq |T|$, we have

$$\begin{aligned} \sum_{uv \in E} (f(u) - f(v))^2 &= \frac{|E(S, \partial(S))|}{|S|^2} + \frac{|E(T, \partial(S))|}{|T|^2} \\ &\leq \Delta |\partial(S)| \max\left(\frac{1}{|S|^2}, \frac{1}{|T|^2}\right) \\ &= \Delta \frac{\psi(S)}{|S|} \end{aligned}$$

as well as

$$\sum_{v\in V} f(v)^2 = \frac{|S|}{|S|^2} + \frac{|T|}{|T|^2} = \frac{1}{|S|} + \frac{1}{|T|} \geq \frac{1}{|S|}.$$

This gives $\psi(G) = \psi(S) \geq \lambda_2'(G)/\Delta$.

The "hard direction" is due to Alon [Alo86]. The proof is by taking an eigenvector f of $L'(G)$, constructing a flow network based on it, and using max-flow min-cut duality to connect $\psi(G)$ and $\lambda_2'(G)$.

Let $f : V \to \mathbb{R}$ be an eigenvector of $L'(G)$ with eigenvalue $\lambda_2'(G)$. Let $V^+ := \{v \in V : f(v) > 0\}$. Without loss of generality, assume that $0 < \mathrm{vol}(V^+) \leq \mathrm{vol}(V)/2$ (otherwise, replace f by $-f$). Define the function $g := f^+ = \max(f, 0)$ which will be useful later. Write ψ for $\psi(G)$ for simplicity. Define the following flow network, with vertex set $\{s,t\} \cup V^+ \cup V'$ where $V' := \{v' : v \in V\}$ is a distinct copy of V, and arcs as follows (see Figure 3.1):

- (s, v) for each $v \in V^+$ with capacity $1 + \psi$;
- (v, v') for each $v \in V^+$ and (v, u') for each $vu \in E$ and $v \in V^+$, with unit capacity;
- (v', t) for each $v \in V$ with unit capacity.

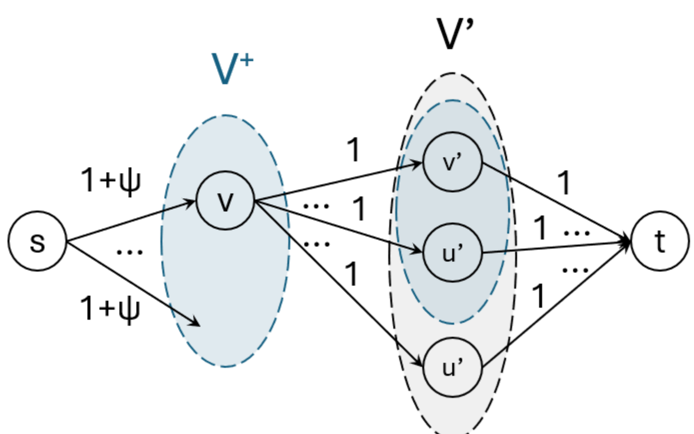


Figure 3.1: The flow network construction. The numbers indicate arc capacities. From $v \in V^+$ the outgoing arcs are (v, v') and (v, u') for all $vu \in E$.

We claim that the minimum s-t cut is $(1 + \psi)|V^+|$. First, this value is attained by the cut $\{s\}$. Second, all other s-t cuts have value at least $(1 + \psi)|V^+|$, as we shall prove.

Suppose that the cut is $\{s\} \cup S \cup T'$ where $S \subseteq V^+$ and $T' \subseteq V'$. Then, the following arcs (and possibly some more) are cut: (s, u) where $u \in V^+ \setminus S$, either (u, u') or (u', t) for all $u \in S$, and either (u, v') or (v', t) for all $uv \in E$ with $u \in S$. The value of the cut is then at least

$$(1 + \psi)|V^+ \setminus S| + |S| + |\partial S| \geq (1 + \psi)(|V^+ \setminus S| + |S|) = (1 + \psi)|V^+|.$$

By max-flow min-cut duality, the maximum s-t flow is $(1+\psi)|V^+|$, meaning that it saturates all arcs (s, v) for $v \in V^+$. Consider a maximum flow θ, with $\theta(u, v')$ being the amount of flow going through the arc (u, v'). Note that $\theta(u, v')$ satisfy:

- $0 \leq \theta(u, v') \leq 1$ for all arcs (u, v');
- $\sum_{v \in V} \theta(u, v') = 1 + \psi$ for all $u \in V^+$;
- $\sum_{u \in V^+} \theta(u, v') \leq 1$ for all $v \in V$.

We relate the Rayleigh quotient

$$\frac{\sum_{uv \in E}(f(u) - f(v))^2}{\sum_{v \in V} f(v)^2}$$

of f to that of g, and apply Cauchy-Schwarz inequality while utilizing the flow values to upper bound the latter quantity. By the choice of f,

$$\sum_{v: uv \in E} (f(u) - f(v)) = \lambda_2'(G) \cdot f(u)$$

for all $u \in V^+$. Multiplying each equation by $f(u)$ and summing over all $u \in V^+$,

$$\sum_{u \in V^+} \sum_{v: uv \in E} f(u)(f(u) - f(v)) = \lambda_2'(G) \sum_{u \in V^+} f(u)^2.$$

The contribution of an edge in $V^+ \times V^+$ to LHS is $(f(u) - f(v))^2 = (g(u) - g(v))^2$. The contribution of an edge $uv \in V^+ \times (V \setminus V^+)$ to LHS is $f(u)(f(u) - f(v)) \geq f(u)^2 = (g(u) - g(v))^2$. All other edges contributes $0 = (g(u) - g(v))^2$ to LHS. Therefore,

$$\sum_{uv \in E} (g(u) - g(v))^2 \leq \sum_{u \in V^+} \sum_{v: uv \in E} f(u)(f(u) - f(v)) = \lambda_2'(G) \sum_{u \in V^+} f(u)^2 = \lambda_2'(G) \sum_{u \in V} g(u)^2.$$

Finally,

$$
\begin{aligned}
\lambda_2'(G) &\geq \frac{\sum_{uv\in E}(g(u)-g(v))^2}{\sum_{v\in V} g(v)^2} \qquad \text{(by the inequality above)} \\
&= \frac{\sum_{uv\in E}(g(u)-g(v))^2 \sum_{uv\in E}\theta^2(u,v')(g(u)+g(v))^2}{\sum_{v\in V} g(v)^2 \sum_{uv\in E}\theta^2(u,v')(g(u)+g(v))^2} \\
&\geq \frac{\left[\sum_{uv\in E}\theta(u,v')(g(u)^2-g(v)^2)\right]^2}{\left[\sum_{v\in V} g(v)^2\right]\left[\sum_{uv\in E}\theta^2(u,v')(g(u)+g(v))^2\right]},
\end{aligned} \tag{3.5}
$$

where we applied Cauchy-Schwarz inequality to obtain the last line. We lower bound the sum in the numerator of (3.5) as

$$
\begin{aligned}
\sum_{uv\in E}\theta(u,v')(g(u)^2-g(v)^2) &= \sum_{u\in V^+} g(u)^2\left[\sum_{v:uv\in E}(\theta(u,v')-\theta(v,u'))\right] \\
&\geq \sum_{u\in V^+} g(u)^2(1+\psi-1) = \psi\sum_{u\in V} g(u)^2,
\end{aligned}
$$

and we upper bound the second sum in the denominator in $(*)$ as

$$
\begin{aligned}
\sum_{uv\in E}\theta^2(u,v')(g(u)+g(v))^2 &\leq 2\sum_{u\in V^+} g(u)^2\left[\sum_{v:uv\in E}(\theta^2(u,v')+\theta^2(v,u'))\right] \\
&\leq 2\sum_{u\in V^+} g(u)^2(2+\psi^2),
\end{aligned}
$$

where for the last inequality we used the properties of $\theta(\cdot,\cdot)$ and the fact that $\sum_v \theta^2(u,v')$ and $\sum_v \theta^2(v,u')$ are maximized by taking as many $\theta(\cdot,\cdot)=1$ as possible.

Plugging these two bounds in $(*)$ then simplifying yields the desired inequality. □

This work established the first connection between the vertex expansion of a graph and its spectral properties. While useful in constructing vertex expanders [Alo86], the dependence on Δ means that $\lambda_2'(G)$ is not the best proxy for $\psi(G)$. Compared to Cheeger's inequality for edge conductance that $\phi(G)^2/2 \leq \lambda_2(G) \leq 2\phi(G)$, there is an extra factor Δ between the upper and lower bounds.

3.2.2 Fastest Mixing Markov Chain

The fastest mixing Markov chain problem is introduced by Boyd, Diaconis, and Xiao [BDX04]. In the fastest mixing time problem, we are given an undirected graph $G = (V, E)$ and a target probability distribution $\pi : V \to \mathbb{R}^+$. The task is to find a time-reversible transition matrix $P \in \mathbb{R}^{n \times n}$ supported on the edges of G, so that the stationary distribution of the random walk with transition matrix P is π. The objective is to find such a transition matrix that minimizes the mixing time to the stationary distribution π. By Proposition 2.6.2, the mixing time to the stationary distribution is approximately inversely proportional to the spectral gap $1 - \alpha_2(P)$ of the time-reversible transition matrix P, where $1 = \alpha_1(P) \geq \alpha_2(P) \geq \cdots \geq \alpha_n(P) \geq -1$ are the eigenvalues of P. The fastest mixing time problem is thus formulated as follows in [BDX04], using the maximum spectral gap achievable through such a "reweighting" P of the input graph G as a proxy:

Definition 3.2.2 (Maximum Reweighted Spectral Gap [BDX04] (restatement of Definition 1.1.1)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted spectral gap is defined as*

$$\begin{aligned}
\lambda_2^*(G) \;:=\; \max_{P \geq 0} \quad & 1 - \alpha_2(P) \\
\text{subject to} \quad & P(u, v) = P(v, u) = 0 && \forall uv \notin E \\
& \sum_{v \in V} P(u, v) = 1 && \forall u \in V \\
& \pi(u)P(u, v) = \pi(v)P(v, u) && \forall uv \in E.
\end{aligned}$$

The graph is assumed to have a self-loop on each vertex, to ensure that the optimization problem for $\lambda_2^*(G)$ is always feasible.

In the context of Markov chains, this corresponds to allowing a nonzero holding probability on each vertex. The last constraint is the time reversible condition to ensure that the transition matrix P corresponds to random walks on an undirected graph (where the edge weight of uv is $\pi(u)P(u, v)$) and that the stationary distribution of P is π. Note that $\lambda_2^*(G) = \max_{P \geq 0}(1 - \alpha_2(P)) = \max_{P \geq 0} \lambda_2(I - P)$, which is the maximum reweighted second smallest eigenvalue of the normalized Laplacian matrix of G (where the edge weight of uv is $\pi(u)P(u, v)$) subject to the above constraints.

They observe that $\lambda_2^*(G)$ may be formulated as an SDP, and so is solvable in polynomial time. This is by writing the program for maximizing the spectral gap as

$$
\begin{aligned}
\min_{s,\mathcal{Q}} \quad & s \\
\text{subject to} \quad & -sI \preceq \mathcal{Q} - \sqrt{\pi}\sqrt{\pi}^T \preceq sI \\
& \mathcal{Q}(u,v) = \mathcal{Q}(v,u) = 0 & \forall uv \notin E \\
& \mathcal{Q}\sqrt{\pi} = \sqrt{\pi} \\
& \mathcal{Q} = \mathcal{Q}^T
\end{aligned}
$$

where $\sqrt{\pi}$ is the column vector whose u-th entry is $\sqrt{\pi(u)}$, $\Pi := \mathrm{diag}(\pi)$, and $\mathcal{Q} := \Pi^{1/2} P \Pi^{-1/2}$ is similar to P, hence having the same eigenvalues as P. Finally, this can be transformed into standard SDP form as

$$
X = \begin{pmatrix} \mathcal{Q} + sI - \sqrt{\pi}\sqrt{\pi}^T & 0 & 0 \\ 0 & -\mathcal{Q} + sI + \sqrt{\pi}\sqrt{\pi}^T & 0 \\ 0 & 0 & s \end{pmatrix}
$$

so that $X \succeq 0$ and the objective and other constraints are linear in X.

Upper Bound by Vertex Expansion

Roch noticed that $\lambda_2^*(G)$ is upper bounded by the vertex expansion of the graph. His main contribution was to provide the following dual characterization of the maximum reweighted spectral gap.

Proposition 3.2.3 (Dual Program for Fastest Mixing [Roc05]). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the following semidefinite program is dual to the primal program in Definition 1.1.1 with strong duality $\lambda_2^*(G) = \gamma(G)$ where*

$$
\begin{aligned}
\gamma(G) := \min_{f:V\to\mathbb{R}^n,\ g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} \pi(v) g(v) \\
\text{subject to} \quad & \sum_{v\in V} \pi(v) \|f(v)\|^2 = 1 \\
& \sum_{v\in V} \pi(v) f(v) = \vec{0} \\
& g(u) + g(v) \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E.
\end{aligned}
$$

We note that this is equivalent to the dual program given in [BDX04], but Roch's program is written in a vector program form that will be more convenient for rounding.

The following proof is inspired by [Roc05], but is presented slightly differently.

Proof. With the change of variables $\mathcal{Q} := \Pi^{1/2} P \Pi^{-1/2}$, and using the variational characterization of eigenvalues, we can write the objective $\max_{P \geq 0} 1 - \alpha_2(P)$ as

$$\max_{\mathcal{Q} \geq 0} \min_{\substack{z: V \to \mathbb{R} \\ z \perp \sqrt{\pi} \\ \|z\|_2 = 1}} z^T (I - \mathcal{Q}) z.$$

Define $Q := \Pi P = \Pi^{1/2} \mathcal{Q} \Pi^{1/2}$ and $Z := \Pi^{-1/2} z z^T \Pi^{-1/2}$ so that

$$\min_{\substack{z: V \to \mathbb{R} \\ z \perp \sqrt{\pi} \\ \|z\| = 1}} z^T (I - \mathcal{Q}) z = \min_{\substack{Z \succeq 0 \\ rank(Z) = 1 \\ Z\pi = \vec{0} \\ \langle Z, \Pi \rangle = 1}} \langle Z, \Pi - Q \rangle.$$

Since the objective and the constraints on Z (save for $Z \succeq 0$) are all linear, removing the rank-one constraint on Z does not affect the objective, and so we can rewrite the program in terms of Q and Z as

$$\begin{aligned}
\lambda_2^*(G) \;=\; & \max_{Q \geq 0} \min_{Z \succeq 0} && \langle Z, \Pi - Q \rangle \\
& \text{subject to} && Q(u,v) = Q(v,u) = 0 && \forall uv \notin E \\
& && Q\mathbb{1} = \pi, Q = Q^T \\
& && Z\pi = \vec{0}, \langle Z, \Pi \rangle = 1
\end{aligned}$$

The relaxation step is crucial to make the feasible region for Z convex. Since the feasible regions for Q and for Z are both compact and convex, and the objective function is linear, we may apply von Neumann minimax theorem in Theorem 2.8.1 to switch the max and min without changing the objective value, i.e.

$$\begin{aligned}
\lambda_2^*(G) \;=\; & \min_{Z \succeq 0} \max_{Q \geq 0} && \langle Z, \Pi - Q \rangle \\
& \text{subject to} && Q(u,v) = Q(v,u) = 0 && \forall uv \notin E \\
& && Q\mathbb{1} = \pi, Q = Q^T \\
& && Z\pi = \vec{0}, \langle Z, \Pi \rangle = 1
\end{aligned}$$

For fixed Z, the inner maximization problem is an LP, for which strong duality holds. To arrive at a clean formulation for the dual program, we write $\Pi(u,u) = \sum_v Q(u,v)$, identify $Q(u,v)$ and $Q(v,u)$ for $uv \in E$, eliminate $Q(u,v)$ for $uv \notin E$, and use $g(u)$ as dual variable for the constraint $\sum_u Q(u,v) = \pi(u)$. After taking dual of the inner LP, we arrive at the following min-min program:

$$\begin{aligned} \min_{Z \succeq 0} \min_{g: V \to \mathbb{R}_{\geq 0}} & \sum_{v \in V} \pi(v) g(v) \\ \text{subject to} \quad & g(u) + g(v) \geq Z(u,u) + Z(v,v) - Z(u,v) - Z(v,u) \qquad \forall uv \in E \\ & Z\pi = \vec{0}, \langle Z, \Pi \rangle = 1. \end{aligned}$$

Note that $g(u) \geq 0$ follows from the dual constraint for $Q(u,u)$, which corresponds to the self loop $u \to u$ added to the graph.

Finally, to write the program in vector form, since $Z \succeq 0$ is constrained to be PSD, there exists vectors $\{f(u)\}_{u \in v}$ in $\mathbb{R}^n$ such that $Z(u,v) = \langle f(u), f(v) \rangle$. Substituting this in the program above and rewriting the constraints

$$Z\pi = \vec{0}, \qquad \langle Z, \Pi \rangle = 1$$

as

$$\sum_{v \in V} \pi(v) f(v) = \vec{0}, \qquad \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1,$$

we have derived the desired dual formulation. □

Then, to prove the "easy direction" that $\lambda_2^*(G) \leq \psi(G)$, it suffices to construct, given a cut $S \subseteq V$, a dual solution (f,g) to $\gamma(G)$ whose objective value is upper bounded by $\psi(S)$. We will, however, prove the easy direction slightly differently in Proposition 3.2.8 in the following subsubsection.

First Cheeger Inequality

Roch's result can be interpreted as saying that a graph with small vertex expansion cannot be reweighted to have fast mixing time to the target distribution. Olesker-Taylor and Zanetti [OZ22] completed the picture by proving that small vertex expansion is the *only* obstacle to fastest mixing: a graph with good vertex expansion can always be reweighted to have fast mixing time. Their main result is a Cheeger inequality for vertex expansion (under uniform vertex weights).

Theorem 3.2.4 (Cheeger Inequality for Vertex Expansion [OZ22]). *For any undirected graph $G = (V, E)$ and the uniform distribution $\pi = \vec{1}/n$,*

$$\frac{\psi(G)^2}{\log n} \lesssim \lambda_2^*(G) \lesssim \psi(G).$$

In terms of the fastest mixing time $\tau_{\mathrm{mix}}^*(G)$ to the uniform distribution, $\frac{1}{\psi(G)} \lesssim \tau_{\mathrm{mix}}^*(G) \lesssim \frac{\log^2 n}{\psi^2(G)}$. (See Section 2.6 for definitions and results for random walks and mixing time.)

The starting point of their proof is the dual characterization of Proposition 3.2.3 obtained by Roch [Roc05]. The proof of Theorem 3.2.4 has two main steps. The first step is to project the above dual program to the following one-dimensional "spectral" program.

Definition 3.2.5 (One-Dimensional Dual Program for Fastest Mixing [OZ22]). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, $\gamma^{(1)}(G)$ is defined to be the program:*

$$\begin{aligned}
\gamma^{(1)}(G) := \min_{f:V\to\mathbb{R},\ g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} \pi(v) g(v) \\
\text{subject to} \quad & \sum_{v\in V} \pi(v) f(v)^2 = 1 \\
& \sum_{v\in V} \pi(v) f(v) = 0 \\
& g(u) + g(v) \geq (f(u) - f(v))^2 \qquad \forall uv \in E.
\end{aligned}$$

Olesker-Taylor and Zanetti use the Johnson-Lindenstrauss lemma in Lemma 2.10.8 to first project the solution in Proposition 3.2.3 to $O(\log n)$ dimensions with constant distortion, and then take the best coordinate to obtain a 1-dimensional solution with the following guarantee. Note that this step works for any probability distribution π on V.

Proposition 3.2.6 ([OZ22, Proposition 2.9]). *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\gamma(G) \leq \gamma^{(1)}(G) \lesssim \log n \cdot \gamma(G).$$

In the second step, Olesker-Taylor and Zanetti observed that the dual program in Definition 3.2.5 is similar to the weighted vertex cover problem with edge weights $(f(u) - f(v))^2$ for each edge $uv \in E$, which is equivalent to the fractional matching problem by linear programming duality. To analyze Definition 3.2.5, they introduced an interesting new concept called "matching conductance", and used some combinatorial arguments about greedy matching as well as some spectral arguments to prove the following Cheeger-type inequality.

Theorem 3.2.7 ([OZ22, Theorem 2.10])**.** *For any undirected graph $G = (V, E)$ and the uniform distribution $\pi = \vec{1}/n$,*

$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi(G).$$

Proposition 3.2.8 (Easy Direction)**.** $\gamma^{(1)}(G) \leq 2\psi(G)$.

Proof. Let $\pi = \vec{1}/n$. Given $S \subseteq V$ with $0 < \pi(S) \leq 1/2$ and $\psi(S) = \psi(G)$, consider the following solution to the $\gamma^{(1)}(G)$ program:

$$f(v) := \begin{cases} \frac{C}{\pi(S)}, & \text{if } v \in S; \\ \frac{-C}{\pi(S^c)}, & \text{if } v \notin S \end{cases}, \qquad g(v) := \begin{cases} \left(\frac{C}{\pi(S)} + \frac{C}{\pi(S^c)}\right)^2, & \text{if } v \in \partial S; \\ 0, & \text{otherwise,} \end{cases}$$

where $C \in \mathbb{R}$ is such that

$$\sum_{v \in V} \pi(v) f(v)^2 = C^2 \left(\frac{1}{\pi(S)} + \frac{1}{\pi(S^c)}\right) = 1.$$

One can readily check that (f, g) is feasible, and the objective value is

$$\sum_{v \in V} \pi(v) g(v) = \pi(\partial S) \cdot \left(\frac{C}{\pi(S)} + \frac{C}{\pi(S^c)}\right)^2 = \pi(\partial S) \cdot \left(\frac{1}{\pi(S)} + \frac{1}{\pi(S^c)}\right) \leq 2\psi(S).$$

□

This implies $\lambda_2^*(G) \lesssim \psi(G)$ by $\lambda_2^*(G) = \gamma(G) \leq \gamma^{(1)}(G) \lesssim \psi(G)$. Essentially the same construction gives a direct proof that $\gamma(G) \lesssim \psi(G)$.

We now turn our attention to the hard direction. As mentioned, they introduced the following concept called matching conductance.

Definition 3.2.9 (Matching Conductance)**.** *Let $G = (V, E, w)$ be an edge-weighted graph. The matching conductance of a set $S \subseteq V$ and of G are defined as*

$$\Upsilon(S) := \frac{\nu(E(S, S^c))}{|S|} \qquad \text{and} \qquad \Upsilon(G) := \min_{0 < |S| \leq n/2} \Upsilon(S),$$

where $\nu(F)$ is the maximum total weight of a matching on $F \subseteq E$.

Proposition 3.2.10 (Matching Conductance and Vertex Expansion)**.** *Let $G = (V, E)$ be a graph with unit edge and vertex weights. Then,* $\Upsilon(G) \le \psi(G) \le 4\Upsilon(G)$.

Proof. The first inequality $\Upsilon(G) \le \psi(G)$ follows from the fact that $\nu(E(S, S^c)) \le |\partial(S)|$ for any $S \subseteq V$. For the second inequality, let $S \subseteq V$ be such that $0 < |S| \le n/2$ and $\Upsilon(S) = \Upsilon(G) \le 1/4$. Let M be a maximum matching on $E(S, S^c)$ and $V(M)$ be the set of vertices incident to M. Then, $|M| \le |S|/4$. Consider $T := S \setminus V(M)$. Then $\partial T \subseteq V(M)$, and so $|\partial T| \le |V(M)| \le 2|M|$. We also have $|T| \ge |S| - |V(M)| = |S| - 2|M| \ge |S|/2$. Hence, we conclude that $\psi(G) \le \psi(T) \le 4|M|/|S| = 4\Upsilon(S) = 4\Upsilon(G)$. □

Proposition 3.2.11 (Hard Direction)**.** $\Upsilon(G)^2 \lesssim \gamma^{(1)}(G)$.

This in turn implies that $\psi(G)^2 \lesssim \gamma^{(1)}(G)$. The proof involves some combinatorial arguments about greedy matching as well as some spectral arguments, that we shall outline below. Note that this only holds under uniform vertex weights.

Proof outline. Let (f, g) be an optimal solution to the $\gamma^{(1)}(G)$ program. We refactor the proof into the same two-part structure as in Theorem 3.1.1, keeping many of the main proof ideas in [OZ22] in the threshold rounding step.

Step 1 (ℓ_2^2 to ℓ_1). In this step, we would like to define an ℓ_1 version of $\gamma^{(1)}(G)$ and relate the objective values of the two. This step holds for any distribution π.

We first go from vertex cover to fractional matching by considering the following equivalent min-max formulation of $\gamma^{(1)}(G)$. This is by LP duality.

$$\begin{aligned} \gamma^{(1)}(G) = \min_{f:V\to\mathbb{R}} \max_{A:E\to\mathbb{R}_{\ge 0}} \quad & \sum_{uv\in E} A(uv)\cdot (f(u)-f(v))^2 \\ \text{subject to} \quad & \sum_{v\in V} \pi(v) f(v)^2 = 1 \\ & \sum_{v\in V} \pi(v) f(v) = 0 \\ & \sum_{u:uv\in E} A(uv) \le \pi(v) \qquad \forall v \in V. \end{aligned}$$

We then define the following ℓ_1 program which we call $\eta(G)$:

$$\begin{aligned}
\eta(G) := \min_{h:V\to\mathbb{R}} \max_{A:E\to\mathbb{R}_{\geq 0}} \quad & \sum_{uv\in E} A(uv)\cdot|h(u)-h(v)| \\
\text{subject to} \quad & \sum_{v\in V} \pi(v)|h(v)| = 1 \\
& \max\Big\{\pi\left(\{v\in V : h(v)>0\}\right), \pi\left(\{v\in V : h(v)<0\}\right)\Big\} \leq \frac{1}{2} \\
& \sum_{u:uv\in E} A(uv) \leq \pi(v) \qquad \forall v\in V.
\end{aligned}$$

The second constraint states that 0 is a π-weighted median of h, which serves the role of the "balance" constraint $\sum_{v\in V}\pi(v)f(v)=0$. Our goal is to show that

$$\eta(G) \lesssim \sqrt{\gamma^{(1)}(G)}.$$

To this end, let (f, A) be an optimal solution to $\gamma^{(1)}(G)$. Construct $h : V \to \mathbb{R}$ from f exactly as in (3.1) in Theorem 3.1.1, and using similar arguments we can show that, for any feasible $A' : E \to \mathbb{R}_{\geq 0}$,

$$\sum_{uv\in E} A'(uv)\cdot|h(u)-h(v)| \leq \sqrt{\sum_{uv\in E} A(uv)(f(u)-f(v))^2 \cdot 2\sum_{v\in V}\pi(v)(f(v)-c)^2},$$

where we have also used the optimality of A with respect to f and the constraint on A'. Dividing throughout by $\sum_{v\in V}\pi(v)|h(v)|$, we obtain

$$\frac{\max_{A'}\sum_{uv\in E} A'(uv)\cdot|h(u)-h(v)|}{\sum_{v\in V}\pi(v)|h(v)|} \leq \sqrt{\frac{2\sum_{uv\in E} A(uv)(f(u)-f(v))^2}{\sum_{v\in V}\pi(v)f(v)^2}} = \sqrt{2\gamma^{(1)}(G)},$$

and we are done after scaling h so that the denominator of LHS is 1.

Step 2 (threshold rounding). In this step, given a feasible solution h to $\eta(G)$, we would like to show that threshold rounding yields a set S with small matching conductance. We restrict to $\pi = \vec{1}/n$, and the desired requirements on S are that $0 < |S| \leq n/2$ and

$$\Upsilon(S) \lesssim \max_A A(uv)\cdot|h(u)-h(v)| =: \eta_h,$$

where A is subject to the constraints in the $\eta(G)$ program. Let $t \in \mathbb{R}$ be a parameter and define $S_t \subseteq V$ as in (3.3) Note that $|S_t| \leq n/2$ always, and the "average" size of S_t is

$$\int_{-\infty}^{\infty} |S_t|\, dt = n\sum_{v\in V}\pi(v)|h(v)|.$$

It suffices to show that

$$\int_{-\infty}^{\infty} \nu(E(S_t, S_t^c))\, dt \lesssim n \cdot \eta_h.$$

The observation in [OZ22] is that, the weighted graphs G_h and $\overrightarrow{G_h}$ defined below have maximum matchings that relate well to the two quantities of interest:

- $G_h = (V, E, w)$ where $w(uv) = |h(u) - h(v)|$;
- $\overrightarrow{G_h} = (V, \overrightarrow{E}, w)$ where $\overrightarrow{E} := \{(u,v) : uv \in E, h(u) > h(v)\}$ and $w(u,v) = h(u) - h(v)$.

Proposition 3.2.12. *For $\pi = \vec{1}/n$, $\int_{-\infty}^{\infty} \nu(E(S_t, S_t^c))\, dt \le 2\nu(\overrightarrow{G_h}) \le 6\nu(G_h) \le 6n \cdot \eta_h$.*

We will sketch the proof here. The last inequality holds because the maximum matching on G_h is upper bounded by the maximum fractional matching on G_h which has value $n \cdot \eta_h$. To prove the second inequality, take the arcs from a maximum matching of $\overrightarrow{G_h}$. Sort them in decreasing order of weight, and add the corresponding arcs one by one, whenever possible, to form a matching of G_h. For each arc added, at most two other arcs are skipped, so this loses at most a factor of 3.
To prove the first inequality, they consider not the maximum matching in $\overrightarrow{G_h}$, but a fixed "greedy" matching $\overrightarrow{M}$, which is by sorting the arcs in $\overrightarrow{G_h}$ in decreasing order of weight and adding them one by one to $\overrightarrow{M}$ whenever possible. The key claim is that $|\overrightarrow{M} \cap E(S_t, S_t^c)|$[2] $\ge \nu(E(S_t, S_t^c))/2$ for any $t \in \mathbb{R}$, which implies the desired inequality after taking integral over $t \in \mathbb{R}$. To see the claim, let M_t be a maximum matching of $E(S_t, S_t^c)$. If an edge $uv \in M_t$ is not added to the greedy matching $\overrightarrow{M}$, then it must be because a prior edge in $\overrightarrow{M}$ is blocking uv, and we can show that indeed that the edge must be in $E(S_t, S_t^c)$ as well. If the edge is added to $\overrightarrow{M}$, then we say that it blocks itself. Therefore, each edge in M_t is blocked by some edge in $\overrightarrow{M} \cap E(S_t, S_t^c)$, and each edge in $\overrightarrow{M} \cap E(S_t, S_t^c)$ can block at most two edges in M_t. This completes the proof of the key claim and hence Proposition 3.2.12.

By Proposition 3.2.12 and the usual averaging argument, we are done. □

Combining Proposition 3.2.3 and Proposition 3.2.6 and Theorem 3.2.7 gives

$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \log n \cdot \gamma(G) = \log n \cdot \lambda_2^*(G) \quad \text{and} \quad \lambda_2^*(G) = \gamma(G) \le \gamma^{(1)}(G) \lesssim \psi(G),$$

[2] This makes sense because, by the definition of $\overrightarrow{G_h}$, at most one of uv and vu will be in $\overrightarrow{E}$ for any $u, v \in V$, so each arc in $\overrightarrow{M}$ uniquely corresponds to an edge in G_h.

proving Theorem 3.2.4.

Note that their proof of Theorem 3.2.7 only works when π is the uniform distribution. Olesker-Taylor and Zanetti discussed some difficulty in generalizing their combinatorial arguments to the weighted setting, and left it as an open question to prove the theorem for any probability distribution π.

3.2.3 Improved Analysis of Projection Step

Jain, Pham, and Vuong [JPV22] improved Proposition 3.2.6 to only an $O(\log \Delta)$ loss for the one-dimensional program $\gamma^{(1)}(G)$, where Δ is the maximum degree of G. Instead of using the Johnson-Lindenstrauss lemma that preserves ℓ_2 distance between all pairs of points, they use a new analysis tailored to the fastest mixing program. Their strategy is to show that the standard Gaussian projection to $O(\log \Delta)$ dimensions (formally defined in Definition 3.1.7 for higher-order Cheeger inequality) has constant distortion, and then choose the best coordinate as in [OZ22] which incurs an $O(\log \Delta)$ loss.

The following properties of the random projection algorithm will be useful.

Lemma 3.2.13 (Gaussian Properties [MMR19, JPV22])**.** *Let $G = (V, E)$ be a graph and $f : V \to \mathbb{R}^n$ be an embedding of the vertices in G. Let $\bar{f} : V \to \mathbb{R}^k$ be the randomly projected solution by applying Definition 3.1.7 to f. There exists a universal constant $c > 0$ that satisfies the following two properties:*

- *For all $u, v \in V$,*

$$\Pr_{\bar{f}}\left[\, \left\|\bar{f}(u) - \bar{f}(v)\right\| \notin e^{\pm\varepsilon} \left\|f(u) - f(v)\right\| \,\right] \le e^{-c\varepsilon^2 k}.$$

- *For all $u, v \in V$, let $\mathcal{E}_{u,v}$ be the event that $\left\|\bar{f}(u) - \bar{f}(v)\right\| \ge e^{\varepsilon} \left\|f(u) - f(v)\right\|$, then*

$$\mathbb{E}_{\bar{f}}\left[\mathbb{1}_{\mathcal{E}_{u,v}}\left(\frac{\left\|\bar{f}(u) - \bar{f}(v)\right\|^2}{\left\|f(u) - f(v)\right\|^2} - e^{2\varepsilon}\right)\right] \le e^{-c\varepsilon^2 k}.$$

They take the min-max form of the maximum reweighted spectral gap in Definition 1.1.1 that

$$\lambda_2^*(G) = \min_{f:V\to\mathbb{R}} \max_{P\ge 0} \sum_{uv\in E} \pi(u)P(u,v)\left\|f(u) - f(v)\right\|^2$$

where P is subject to the constraints of Definition 1.1.1. Regarding the inner maximization program as a maximum weighted fractional matching problem, they used the simple observation that the maximum matching has objective value at least $1/(\Delta+1)$ times the trivial upper bound of

$$\sum_{uv\in E} \min(\pi(u), \pi(v)) \cdot \|f(u) - f(v)\|^2$$

to obtain the following projection loss upper bound.

Theorem 3.2.14 (Dimension Reduction for Maximum Fractional Matching Program [JPV22])**.** *Let $G = (V, E)$ be a graph, and let $\pi : V \to \mathbb{R}^+$ be a distribution on V. Define the k-dimensional program $\lambda_2^{(k)}(G)$ as*

$$\begin{aligned}
\min_{f:V\to\mathbb{R}^k} \max_{P\geq 0} \quad & \sum_{uv\in E} \pi(u)P(u,v)\,\|f(u) - f(v)\|^2 \\
\text{subject to} \quad & \sum_{u\in V} \pi(u)f(u) = \vec{0}, \quad \sum_{u\in V} \pi(u)\,\|f(u)\|^2 = 1 \\
& P(u,v) = P(v,u) = 0 \qquad \forall uv \notin E \\
& \sum_{v\in V} P(u,v) = 1 \qquad \forall u \in V \\
& \pi(u)P(u,v) = \pi(v)P(v,u) \qquad \forall u \in V,
\end{aligned}$$

and let $\nu_f^{(k)}(G)$ be the inner maximization problem for a feasible $f : V \to \mathbb{R}^k$. Then, there exists a constant $C > 0$ such that

$$\lambda_2^{(C\log\Delta)}(G) \lesssim \lambda_2^{(n)}(G) = \lambda_2^*(G).$$

Proof. Let $f : V \to \mathbb{R}^n$ be an optimal embedding of G such that $\nu_f^{(n)}(G) = \lambda_2^*(G)$. Let $d = C\log\Delta$ and let $\bar{f} : V \to \mathbb{R}^d$ be obtained from f via Gaussian projection (as in Definition 3.1.7). We would like to use $\bar{f}$ as a solution to $\lambda_2^{(d)}(G)$.

First, note that $\bar{f}$ is obtained by applying a (random) linear operator to f, and so the $\sum_{v\in V} \pi(v)\cdot f(v) = \vec{0}$ constraint in the n-dimensional program is also satisfied by $\bar{f}$ in the d-dimensional program. But the normalization constraint $\sum_{v\in V} \pi(v)\cdot \left\|\bar{f}(v)\right\|^2 = 1$ may not be satisfied, and the objective value $\nu_{\bar{f}}^{(d)}(G) = \max_{P\geq 0} \sum_{uv\in E} \pi(u)P(u,v)\cdot \left\|\bar{f}(u) - \bar{f}(v)\right\|^2$ may be bigger than $\nu_f^{(n)}(G)$. Our plan is to prove that

$$\lambda_2^{(d)}(G) \leq \frac{\nu_{\bar{f}}^{(d)}(G)}{\sum_{v\in V} \pi(v)\left\|\bar{f}(v)\right\|^2} \lesssim \frac{\nu_f^{(n)}(G)}{\sum_{v\in V} \pi(v)\,\|f(v)\|^2} = \lambda_2^*(G), \tag{3.6}$$

and this would imply that a scaled version of $\bar{f}$ will satisfy the constraint for $\lambda_2^{(d)}(G)$ with objective value at most $O\big(\lambda_2^*(G)\big)$.

The main job is to upper bound $\nu_{\bar{f}}^{(d)}(G)$. Given $\bar{f} : V \to \mathbb{R}^d$, and let $\varepsilon > 0$, consider the set of "bad edges" $\mathcal{B} := \{uv \in E \mid \left\|\bar{f}(u) - \bar{f}(v)\right\|^2 \geq e^{2\varepsilon} \cdot \|f(u) - f(v)\|^2\}$ where the projected length is considerably longer than the original length. We can bound $\nu_{\bar{f}}^{(d)}(G)$ in terms of the edges in $\mathcal{B}$ as follows. For any feasible P to the $\nu_{\bar{f}}^{(d)}(G)$ program, its objective value is

$$
\begin{aligned}
& \sum_{uv \notin \mathcal{B}} \pi(u)P(u,v) \left\|\bar{f}(u) - \bar{f}(v)\right\|^2 + \sum_{uv \in \mathcal{B}} \pi(u)P(u,v) \left\|\bar{f}(u) - \bar{f}(v)\right\|^2 \\
= \; & \sum_{uv \notin \mathcal{B}} \pi(u)P(u,v) \left\|\bar{f}(u) - \bar{f}(v)\right\|^2 \\
& \quad + \sum_{uv \in \mathcal{B}} \pi(u)P(u,v)\Big(\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 + e^{2\varepsilon} \|f(u) - f(v)\|^2 \Big) \\
\leq \; & 2e^{2\varepsilon} \sum_{uv \in E} \pi(u)P(u,v) \|f(u) - f(v)\|^2 \\
& \quad + \sum_{uv \in \mathcal{B}} \pi(u)P(u,v)\Big(\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 \Big) \\
\leq \; & 2e^{2\varepsilon}\nu_f^{(n)}(G) + \sum_{uv \in \mathcal{B}} \min(\pi(u), \pi(v))\Big(\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 \Big),
\end{aligned}
$$

where the last inequality is because $P(u,v) \leq 1$ and $\pi(u)P(u,v) = \pi(v)P(v,u)$. Recall the definition in Lemma 3.2.13 that $\mathcal{E}_{u,v}$ is the event that $\left\|\bar{f}(u) - \bar{f}(v)\right\| \geq e^{\varepsilon} \|f(u) - f(v)\|$, which is equivalent to $uv \in \mathcal{B}$. Since the upper bound on the last line no longer depends on P, it follows that

$$
\begin{aligned}
& \mathbb{E}_{\bar{f}}[\nu_{\bar{f}}^{(d)}(G)] \\
\leq \; & 2e^{2\varepsilon}\nu_f^{(n)}(G) + \mathbb{E}_{\bar{f}}\Big[\sum_{uv \in \mathcal{B}} \min(\pi(u), \pi(v))\Big(\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 \Big)\Big] \\
= \; & 2e^{2\varepsilon}\nu_f^{(n)}(G) + \sum_{uv \in E} \min(\pi(u), \pi(v)) \cdot \mathbb{E}_{\bar{f}}\Big[\mathbb{1}_{\mathcal{E}_{u,v}}\Big(\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 \Big)\Big] \\
\leq \; & 2e^{2\varepsilon}\nu_f^{(n)}(G) + e^{-c\varepsilon^2 d} \sum_{uv \in E} \min(\pi(u), \pi(v)) \|f(u) - f(v)\|^2 \\
\leq \; & 2e^{2\varepsilon}\nu_f^{(n)}(G) + 2e^{-c\varepsilon^2 d} \cdot (\Delta + 1) \cdot \nu_f^{(n)}(G),
\end{aligned}
$$

where the second last inequality is by the second property in Lemma 3.2.13, and the last inequality is by the property that the value of $\nu_f^{(n)}(G)$ is at least the objective of the following "uniform" solution

$$P'(u,v) := \begin{cases} \frac{1}{\Delta+1}, & \text{if } u = v \text{ or } uv \in E \\ 0, & \text{otherwise,} \end{cases}$$

which in turn satisfies

$$\sum_{uv\in E} \pi(u)P'(u,v) \left\| f(u) - f(v) \right\|^2 \geq \frac{1}{\Delta+1} \sum_{uv\in E} \min(\pi(u), \pi(v)) \left\| f(u) - f(v) \right\|^2.$$

By choosing some constant $\varepsilon \leq 1/4$ and $d \gtrsim \frac{1}{c\varepsilon^2}\log(\Delta+1)$, it follows that

$$\mathbb{E}_{\bar{f}}[\nu_{\bar{f}}^{(d)}(G)] \leq 2\big(e^{2\varepsilon} + e^{-c\varepsilon^2 d}\Delta\big) \cdot \nu_f^{(n)}(G) \lesssim \nu_f^{(n)}(G).$$

Finally, we lower bound the denominator. Let $\mathcal{E}'_v$ be the event that $\left\|\bar{f}(v)\right\|^2 < e^{-2\varepsilon}\left\|f(v)\right\|^2$. Using a similar argument as above,

$$\sum_{v\in V} \pi(v) \cdot \left\|\bar{f}(v)\right\|^2 \geq e^{-2\varepsilon} \sum_{v\in V} \pi(v) \left\|f(v)\right\|^2 - \sum_{v\in V} \pi(v) \cdot \mathbb{1}_{\mathcal{E}'_v}\big(e^{-2\varepsilon}\left\|f(v)\right\|^2 - \left\|\bar{f}(v)\right\|^2\big).$$

We can view the event $\mathcal{E}'_v$ as $\mathcal{E}'_{v,0}$ where the zero vector is one of the embedding vectors, so that $\left\|\bar{f}(v)\right\|^2 < e^{-2\varepsilon}\left\|f(v)\right\|^2$ is equivalent to $\left\|\bar{f}(v) - \bar{f}(0)\right\|^2 < e^{-2\varepsilon}\left\|f(v) - f(0)\right\|^2$. Thus, we can apply the first property in Lemma 3.2.13 to bound $\mathbb{E}_h[\mathbb{1}_{\mathcal{E}'_v}] = \Pr[\mathbb{1}_{\mathcal{E}'_v}]$, so that

$$\begin{aligned} \mathbb{E}_{\bar{f}}\Big[\sum_{v\in V} \pi(v) \cdot \mathbb{1}_{\mathcal{E}'_v}\big(e^{-2\varepsilon}\left\|f(v)\right\|^2 - \left\|\bar{f}(v)\right\|^2\big)\Big] &\leq \sum_{v\in V} e^{-2\varepsilon}\pi(v)\left\|f(v)\right\|^2 \cdot \mathbb{E}_{\bar{f}}[\mathbb{1}_{\mathcal{E}'_v}] \\ &\leq e^{-c\varepsilon^2 d - 2\varepsilon} \sum_{v\in V} \pi(v)\left\|f(v)\right\|^2. \end{aligned}$$

By Markov's inequality and the same choice of ε and k, with probability at least 9/10,

$$\sum_{v\in V} \pi(v) \cdot \left\|\bar{f}(v)\right\|^2 \geq e^{-2\varepsilon}\big(1 - 10e^{-c\varepsilon^2 d}\big) \sum_{v\in V} \pi(v)\left\|f(v)\right\|^2 \gtrsim \sum_{v\in V} \pi(v)\left\|f(v)\right\|^2.$$

Therefore, (3.6) follows by combining the upper bound on the numerator and this lower bound on the denominator. □

3.2.4 λ_∞ and SDP Relaxation

Bobkov, Houdré, and Tetali [BHT00] defined a spectral quantity called λ_∞:

$$\lambda_\infty(G) := \min_{\substack{f:V\to\mathbb{R}\\ \sum_v f(v)=0}} \frac{\sum_{v\in V} \max_{u:uv\in E}(f(u)-f(v))^2}{\sum_{v\in V} f(v)^2}$$

Observe that λ_2 can be written in a similar form to λ_∞ as follows:

$$\lambda_2(G) = \frac{1}{2} \min_{\substack{f:V\to\mathbb{R}\\ \sum_v f(v)=0}} \frac{\sum_{v\in V} \sum_{u:uv\in E}(f(u)-f(v))^2}{\sum_{v\in V} f(v)^2},$$

which explains the notation λ_∞ as the sum is replaced by the maximum, an "ℓ_∞" quantity. λ_∞ satisfies an exact analogue of Cheeger's inequality for *symmetric* vertex expansion:

Theorem 3.2.15 ([BHT00])**.** *For any graph $G = (V, E)$,*

$$\frac{1}{2}\lambda_\infty(G) \le \psi_{sym}(G) \lesssim \sqrt{\lambda_\infty(G)},$$

where $\psi_{sym}(G)$ is defined as in Definition 2.3.5 with $\pi = \frac{1}{n}\vec{1}$.

Proof. For the easy direction, given $S \subseteq V$ with $0 < |S| \le n/2$ that minimizes $\psi_{sym}(S)$, define the following vector solution $f : V \to \mathbb{R}$ to $\lambda_\infty(G)$:

$$f(v) := \begin{cases} \frac{1}{|S|}, & \text{if } v \in S \\ \frac{-1}{|S^c|}, & \text{if } v \notin S. \end{cases}$$

Then, $\sum_{v\in V} f(v) = 0$, and so

$$\begin{aligned} \lambda_\infty(G) &\le \frac{\sum_{v\in V} \max_{u:uv\in E}(f(u)-f(v))^2}{\sum_{v\in V} f(v)^2} \\ &= \frac{|\partial_{sym}(S)| \cdot (1/|S| + 1/|S^c|)^2}{1/|S| + 1/|S^c|} \le 2 \cdot \frac{|\partial_{sym}(S)|}{|S|} = 2\psi_{sym}(G). \end{aligned}$$

Now we prove the hard direction, which we redo using the two-step framework.

Step 1 (ℓ_2^2 to ℓ_1). Given optimal solution $f : V \to \mathbb{R}$ to $\lambda_\infty(G)$, we would like to construct $h : V \to \mathbb{R}$ such that 0 is a median of h, and

$$\frac{\sum_{v\in V} \max_{u:uv\in E} |h(u)-h(v)|}{\sum_{v\in V} |h(v)|} \lesssim \sqrt{\frac{\sum_{v\in V} \max_{u:uv\in E}(f(u)-f(v))^2}{\sum_{v\in V} f(v)^2}}$$

The construction of h is exactly the same as in (3.1) in Theorem 3.1.1, with c this time being a simple median of f:

$$h(u) := \begin{cases} (f(u) - c)^2, & \text{if } f(u) \geq c, \\ -(f(u) - c)^2, & \text{if } f(u) < c. \end{cases}$$

We again have

$$\sum_{v \in V} |h(v)| = \sum_{v \in V} (f(v) - c)^2 \geq \sum_{v \in V} f(v)^2 \tag{3.7}$$

in the denominator, and applying (3.2) from Theorem 3.1.1 to the numerator we get

$$\begin{aligned} & \sum_{v \in V} \max_{u:uv \in E} |h(u) - h(v)| \\ \leq\ & \sum_{v \in V} \max_{u:uv \in E} |f(u) - f(v)| \left(|f(u) - c| + |f(v) - c|\right) \\ \leq\ & \sum_{v \in V} \max_{u:uv \in E} |f(u) - f(v)| \left(2|f(v) - c| + |f(u) - f(v)|\right) \\ \leq\ & \sum_{v \in V} \max_{u:uv \in E} (f(u) - f(v))^2 + 2\sqrt{\sum_{v \in V} \max_{u:uv \in E} (f(u) - f(v))^2 \cdot \sum_{v \in V} (f(v) - c)^2}, \end{aligned}$$

where we used Cauchy-Schwarz inequality in the last step. Dividing both sides by $\sum_{v \in V} |h(v)|$ and using (3.7), we get

$$\begin{aligned} & \frac{\sum_{v \in V} \max_{u:uv \in E} |h(u) - h(v)|}{\sum_{v \in V} |h(v)|} \\ \leq\ & \frac{\sum_{v \in V} \max_{u:uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} f(v)^2} + 2\sqrt{\frac{\sum_{v \in V} \max_{u:uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} f(v)^2}} \\ \lesssim\ & \sqrt{\frac{\sum_{v \in V} \max_{u:uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} f(v)^2}}, \end{aligned}$$

where the asymptotic inequality uses the easy direction that $\lambda_\infty(G) \lesssim \psi_{sym}(G)$, which in turn is $\leq O(1)$.

Step 2 (threshold rounding). Given h, we would like to apply threshold rounding to extract a set $S \subseteq V$ with $0 < |S| \leq n/2$ and $\psi_{sym}(S)$ small. Let $t \in \mathbb{R}$ be a parameter and

$S_t \subseteq V$ defined exactly as in (3.3) in Theorem 3.1.1. Note that $|S_t| \leq n/2$ for any $t \in \mathbb{R}$ since 0 is a median of h. The "average" size of S_t is

$$\int_{-\infty}^{\infty} |S_t|\, dt = \sum_{v \in V} \int_{-\infty}^{\infty} \mathbb{1}[v \in S_t]\, dt = \sum_{v \in V} |h(v)|,$$

and the "average" size of $\partial_{sym}(S_t)$ is

$$\begin{aligned} \int_{-\infty}^{\infty} |\partial_{sym}(S_t)|\, dt = \sum_{v \in V} \int_{-\infty}^{\infty} \mathbb{1}[v \in \partial_{sym}(S_t)]\, dt &\overset{(*)}{=} \sum_{v \in V} \max_{u,u' \in \partial(V)} |h(u) - h(u')| \\ &\leq 2 \sum_{v \in V} \max_{u: uv \in E} |h(u) - h(v)|. \end{aligned}$$

The step $(*)$ is due to the observation that $v \in \partial_{sym}(S_t)$ if and only if $\min_{u: uv \in E} h(u) \leq t \leq \max_{u: uv \in E} h(u)$. Note that $|\partial_{sym}(S_t)| = 0$ when $|S_t| = 0$. Hence, there exists some $t \in \mathbb{R}$ such that $0 < |S_t| \leq n/2$ and

$$\psi_{sym}(S_t) \leq \frac{2 \sum_{v \in V} \max_{u: uv \in E} |h(u) - h(v)|}{\sum_{v \in V} |h(v)|}.$$

This completes the proof of the hard direction and hence the theorem. □

While Theorem 3.2.15 has the exact same form as the classical Cheeger's inequality, it is not known how to compute λ_∞ efficiently, and Farhadi, Louis, and Tetali [FLT20] recently showed that λ_∞ is NP-hard to compute. Therefore, λ_∞ alone does not yield an algorithmic spectral theory. To design an approximation algorithm for $\psi(G)$, Louis, Raghavendra and Vempala [LRV13] defined the following semidefinite programming relaxation for λ_∞, denoted by sdp_∞.

Definition 3.2.16 (sdp_∞ in [LRV13])**.** *Given an undirected graph $G = (V, E)$,*

$$\begin{aligned} \mathsf{sdp}_\infty(G) := \min_{f: V \to \mathbb{R}^n,\ g: V \to \mathbb{R}} \quad & \sum_{v \in V} g(v) \\ \text{subject to} \quad & \sum_{v \in V} \|f(v)\|^2 = 1 \\ & \sum_{v \in V} f(v) = \vec{0} \\ & g(v) \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E \end{aligned}$$

Observe that λ_∞ is the 1-dimensional version of sdp_∞, with $f : V \to \mathbb{R}^n$ replaced by $f : V \to \mathbb{R}$. Therefore, $\mathsf{sdp}_\infty(G)$ is a relaxation of $\lambda_\infty(G)$. To see that sdp_∞ is indeed an SDP, note that it can be rewritten as

$$\begin{aligned}
\mathsf{sdp}_\infty(G) \;=\; &\min_{\substack{Z \succeq 0 \\ g:V\to\mathbb{R}}} && \sum_{v\in V} g(v) \\
&\text{subject to} && g(v) \geq Z(u,u) + Z(v,v) - Z(u,v) - Z(v,u) \quad \forall uv \in E \\
& && Z\mathbb{1} = \vec{0}, \langle Z, I\rangle = 1
\end{aligned}$$

where Z is the Gram matrix of f so that $Z(u,v) = \langle f(u), f(v)\rangle$. The program implicitly requires that $g \geq 0$, so that

$$X := \begin{pmatrix} Z & 0 \\ 0 & \operatorname{diag}(g) \end{pmatrix}$$

is PSD, and both the objective and the constraints of $\mathsf{sdp}_\infty(G)$ are linear in X.

The rounding algorithm in [LRV13] is to project the solution to sdp_∞ into a 1-dimensional solution by setting $x(v) = \langle f(v), z\rangle$ where $z \sim N(0,1)^n$ is a random Gaussian vector. They proved that the 1-dimensional solution is a $O(\log\Delta)$-approximation to sdp_∞ where Δ is the maximum degree of the graph. For the analysis, they used Fact 2.10.6 and Fact 2.10.7 about Gaussian random variables. The first fact is for the analysis of the numerator and the second fact is for the analysis of the denominator of λ_∞.

Theorem 3.2.17 ([LRV13, Lemma 9.6])**.** *For any undirected graph $G = (V,E)$ with maximum degree Δ,*

$$\mathsf{sdp}_\infty(G) \leq \lambda_\infty(G) \lesssim \mathsf{sdp}_\infty(G) \cdot \log\Delta.$$

Proof. We have already observed that $\mathsf{sdp}_\infty(G)$ is a relaxation of $\lambda_\infty(G)$, and therefore $\mathsf{sdp}_\infty(G) \leq \lambda_\infty(G)$. For the other direction, given an optimal solution f to $\mathsf{sdp}_\infty(G)$ (the corresponding optimal choice for $g(v)$ would be $\max_{u:uv\in E} \|f(u) - f(v)\|^2$), let

$$x(v) = \langle f(v), z\rangle$$

where $z \sim N(0,1)^n$ is a random Gaussian vector. We have

$$\sum_{v\in V} x(v) = \left\langle \sum_{v\in V} f(v), z \right\rangle = 0,$$

so x is a feasible solution to $\lambda_\infty(G)$. To analyze the numerator, apply Fact 2.10.6 with $Y_u := x(u) - x(v)$, where u runs over the at most Δ neighbors of $v \in V$. Each Y_u is a centered Gaussian with variance $\|f(u) - f(v)\|^2 \leq g(v)$, and therefore

$$\mathbb{E}\left[\max_{u:uv\in E}(x(u) - x(v))^2\right] \leq 4g(v) \cdot \log \Delta.$$

Summing over $v \in V$ and applying Markov's inequality gives

$$\sum_{v\in V} \max_{u:uv\in E}(x(u) - x(v))^2 \leq 24 \cdot 4 \sum_{v\in V} g(v) \cdot \log \Delta = 96\,\mathsf{sdp}_\infty(G) \cdot \log \Delta$$

with probability at least $23/24$.

To analyze the denominator, take $Y_v := x(v)$ in Fact 2.10.7 for $v \in V$. Then,

$$\mathbb{E}\left[\sum_{v\in V} Y_v^2\right] = \mathbb{E}\left[\sum_{v\in V} f(v)^T zz^T f(v)\right] = \sum_{v\in V} \|f(v)\|^2 = 1,$$

and so with probability at least $1/12$, $\sum_{v\in V} x(v)^2 \geq \frac{1}{2}$. By union bound, then, with probability at least $\frac{1}{24}$ (in particular $\geq \Omega(1)$), both the numerator and denominator bounds hold simultaneously. Hence we see that the objective value for $\lambda_\infty(G)$ is at most

$$192\,\mathsf{sdp}_\infty(G) \cdot \log \Delta,$$

proving the theorem. □

Combining Theorem 3.2.17 with Theorem 3.2.15 immediately yields the following Cheeger inequality:

Theorem 3.2.18 ([LRV13])**.** *For any graph $G = (V, E)$ with maximum degree Δ,*

$$\frac{\psi_{sym}(G)^2}{\log \Delta} \lesssim \mathsf{sdp}_\infty(G) \lesssim \psi_{sym}(G).$$

Then, by constructing a graph G' such that $\psi_{sym}(G') = \Theta(\psi(G))$, they reduce vertex expansion to symmetric vertex expansion and obtain a Cheeger-type inequality for $\psi(G)$ with the same guarantee. See Proposition 2.3.6 for the construction. They also show that it is impossible to improve on the $\log \Delta$ factor under the small-set expansion hypothesis (c.f. Section 3.1.5).

Finally, we remark that all the results discussed in this subsection apply to graphs equipped with any vertex measure $\pi : V \to \mathbb{R}^+$, by modifying the definitions accordingly.

3.2.5 Reduction from Hypergraph Spectral Theory

Louis [Lou15] and Chan, Louis, Tang, Zhang [CLTZ18] developed a spectral theory for hypergraphs. Their main results are surveyed in Section 3.4.1 in a later section. Their theory may be applied to obtain a spectral theory for vertex expansion via a reduction.

Fact 3.2.19 (Reducing Vertex Expansion to Hypergraph Expansion [CLTZ18, LM14b])**.** *Given a graph $G = (V, E)$, construct a hypergraph $H = (V, E')$ by adding the hyperedge $\{v\} \cup \partial(\{v\})$ for each $v \in V$ to E'. If the maximum degree of G is $\Delta_{\max} = \Delta$ and the minimum degree $\Delta_{\min}$, then the rank of H is at most $\Delta_{\max} + 1$, and*

$$\Delta_{\min} \cdot \phi_H(S) \leq \frac{\Delta_{\max}}{\Delta_{\max} + 1} \cdot \psi(S) \leq \Delta_{\max} \cdot \phi_H(S) \qquad \forall S \subset V.$$

Since this relation holds for all subsets S, their results on small-set hypergraph expansion and higher order hypergraph expansion translate to the vertex setting as well. In Section 4.1, we will compare in more detail our approach of reweighted eigenvalues to this approach of reducing directly from hypergraph spectral theory.

3.3 Directed Graphs

3.3.1 Cheeger Constant for Directed Graphs

One obstacle to developing a spectral theory for directed graphs is that the Laplacian may not have real eigenvalues. Fill [Fil91] and Chung [Chu05] provided a workaround, shifting attention to a carefully weighted version G_π of G and relating expansion properties of G_π to the spectrum of the symmetrized Laplacian $(L(G_\pi) + L(G_\pi^T))/2$, which must have real eigenvalues. While the two formulations are similar, we will follow Chung's formulation as it is closer and more consistent with ours, and also because her work is based on an Eulerian reweighted subgraph (which was called a circulation in [Chu05]) that we describe below, which is the main theme for Chapter 5.

Let $G = (V, E, w)$ be a weighted directed graph with random walk matrix P (see Section 2.6.1). Suppose that G is strongly connected and aperiodic. While P is not necessarily time-reversible, a unique stationary distribution $\pi : V \to \mathbb{R}^+$ nevertheless exists such that $\pi^T P = \pi^T$ (see Proposition 2.6.1). Let $\Pi := \operatorname{diag}(\pi)$. Fill [Fil91] defined the product and the sum matrices as $M(P) := P\Pi^{-1}P^T\Pi$ and $A(P) := (P + \Pi^{-1}P^T\Pi)/2$. Chung [Chu05] noted that if the weight of an arc uv is defined as $F(u, v) = \pi(u)P(u, v)$, then the weighted directed graph $G_\pi = (V, E, F)$ is Eulerian such that $\sum_{u:uv\in E} F(u,v) = \sum_{u:vu\in E} F(v,u)$ for all $v \in V$.[3] Then she used the underlying weighted undirected graph to define the Laplacian of the directed graph G as

$$\tilde{\mathcal{L}}(G) = I - \left(\Pi^{1/2} P \Pi^{-1/2} + \Pi^{-1/2} P^T \Pi^{1/2}\right)/2 = I - \Pi^{-\frac{1}{2}}\left(A_\pi + A_\pi^T\right)\Pi^{-\frac{1}{2}}/2. \tag{3.8}$$

where $A_\pi = \Pi P$ is the adjacency matrix of G_π. Note that the spectra of $I - A(P)$ and $\tilde{\mathcal{L}}(G)$ are the same, as $P + \Pi^{-1}P^T\Pi$ and $\Pi^{1/2}P\Pi^{-1/2} + \Pi^{-1/2}P^T\Pi^{1/2}$ are similar matrices. The Cheeger constant of a directed graph [Fil91, Chu05] is defined as

$$h(G) := \min_{S:S\neq\emptyset, S\neq V} h(S) \quad \text{where} \quad h(S) = \frac{\sum_{u,v:u\in S, v\notin S} \pi(u)P(u,v)}{\min\{\pi(S), \pi(S^c)\}}, \tag{3.9}$$

and Chung [Chu05] proved a Cheeger inequality relating $h(G)$ to the second smallest eigenvalue of $\tilde{\mathcal{L}}(G)$.

Theorem 3.3.1 (Cheeger Inequality for $h(G)$ [Fil91, Chu05])**.** *Let $G = (V, E, w)$ be a strongly connected, aperiodic directed graph, with stationary distribution $\pi : V \to \mathbb{R}^+$. Let $\tilde{\mathcal{L}}(G)$ be as defined in (3.8) and $h(G)$ be as defined in (3.9). Then,*

$$\lambda_2(\tilde{\mathcal{L}}(G))/2 \leq h(G) \leq \sqrt{2 \cdot \lambda_2(\tilde{\mathcal{L}}(G))}.$$

As in the undirected case (see Section 3.1.1), this result is supplemented with a mixing time bound using $\lambda_2(\tilde{\mathcal{L}}(G))$ to relate $h(G)$ to mixing time: random walk on G mixes fast if $h(G)$ is big. We state the result using Chung's formulation. Note that if G is strongly connected but not aperiodic, a unique stationary distribution still exists for the *lazy* random walk $(I + P)/2$ where P is the transition matrix of the ordinary random walk on G, and we can apply Theorem 3.3.1 to the graph corresponding to $(I + P)/2$.

Theorem 3.3.2 (Bounding Mixing Time by Second Eigenvalue of Directed Graphs [Fil91, Chu05])**.** *Let $G = (V, E, w)$ be a directed graph with positive edge weights. Suppose that G is strongly connected. Let P' be the transition matrix of the ordinary random walk on G with $P'(u,v) = w(uv)/\sum_{v\in V} w(uv)$ for $uv \in E$. Then the mixing time of the lazy random walks of G with transition matrix $P = (I + P')/2$ to the stationary distribution π is*

$$\tau_{\text{mix}}\left(\frac{I + P'}{2}\right) \lesssim \frac{1}{\lambda_2(\tilde{\mathcal{L}}(G))} \cdot \log\left(\frac{1}{\pi_{\min}}\right)$$

where $\pi_{\min} = \min_{v\in V} \pi(v)$.

[3] The proof is by noting that $\sum_{u:uv\in E} F(u,v) = (\pi^T P)(v) = \pi(v) = \sum_{u:vu\in E} F(v,u)$.

We note that Theorem 3.3.1 is algorithmic, since the stationary distribution π can be computed in polynomial time and the proof of the hard direction again shows that the sweep-cut algorithm is applicable here. Below we prove Theorem 3.3.1 and Theorem 3.3.2.

Proof of Theorem 3.3.1. First, verify that $\lambda_2(\tilde{\mathcal{L}}(G))$ has a similar variational characterization to Proposition 2.5.2:

$$\lambda_2(\tilde{\mathcal{L}}(G)) = \min_{f: f \perp \pi} \frac{\sum_{uv \in E} F(u,v)/2 \cdot (f(u) - f(v))^2}{\sum_{v \in V} \pi(v) f(v)^2}. \tag{3.10}$$

For the easy direction that $\lambda_2(\tilde{\mathcal{L}}(G))/2 \leq h(G)$, given $\emptyset \neq S \subset V$ with $h(S) = h(G)$, we define $f : V \to$ so that

$$f(v) := \begin{cases} \frac{1}{\pi(S)}, & \text{if } v \in S; \\ \frac{-1}{\pi(S^c)}, & \text{otherwise.} \end{cases}$$

Then, $f \perp \pi$, and

$$\begin{aligned} \lambda_2(\tilde{\mathcal{L}}(G)) &\leq \frac{\sum_{uv \in E} F(u,v)/2 \cdot (f(u) - f(v))^2}{\sum_{v \in V} \pi(v) f(v)^2} \\ &= \frac{\sum_{u,v: u \in S, v \notin S} \pi(u) P(u,v) + \sum_{u,v: u \notin S, v \in S} \pi(u) P(u,v)}{2} \cdot \left(\frac{1}{\pi(S)} + \frac{1}{\pi(S^c)} \right) \\ &= \sum_{u,v: u \in S, v \notin S} \pi(u) P(u,v) \cdot \left(\frac{1}{\pi(S)} + \frac{1}{\pi(S^c)} \right) \leq 2h(S), \end{aligned}$$

where the last equality is because G_π is Eulerian.

For the hard direction that $h(G) \leq \sqrt{2 \cdot \lambda_2(\tilde{\mathcal{L}}(G))}$, we again follow the same two-step approach of Theorem 3.1.1. Given $f : V \to \mathbb{R}$ with $f \perp \pi$. The first step of reducing to an ℓ_1 solution $g : V \to \mathbb{R}$ such that

$$\frac{\sum_{uv \in E} F(u,v)/2 \cdot |g(u) - g(v)|}{\sum_{v \in V} \pi(v) |g(v)|} \leq \sqrt{2 \cdot \frac{\sum_{uv \in E} F(u,v)/2 \cdot (f(u) - f(v))^2}{\sum_{v \in V} \pi(v) f(v)^2}}$$

and 0 is a π-weighted median of g, is essentially the same and is omitted. For the second step of threshold rounding, let $t \in \mathbb{R}$ be a parameter and define $S_t := \{v \in V : g(v) > t\}$. Since 0 is a π-weighted median of g, the "average" denominator of $h(S_t)$ amounts to

$$\int_{-\infty}^{\infty} \min\{\pi(S_t), \pi(S_t^c)\}\, dt = \int_0^{\infty} \pi(S_t)\, dt + \int_{-\infty}^{0} \pi(S_t^c)\, dt = \sum_{v \in V} \pi(v) |g(v)|.$$

The "average" numerator of $h(S_t)$ can be bounded as follows:

$$\begin{aligned}\int_{-\infty}^{\infty}\sum_{u,v:u\in S_t,v\notin S_t}\pi(u)P(u,v)\,dt &= \int_{-\infty}^{\infty}\frac{1}{2}\left(\sum_{u,v:u\in S_t,v\notin S_t}F(u,v)+\sum_{u,v:u\notin S_t,v\in S_t}F(u,v)\right)dt\\ &= \frac{1}{2}\sum_{uv\in E}F(u,v)\int_{-\infty}^{\infty}\mathbb{1}[(u,v)\in\delta(S_t)]\,dt\\ &= \frac{1}{2}\sum_{uv\in E}F(u,v)\cdot|g(u)-g(v)|.\end{aligned}$$

where the first equality is again because G_π is Eulerian. Note that the numerator of $h(S_t)$ is zero when $S_t=\emptyset$ or $S_t=V$. Therefore, threshold rounding yields an $S=S_t$ such that $S\neq\emptyset$, $S\neq V$, and

$$h(S)\leq\frac{\sum_{uv\in E}F(u,v)/2\cdot|g(u)-g(v)|}{\sum_{v\in V}\pi(v)|g(v)|}\leq\sqrt{2\cdot\frac{\sum_{uv\in E}F(u,v)/2\cdot(f(u)-f(v))^2}{\sum_{v\in V}\pi(v)f(v)^2}}.$$

Taking f to be optimal for $\lambda_2(\tilde{\mathcal{L}}(G))$ in (3.10) yields the hard direction. □

Proof Outline of Theorem 3.3.2. In the case where the graph is undirected, the Markov chain is reversible, and as we can see in the proof of Proposition 2.6.2, we can use the spectral decomposition of the random walk matrix, which gives precise control of the rate of decay of each direction orthogonal to the first eigenspace. Here, we cannot do the same. Instead, we employ the following result in [Chu05] which bounds the length of a vector after applying the random walk matrix, using $\lambda_2(\tilde{\mathcal{L}}(G))$.

Lemma 3.3.3 ([Chu05, Theorem 6]). *Under the setting of Theorem 3.3.2, let* $M:=\Pi^{1/2}((I+P')/2)\Pi^{-1/2}$. *Then,*

$$\|fM\|^2\leq\left(1-\frac{1}{2}\lambda_2(\tilde{\mathcal{L}}(G))\right)\|f\|^2$$

for all vectors $f:V\to\mathbb{R}$ *such that* $f\perp\Pi^{1/2}\mathbb{1}$.

The second eigenvalue $\lambda_2(\tilde{\mathcal{L}}(G))$ features here due to the appearance of the sum matrix $\tilde{P}+\tilde{P}^T$ in the expansion of $\|fM\|^2$ where $\tilde{P}:=\Pi^{1/2}P'\Pi^{-1/2}$, and $\tilde{\mathcal{L}}(G)$ is exactly the Laplacian of the undirected graph $(\tilde{P}+\tilde{P}^T)/2$.

Using this lemma, we can now finish the proof. Given any starting distribution p_0 we can decompose it as

$$p_0=\pi+f_\perp$$

where $\langle f_\perp, \mathbb{1}\rangle = (p_0 - \pi)^T \mathbb{1} = 1 - 1 = 0$. Then, writing λ_2 for $\lambda_2(\tilde{\mathcal{L}}(G))$,

$$\begin{aligned} d_{TV}(\pi, (P^t)^T p_0) &= \frac{1}{2}\left\|\pi - (P^t)^T p_0\right\|_1 \\ &= \left\|f_\perp^T \Pi^{-1/2}(M^t)\Pi^{1/2}\right\|_1 \\ &\leq \left\|f_\perp^T \Pi^{-1/2}(M^t)\right\| \cdot \left\|\Pi^{1/2}\mathbb{1}\right\| \quad \text{(Cauchy-Schwarz)} \\ &\leq \left\|f_\perp^T \Pi^{-1/2}\right\| \cdot \left(1 - \frac{\lambda_2}{2}\right)^t \quad \text{(by Lemma 3.3.3 since } f_\perp^T \Pi^{-1/2} \perp \Pi^{1/2}\mathbb{1}) \\ &\lesssim \frac{1}{\sqrt{\pi_{\min}}} \cdot \left(1 - \frac{\lambda_2}{2}\right)^t, \end{aligned}$$

where the last asymptotic inequality uses the fact that $\|f_\perp\| \leq \|f_\perp\|_1 \leq \|p_0\|_1 + \|\pi\|_1 \leq 2$. Taking $t \geq \Omega(\lambda_2^{-1} \cdot \log(1/\pi_{\min}))$, we have that $d_{TV}(\pi, (P^t)^T p_0) \leq 1/e$, and so the desired mixing time bound follows. □

Higher-order Cheeger Inequality for Directed Graphs

Chan, Tang and Zhang [CTZ15] gave a higher-order Cheeger inequality for directed graphs. Roughly speaking, they showed that in a directed graph $G = (V, E)$, there are k disjoint subsets $S_1, \ldots, S_k \subseteq V$ with $\lambda_k(\tilde{\mathcal{L}}(G)) \lesssim h(S_i) \lesssim k^2 \cdot \sqrt{\lambda_k(\tilde{\mathcal{L}}(G))}$ for $1 \leq i \leq k$, where $h(S_i)$ is the Cheeger constant in (3.9) and $\lambda_k(\tilde{\mathcal{L}}(G))$ is the k-th smallest eigenvalue of the Laplacian in (3.8).[4] The proof is a direct application of the higher-order Cheeger inequality [LOT12, LRTV12] in Section 3.1.3 on the reweighted subgraph $G' := (G_\pi + G_\pi^T)/2$ by the stationary distribution. This gives disjoint subsets $S_1, \ldots, S_k \subseteq V$ such that

$$\lambda_k(\tilde{\mathcal{L}}(G)) \lesssim \phi_{G'}(S_i) \lesssim k^2 \cdot \sqrt{\lambda_k(\tilde{\mathcal{L}}(G))},$$

where we note that $\tilde{\mathcal{L}}(G)$ is the ordinary graph Laplacian of G'. Finally,

$$\phi_{G'}(S) = \frac{\sum_{v \in S, u \notin S}(F(u,v) + F(v,u))/2}{\sum_{v \in S, u \in V}(F(u,v) + F(v,u))/2} = \frac{\sum_{v \in S, u \notin S} F(v,u)}{\pi(S)} = h(S)$$

for any nonempty subset $S \subseteq V$ with $\pi(S) \leq 1/2$.

[4] The main technical contribution of [CTZ15] is to apply the results of [Fil91, Chu05] when the graph is not strongly connected, which we do not discuss in depth here.

3.3.2 Nonlinear Laplacian and SDP Relaxation

Yoshida [Yos16] introduced a nonlinear Laplacian operator for directed graphs and used it to define the following second eigenvalue

$$\lambda_G = \inf_{x \perp \mu_G} \frac{\sum_{uv \in E} \left(\left[x(u)/\sqrt{\deg(u)} - x(v)/\sqrt{\deg(v)} \right]^+ \right)^2}{\sum_{u \in V} x(u)^2}$$

where μ_G denotes the first (trivial) eigenvector, $[a-b]^+$ denotes $\max\{a-b, 0\}$, and $\deg(u)$ is the total degree of u.[5]

He considered the directed edge conductance $\vec{\phi}$ as defined in Definition 2.3.1 and proved following Cheeger inequality for $\vec{\phi}(G)$:

Theorem 3.3.4 ([Yos16])**.** *For any directed graph $G = (V, E)$, $\lambda_G/2 \leq \vec{\phi}(G) \leq 2\sqrt{\lambda_G}$.*

An approximation algorithm for computing λ_G was not given in [Yos16]. Later on, Yoshida [Yos19] gave an SDP approximation algorithm for computing λ_G and this gives a polynomial time computable quantity $\tilde{\lambda}_G$ that satisfies $\tilde{\lambda}_G \lesssim \vec{\phi}(G) \lesssim \sqrt{\tilde{\lambda}_G \cdot \log n}$. As [Yos19] will be our primary reference in Chapter 6 and also that his work covers the more general setting of submodular transformations, we leave further discussion of his work to that chapter.

Note that the directed edge conductance $\vec{\phi}(G)$ and the Cheeger constant $h(G)$ in (3.9) are two very different quantities, one reason being that the stationary distribution π can have exponentially small values on some vertices. We discuss this further in Chapter 5 and also refer the interested reader to [Yos16].

Below we prove Theorem 3.3.4. The details are a bit different than previous proofs in this chapter because of the different normalization in the definition of λ_G.

Proof of Theorem 3.3.4. First, note that $\mu_G(v) = \sqrt{\deg(v)}$ is the trivial eigenvector to the nonlinear Laplacian, with eigenvalue 0. For the easy direction, given $S \subseteq V$ such that (without loss of generality) $0 < \mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ and $\vec{\phi}(S) = \vec{\phi}(G)$, consider the following solution to λ_G:

$$x(v) := \begin{cases} \sqrt{\deg(v)}/\mathrm{vol}(S), & \text{if } v \in S; \\ -\sqrt{\deg(v)}/\mathrm{vol}(S^c), & \text{otherwise.} \end{cases}$$

[5] A change of variables $f(u) := x(u)/\sqrt{\deg(u)}$ gives an equivalent definition that is closer to (2.2) for undirected graphs.

Then, $x \perp \mu_G$, and

$$\frac{\sum_{uv\in E}\left(\left[x(u)/\sqrt{\deg(u)} - x(v)/\sqrt{\deg(v)}\right]^+\right)^2}{\sum_{u\in V} x(u)^2} = |\delta^+(S)|\cdot\left(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(S^c)}\right) \le 2\cdot\frac{|\delta^+(S)|}{\mathrm{vol}(S)}.$$

If we replace x by $-x$, then $-x \perp \mu_G$ and the above upper bound becomes $2|\delta^+(S^c)|/\mathrm{vol}(S)$. Therefore, we have shown that

$$\lambda_G \le 2\cdot\frac{\min\{|\delta^+(S)|, |\delta^+(S^c)|\}}{\mathrm{vol}(S)} = \vec{\phi}(S).$$

For the hard direction, we again follow the two-step proof flow in Theorem 3.1.1.

Step 1 (ℓ_2^2 to ℓ_1). The goal is to construct a vector $y : V \to \mathbb{R}$ such that

$$\frac{\sum_{uv\in E}[y(u) - y(v)]^+}{\sum_{u\in V}\deg(u)|y(u)|} \le \left(2\cdot\frac{\sum_{uv\in E}\left(\left[x(u)/\sqrt{\deg(u)} - x(v)/\sqrt{\deg(v)}\right]^+\right)^2}{\sum_{u\in V} x(u)^2}\right)^{1/2}$$

and that 0 is a degree-weighted median of y. Let $x'(v) := x(v)/\sqrt{\deg(v)}$, and apply the same procedure of obtaining h from f in (3.1) in Theorem 3.1.1 to obtain y from x', with c being the degree-weighted median of x'. The analysis of the denominator gives

$$\sum_{u\in V}\deg(u)|y(u)| = \sum_{u\in V}\deg(u)(x'(u) - c)^2 \ge \sum_{u\in V}\deg(u)x'(u)^2,$$

by noting that $\sum_{u\in V}\deg(u)x'(u) = 0$. For the numerator, note that the construction of y preserves the order of the values, i.e. $y(u) \le y(v) \iff x'(u) \le x'(v)$. Therefore, by restricting the summation to $E' := \{uv \in E : x'(u) \ge x'(v)\}$ and following the calculations in the proof of Theorem 3.1.1, we get

$$\sum_{uv\in E}(y(u) - y(v))^+ \le \sqrt{2\sum_{uv\in E}[(x'(u) - x'(v))^+]^2 \cdot \sum_{u\in V}\deg(u)|y(u)|}.$$

Dividing both sides by $\sum_{u\in V}\deg(u)|y(u)|$ and substituting $x(v) = x'(v)\sqrt{\deg(v)}$ back in, we get the desired bound on y.

Step 2 (threshold rounding). Given vector $y : V \to \mathbb{R}$ from above, the goal is to show that

$$\vec{\phi}(S) \le \frac{\sum_{uv\in E}(y(u) - y(v))^+}{\sum_{u\in V}\deg(u)|y(u)|}$$

for some $S \subseteq V$. Let $t \in \mathbb{R}$ be a parameter and define $S_t \subseteq V$ as $S_t := \{v \in V : y(v) > t\}$. Note that 0 is a degree-weighted median of y, and so the "average" denominator of $\vec{\phi}(S_t)$ is

$$\int_{-\infty}^{\infty} \min\{\mathrm{vol}(S_t), \mathrm{vol}(S_t^c)\}\, dt = \int_0^{\infty} \mathrm{vol}(S_t)\, dt + \int_{-\infty}^{0} \mathrm{vol}(S_t^c)\, dt = \sum_{v \in V} \deg(v)|y(v)|.$$

As for the numerator of $\vec{\phi}(S_t)$, the "average" size is

$$\begin{aligned} \int_{-\infty}^{\infty} \min\{|\delta^+(S_t)|, |\delta^+(S_t^c)|\}\, dt &\le \int_{-\infty}^{\infty} |\delta^+(S_t)|\, dt \\ &= \sum_{uv \in E} \int_{-\infty}^{\infty} \mathbb{1}[y(u) > t \ge y(v)]\, dt = \sum_{uv \in E} (y(u) - y(v))^+. \end{aligned}$$

Therefore, for some $S = S_t \subseteq V$ the desired bound on $\vec{\phi}(S)$ holds, and we are done. □

3.4 Hypergraphs

3.4.1 Spectral Theory via Diffusion

Louis [Lou15] and Chan, Louis, Tang, Zhang [CLTZ18] developed a spectral theory for hypergraphs. They defined a continuous time diffusion process on a hypergraph $H = (V, E, w)$ and used it to define the Laplacian operator and its eigenvalues $\gamma_1 \le \gamma_2 \le \ldots \le \gamma_n$. The formulation is similar to the "spectral" quantity λ_∞ in [BHT00] for vertex expansion (c.f. Section 3.2.4), and they proved that there is an exact analog of Cheeger's inequality for hypergraphs:

Theorem 3.4.1 ([CLTZ18, Theorem 6.1]). *For any hypergraph $H = (V, E, w)$,*

$$\frac{1}{2}\gamma_2(H) \le \phi(H) \le \sqrt{2\gamma_2(H)},$$

where $\phi(H)$ is the hypergraph edge conductance of H defined in Section 2.3.3.

As in [BHT00], the quantity γ_2 is not polynomial time computable, and a semidefinite programming relaxation $\tilde{\gamma}_2$ (see [CLTZ18, SDP 8.3]) was used to prove the following Cheeger inequality and to design a $O(\sqrt{\phi(H) \log r})$-approximation algorithm for hypergraph edge conductance where r is the maximum size of a hyperedge.

Theorem 3.4.2 (Cheeger Inequality for Hypergraphs [CLTZ18, Theorem 8.1]). *For any hypergraph $H = (V, E, w)$ of rank r,*

$$\tilde{\gamma}_2(H) \lesssim \phi(H) \lesssim \sqrt{\tilde{\gamma}_2(H) \cdot \log r},$$

Using this spectral theory, they also obtain an analog of higher-order Cheeger inequality for hypergraph edge conductance, and also an approximation algorithm for small-set hypergraph edge conductance. Given a hypergraph $H = (V, E, w)$, the k-way edge conductance of H is defined as $\phi_k(H) := \min_{S_1, S_2, \ldots, S_k} \max_{1 \le i \le k} \phi(S_i)$ where $S_1, S_2, \ldots, S_k$ are over pairwise disjoint subsets of V. [CLTZ18, Theorem 6.14] states that

$$\tilde{\gamma}_k \lesssim \phi_k(H) \lesssim k^4 \cdot \log k \cdot \log\log k \cdot \log r \cdot \sqrt{\tilde{\gamma}_k} \text{ and } \phi_{(1-\varepsilon)k}(H) \lesssim \frac{k^{2.5}}{\varepsilon^{1.5}} \cdot \log k \cdot \log\log k \cdot \log r \cdot \sqrt{\tilde{\gamma}_k} \tag{3.11}$$

for any $\varepsilon \ge 1/k$, where $\tilde{\gamma}_k$ is an SDP relaxation of γ_k which can be computed in polynomial time. Furthermore, they proved a stronger bound in [CLTZ18, Corollary 3.23] about small-set conductance that there is a subset $S \subseteq V$ with $|S| = \Theta(n/k)$ and

$$\phi(S) \lesssim k^{1.5} \cdot \log k \cdot \log\log k \cdot \log r \cdot \sqrt{\tilde{\gamma}_k}. \tag{3.12}$$

In the rest of the section, we define γ_2 and $\tilde{\gamma}_2$, then present the proof for Theorem 3.4.1 and for the following relation between γ_2 and $\tilde{\gamma}_2$, which together imply Theorem 3.4.2. We shall not discuss in further detail the results involving the higher eigenvalues γ_k and $\tilde{\gamma}_k$.

Definition 3.4.3 (γ_2 and $\tilde{\gamma}_2$). *Given a hypergraph $H = (V, E, w)$, $\gamma_2(H)$ is defined as*

$$\gamma_2(H) := \inf_{\substack{f: V \to \mathbb{R} \\ f \perp \deg_w}} \frac{\sum_{e \in E} w(e) \cdot \max_{u,v \in e} (f(u) - f(v))^2}{\sum_{u \in V} \deg_w(u) f(u)^2},$$

and $\tilde{\gamma}_2(H)$ is defined as

$$\tilde{\gamma}_2(H) := \inf_{\substack{f: V \to \mathbb{R}^n \\ D_w f = \vec{0}}} \frac{\sum_{e \in E} w(e) \cdot \max_{u,v \in e} \|f(u) - f(v)\|^2}{\sum_{u \in V} \deg_w(u) \|f(u)\|^2}$$

where $D_w = \mathrm{diag}(\deg_w(v))$ and so $D_w f = \sum_{v \in V} \deg_w(v) f(v)$.

Proposition 3.4.4 ([CLTZ18, Proposition 8.5]). *For any hypergraph $H = (V, E, w)$ of rank r,*

$$\tilde{\gamma}_2(H) \le \gamma_2(H) \lesssim \tilde{\gamma}_2(H) \cdot \log r.$$

Proof of Theorem 3.4.1. We follow the template in the proof of Theorem 3.1.1. For the easy direction, given $S \subseteq V$ with $0 < \mathrm{vol}_w(S) \le \mathrm{vol}_w(V)/2$ consider

$$f(v) := \begin{cases} \frac{1}{\mathrm{vol}_w(S)}, & \text{if } v \in S \\ \frac{-1}{\mathrm{vol}_w(S^c)}, & \text{otherwise.} \end{cases}$$

Verify that, by the definition of $\delta(S)$, for a hyperedge $e \in E$,

$$\max_{u,v\in e}(f(u)-f(v))^2 = \begin{cases} \left(\frac{1}{\mathrm{vol}_w(S)} + \frac{1}{\mathrm{vol}_w(S^c)}\right)^2, & \text{if } e \in \delta(S); \\ 0, & \text{otherwise,} \end{cases}$$

and the rest of the proof proceeds as usual.

For the hard direction, the ℓ_2^2 to ℓ_1 step of establishing an $h : V \to \mathbb{R}$, such that

$$\frac{\sum_{e\in E} w(e)\cdot \max_{u,v\in e}|h(u)-h(v)|}{\sum_{u\in V}\deg_w(u)|h(u)|} \le \sqrt{2\cdot \frac{\sum_{e\in E} w(e)\cdot \max_{u,v\in e}(f(u)-f(v))^2}{\sum_{u\in V}\deg_w(u) f(u)^2}}$$

and that 0 is a degree-weighted median of h, is essentially the same as in the template proof, where the only modification comes from the numerator bound:

$$\begin{aligned}
&\sum_{e\in E} w(e)\cdot\max_{u,v\in e}|h(u)-h(v)| \\
\le\ &\sum_{e\in E} w(e)\cdot\max_{u,v\in e}\Big[|f(u)-f(v)|(|f(u)-c|+|f(v)-c|)\Big] \qquad \text{(by (3.2))} \\
\le\ &\sqrt{\sum_{e\in E} w(e)\cdot\max_{u,v\in e}(f(u)-f(v))^2}\cdot\sqrt{2\sum_{e\in E} w(e)\cdot\max_{\substack{u,v\in e\\ u\neq v}}\left(|f(u)-c|^2+|f(v)-c|^2\right)} \\
\le\ &\sqrt{\sum_{e\in E} w(e)\cdot\max_{u,v\in e}(f(u)-f(v))^2}\cdot\sqrt{2\sum_{e\in E} w(e)\cdot\sum_{u\in e}|f(u)-c|^2} \\
=\ &\sqrt{\sum_{e\in E} w(e)\cdot\max_{u,v\in e}(f(u)-f(v))^2}\cdot\sqrt{2\sum_{v\in V}\deg_w(v)(f(v)-c)^2}.
\end{aligned}$$

As for the threshold rounding step, the only modification needed is the "average" cut size bound. With S_t defined the same way, we have

$$\int_{-\infty}^{\infty} w(\delta(S_t))\,dt = \sum_{e\in E} w(e)\cdot\int_{-\infty}^{\infty}\mathbb{1}[e\in\delta(S_t)]\,dt = \sum_{e\in E} w(e)\cdot\max_{u,v\in e}|h(u)-h(v)|.$$

This completes the proof of the hard direction and thus the theorem. □

Proof of Proposition 3.4.4. The proof is basically the same as that of Theorem 3.2.17 in [LRV13]. First, $\tilde{\gamma}_2(H) \leq \gamma_2(H)$ because $\gamma_2(H)$ can be regarded as $\tilde{\gamma}_2(H)$ but further restricting all but the first coordinates of $f(v)$ to be zero. Next, given optimal $f : V \to \mathbb{R}^n$ to $\tilde{\gamma}_2(H)$, by defining the following one-dimensional solution $x : V \to \mathbb{R}$ where

$$x(v) = \langle f(v), z \rangle$$

where $z \sim N(0,1)^n$ is a random Gaussian vector, we can again establish the feasibility of x, apply Fact 2.10.6, linearity of expectations, and Markov's inequality to show that

$$\sum_{e \in E} w(e) \max_{u,v \in e} (x(u) - x(v))^2 \leq 96 \log(r^2) \cdot \sum_{e \in E} w(e) \max_{u,v \in E} \|f(u) - f(v)\|^2 ,$$

with probability at least 23/24, and apply Fact 2.10.7 to show that $\sum_{v \in V} \deg_w(v) x(v)^2 \geq \frac{1}{2} \sum_{v \in V} \deg_w(v) \|f(v)\|^2$ with probability at least 1/12. This proves that the objective value of the projected solution is at most $384 \tilde{\gamma}_2(H) \cdot \log r$ with probability at least 1/24. □

3.5 Eigenvalue Bounds for Special Graphs

In this section, we review previous work that upper bounds Laplacian eigenvalues for special classes of graphs. We first give a description of edge-based and vertex-based spectral partitioning and discuss the significance of eigenvalue upper bounds in analyzing the performances of these algorithms. Then, we review the work of Spielman and Teng [ST96] proving that spectral partitioning gives balanced separators of size $O(\sqrt{n})$ in bounded degree planar graphs, and the results of Kelner [Kel06], Biswal, Lee, Rao [BLR10], and Kelner, Lee, Price, Teng [KLPT11] that give upper bounds on the second and higher eigenvalues of graph Laplacians in bounded genus graphs and excluded minor graphs.

3.5.1 Edge-Based Spectral Partitioning

Given a graph $G = (V, E)$, the sweep-cut algorithm Algorithm 1 on the second eigenvector of $L(G)$ finds a nonempty vertex subset $S \subset V$ such that $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ and

$$\phi(S) \leq \sqrt{2\lambda_2(G)}.$$

Suppose that G belongs to a graph class $\mathcal{C}$ such that $\mathcal{C}$ is closed under edge and vertex removal, and $\lambda_2(G') \leq \lambda_{\mathcal{C}}$ for all $G' \in \mathcal{C}$. Then, we can remove the at most $O(\mathrm{vol}(S) \cdot \sqrt{\lambda_{\mathcal{C}}})$

edges crossing S and disconnect S from its complement $V \setminus S$. Recurse on $G[V \setminus S]$ until the total volume of the disconnected vertex subsets is at least $\mathrm{vol}(V)/3$. The total number of edges removed is thus $O(m\sqrt{\lambda_{\mathcal{C}}})$, and after the edge removals, each connected component has volume at most two thirds of the total. However, known results were unable to directly upper bound $\lambda_2(G)$, and involve instead the second smallest eigenvalue for the unnormalized Laplacian $\lambda_2'(G)$. While this implies an upper bound on $\lambda_2(G)$ via the following inequality:

$$\lambda_2(G) = \min_{f \perp \mathbb{1}} \frac{\sum_{uv \in E}(f(u) - f(v))^2}{\sum_{v \in V} \deg(v) f(v)^2} \leq \min_{f \perp \mathbb{1}} \frac{\sum_{uv \in E}(f(u) - f(v))^2}{\sum_{v \in V} f(v)^2} = \lambda_2'(G),$$

$\lambda_2'(G)$ may be Δ times larger than $\lambda_2(G)$ where Δ is the maximum degree of G.

To summarize, eigenvalue bounds are useful for ensuring that recursive spectral partitioning finds a balanced separator of small conductance, but existing indirect bounds via $\lambda_2'(G)$ may be suboptimal for the purpose of analyzing spectral partitioning.

3.5.2 Vertex-Based Spectral Partitioning

For vertex-based spectral partitioning, it is more apt to consider obtaining cuts with small vertex expansion. Below we present one spectral partitioning algorithm using the second eigenvector of the unnormalized Laplacian, which is based on the "classical" vertex Cheeger inequality [Tan84, AM85, Alo86] in Section 3.2.1.

Given a graph $G = (V, E)$, the flow-based algorithm in the proof of Theorem 3.2.1 takes the second eigenvector of the unnormalized Laplacian $L'(G)$ as input and finds a nonempty vertex subset $S \subset V$ such that $|S| \leq |V|/2$ and

$$\psi(S) \lesssim \sqrt{\lambda_2'(G)}.$$

Suppose that G belongs to a graph class $\mathcal{C}$ such that $\mathcal{C}$ is closed under edge and vertex removal, and $\lambda_2'(G') \leq \lambda_{\mathcal{C}}'$ for all $G' \in \mathcal{C}$. Proceeding as for edge-based partitioning, by repeatedly removing ∂S and disconnecting S from the rest of the graph, until the total number of disconnected vertices is at least $n/3$, the total number of vertices removed is $O(n\sqrt{\lambda_{\mathcal{C}}'})$, and after the vertex removals, each connected component has size at most $2n/3$. The set of removed vertices is called a *balanced separator*, since its removal separates the rest of the graph into balanced components, i.e. none of the components are too big.

Like edge-based spectral partitioning, separator size bounds for vertex-based spectral partitioning using $\lambda_2'(G)$ has a nontrivial dependence on the maximum degree Δ of the graph, and such bounds are only useful when Δ is small.

3.5.3 Planar Separation via Spectral Partitioning

Spielman and Teng [ST96] gave an upper bound on the second smallest eigenvalue for the *unnormalized* Laplacian $\lambda'_2(G)$ when G is a planar graph.

Theorem 3.5.1 (Restatement of Theorem 1.4.1)**.** *For any planar graph $G = (V, E)$ with maximum degree Δ, the second smallest eigenvalue $\lambda'_2(G)$ of the unnormalized Laplacian satisfies $\lambda'_2(G) \lesssim \Delta/n$.*

Using this theorem and the flow-based partitioning algorithm in the previous subsection, this implies that the algorithm produces an $O(\sqrt{\Delta n})$-sized balanced separator. In the paper, they used a theorem of Mihail [Mih89] which states that the sweep-cut algorithm on the second eigenvector of the unnormalized Laplacian produces a vertex subset $S \subset V$ with $0 < |S| \leq n/2$ and whose *edge expansion* $\varphi(S)$, and hence[6] vertex expansion $\psi(S)$, is at most $O(\sqrt{\Delta \lambda'_2(G)})$. Then, using this analysis, recursive application of the sweep-cut algorithm produces an $O(\Delta\sqrt{n})$-sized balanced separator.

Corollary 3.5.2 (Balanced Separator on Planar Graphs [ST96])**.** *For any planar graph $G = (V, E)$ of bounded maximum degree, i.e. $\Delta = O(1)$, the spectral partitioning algorithm described above using $\lambda'_2(G)$ produces a balanced separator of size $O(\sqrt{n})$.*

Both analyses match the optimal guarantee of $O(\sqrt{n})$-sized balanced separators asserted by the planar separator theorem of Lipton and Tarjan [LT79] only in the bounded degree case, and while the average degree of a planar graph is $O(1)$, the maximum degree can still go up to $O(n)$.

Now we outline the proof of the eigenvalue bound in Theorem 3.5.1. The idea is to produce a vector $f : V \to \mathbb{R}$, such that $f \perp \mathbb{1}$ and

$$\frac{\sum_{uv \in E} (f(u) - f(v))^2}{\sum_{v \in V} f(v)^2} \lesssim \frac{\Delta}{n}.$$

This will prove the theorem by the variational characterization of $\lambda'_2(G)$.

Their key observation is that the Koebe-Andreev-Thurston "kissing disks" embedding [Koe36, And70a, And70b, Thu78] is a good starting point towards producing such vector f. Given a graph $G = (V, E)$ and a surface Σ which we restrict to be either the plane $\mathbb{R}^2$ or the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$, a disk packing of G in Σ is a mapping that takes each

[6] It is easy to see that, in the unweighted setting, $\psi(S) \leq \varphi(S)$ for all vertex subsets S.

vertex $v \in V$ to a disk (when Σ is the plane) or a spherical cap (when Σ is the unit sphere) $D_v \subseteq \Sigma$, such that the interiors of D_v are pairwise disjoint. A "kissing disks" embedding of G in Σ is a disk packing such that D_u and D_v touch at the boundary if and only if $uv \in E$. Refer to Figure 3.2 for an illustration.

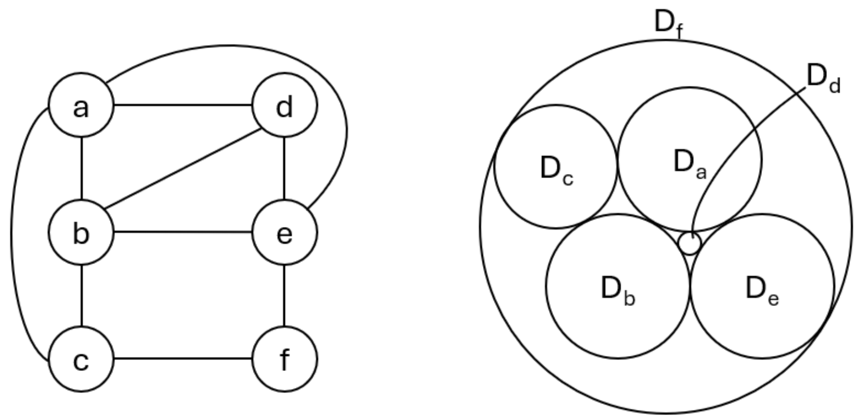


Figure 3.2: A planar graph and its "kissing disks" embedding on the plane. Note that in this case, the "disk" D_f is considered to be the unbounded region "outside" the circular boundary. Such irregularities do not appear when embedding to $\mathbb{S}^2$.

It is a powerful theorem by Koebe [Koe36], Andreev [And70a, And70b], and Thurston [Thu78] that a planar graph always admits a "kissing disks" embedding in Σ. Colin de Verdière [Col91] and Mohar [Moh93] are the first ones to design polynomial-time algorithms for computing such an embedding, after which many more proofs follow. The current fastest algorithm is by Dong, Lee, and Quanrud [DLQ20] and has worst-case runtime $\tilde{O}(n^2)$.[7]

Starting with any "kissing disks" embedding of G in $\mathbb{S}^2$, Spielman and Teng try to adjust the embedding so that the center of gravity of the centres $z(u)$ of the disks D_u is the origin. Such an embedding $z : V \to \mathbb{R}^3$ satisfies the conditions that $\sum_{u\in V} z(u) = \vec{0}$ and $\sum_{u\in V} \|z(u)\|^2 = \sum_{u\in V} 1 = n$. Moreover, if $r(u)$ is the geodesic radius of the disk D_u then

$$\sum_{uv\in E} \|z(u) - z(v)\|^2 = \sum_{uv\in E} (r(u) + r(v))^2 \leq 2 \sum_{uv\in E} (r(u)^2 + r(v)^2) \leq 2\Delta \sum_{v\in V} r(v)^2 \lesssim \Delta,$$

where the last inequality is because the area of the disk D_u is $\Theta(r(u)^2)$, and since the disks have disjoint interiors their areas sum to at most 4π (the area of the unit sphere), which is

[7] More precisely, the runtime depends on the ratio between radius of the largest disk and the radius of the smallest disk in a "true" embedding, as well as the desired accuracy. Refer to [DLQ20, Theorem 1.6] for more details.

$O(1)$. Therefore, by choosing the best coordinate, they obtain a one-dimensional solution $f : V \to \mathbb{R}$ such that $f \perp \mathbb{1}$ and

$$\frac{\sum_{uv \in E}(f(u) - f(v))^2}{\sum_{v \in V} f(v)^2} \lesssim \frac{\Delta}{n}.$$

Now we briefly explain the missing detail about how to adjust the embedding. They considered a family of maps $\Phi_{w,\rho}$ for $w \in \mathbb{S}^2$ and $\rho \in \mathbb{R}^+$. The map is by stereographically projecting the embedding in $\mathbb{S}^2$ to the plane in the $-w$ direction (so that the plane is tangent to the sphere at the point $-w$), scaling the points on the plane by a factor of ρ, then apply inverse stereographic projection to send the points back to the sphere.[8] Refer to Figure 3.3 for an illustration of the stereographic projection.

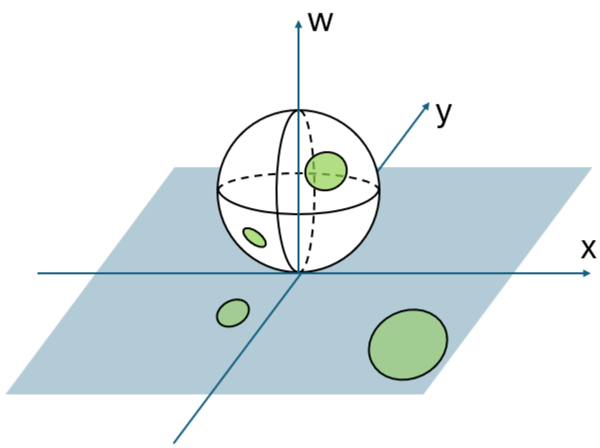


Figure 3.3: The stereographic projection. Disks on the sphere are mapped to disks on the plane.

The two key properties of these maps are that they are injective and that they preserve circles and disks, so that a "kissing disks" embedding gets mapped to another "kissing disks" embedding. They prove the existence of such a map $\Phi_{w,\rho}$ such that $\sum_{v \in V} \bar{z}(v) = \vec{0}$, where $\bar{z}(v)$ is the center of the image of D_v (which is again a disk) under $\Phi_{w,\rho}$. The proof is non-constructive and relies on the Brouwer fixed point theorem; we refer the reader to [ST96] for details.

[8]The point w is sent to a "point of infinity" ∞, then back to w.

Spielman and Teng conjectured that similar eigenvalue bounds should hold for graphs with bounded genus g and graphs which are K_h-minor free. The genus of a graph G is the smallest integer $g \geq 0$ such that G can be embedded in a torus with g holes without crossing edges. Formally:

Definition 3.5.3 (Non-crossing Embedding)**.** *Let $G = (V, E)$ be a graph and $\Sigma \subset \mathbb{R}^3$ be a surface. An embedding of G in Σ is a pair of maps (ρ, κ), where $\rho : V \to \Sigma$ is injective so that it maps vertices to distinct locations on the surface Σ, and $\kappa : E \to \mathcal{C}([0,1], \Sigma)$ maps edges to continuous simple curves on Σ with $\kappa(uv)(0) = \rho(u)$ and $\kappa(uv)(1) = \rho(v)$. The embedding (ρ, κ) is said to be non-crossing if the curves $\kappa(e_1)$ and $\kappa(e_2)$ do not intersect except possibly at the common endpoint of the curves.*

We say that G can be embedded without crossing in Σ if such an embedding (ρ, κ) exists.

A planar graph is a graph that is can be embedded in a plane without crossing, which is equivalent to embeddability in the 2-sphere without crossing (via stereographic projection). Therefore, planar graphs have genus 0, and the first part of the conjecture can be seen as a generalization of Theorem 1.4.1.

A graph G contains another graph H as a minor if we can obtain H from G by (1) contracting edges[9], (2) deleting edges, and (3) deleting vertices, otherwise we say that G is H-minor free. This definition is useful because H-minor free graphs are considered generalizations of planar graphs and bounded-genus graphs: it is a well-known result by Kuratowski [Kur30] that a graph G is planar if and only if G is K_5-minor free and $K_{3,3}$-minor free.[10] It is also known that K_h has genus $\Theta(h^2)$ [RY68], and so a graph that has genus g is K_h-minor free for some $h = O(\sqrt{g})$. Therefore, an eigenvalue upper bound on minor-free graphs would be a further generalization of Theorem 1.4.1.

3.5.4 Second Eigenvalue Bound for Bounded-Genus Graphs

Kelner [Kel06] first proved the conjecture for graphs with bounded degree and bounded genus.

Theorem 3.5.4 ([Kel06, Theorem 2.3])**.** *For any graph G with genus $g \geq 1$ and maximum degree Δ, $\lambda_2'(G) \lesssim \mathrm{poly}(\Delta) \cdot g/n$.*

[9] To contract an edge $uv \in E$ means to identify the vertices u and v, so that a "super-vertex" $\{u, v\}$ is created, the edge uv and the vertices u and v removed, and every other edge connected to either u or v now being connected to the super-vertex $\{u, v\}$.

[10] $K_{3,3}$ is the complete bipartite graph with three vertices on each side.

The exact dependence on the maximum degree Δ was not specified in the paper. The theorem implies that the vertex-based spectral partitioning algorithm described in Section 3.5.2 produces a balanced separator of size $O(\sqrt{gn})$ for bounded degree graphs and $O(\text{poly}(\Delta) \cdot \sqrt{gn})$ for general graphs of genus g.

Kelner's idea is to start with a "kissing disk" embedding of the graph G in the torus with g holes, then construe a circle-preserving (and locally injective) map to the 2-sphere so that at most $O(g)$ disks overlap at any point. If this is possible, then the same argument as before implies Theorem 3.5.4. To make this idea go through, however, it is necessary to subdivide the graph to obtain an approximately "continuous" structure and come up with a suitable notion of approximately circle-preserving map. In brief, the approach relies on nontrivial results in the theory of Riemann surfaces, and it is not clear how to generalize it to the K_h-minor free case.

3.5.5 Flow-Based Techniques for Bounding Second Eigenvalue

Biswal, Lee, and Rao [BLR10] finally settled the conjecture of Spielman and Teng [ST96], by proving a second eigenvalue upper bound on K_h-minor free graphs.

Theorem 3.5.5 (Upper Bound on $\lambda_2'(G)$, [BLR10, Theorem 5.2, 5.3])**.** *Let $G = (V, E)$ be a graph with maximum degree Δ. Then,*

- *If G is of genus $g \geq 1$, then $\lambda_2'(G) \leq O(g \log^2 g \cdot \frac{\Delta}{n})$.*[11]
- *If G is K_h minor free, then $\lambda_2'(G) \leq O(h^6 \log h \cdot \frac{\Delta}{n})$.*

The result implies that vertex-based spectral partitioning produces balanced separators of size $O\left(\sqrt{g \log^2 g \cdot \Delta n}\right)$ and $O\left(\sqrt{h^6 \log h \cdot \Delta n}\right)$ respectively.

Instead of passing to the extrinsic geometry of the graph via embeddings, their approach is more combinatorial and is based on network flows. They relate the second eigenvalue to a metric quantity using the Rayleigh quotient, then establish a flow/metric duality to pass to flows. The flow problem under consideration has demand graph K_n, so that we are sending flows between every pair of vertices in G. The less well-connected G is, the more congested the flows are, and by relating a measure of flow congestion to the graph parameters via the "crossing number" of flows, the result is proven.

We provide an outline of the proof by [BLR10] in the remainder of the subsection.

[11] In [BLR10] a weaker statement is proved, that $\lambda_2'(G) \leq O(g^3 \cdot \frac{\Delta}{n})$. The bound in the theorem statement appears in [KLPT11] and relies upon a stronger result in [LS10] concerning the decomposition of shortest-path metrics on genus g graphs.

Rayleigh Quotient to Shortest Path Metric

First, write the target quantity $\lambda_2'(G)$ as a Rayleigh quotient, then relate it to a minimization problem involving metrics. The easiest to relate to is the ℓ_2^2 distance, and afterwards they pass to the shortest path metric which is dual (in an appropriate sense) to flows. The following proposition details the relation between eigenvalues and ℓ_2^2 metrics.

Proposition 3.5.6. *Let $G = (V, E)$ be a graph. Then,*

$$\frac{1}{2n}\lambda_2'(G) = \min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E}(f(u)-f(v))^2}{\sum_{u,v\in V}(f(u)-f(v))^2}. \tag{3.13}$$

Proof. By Rayleigh quotient definition of $\lambda_2'(G)$ similar to Proposition 2.5.3, we have

$$\lambda_2'(G) = \min_{f\perp\mathbb{1}} \frac{\sum_{uv\in E}(f(u)-f(v))^2}{\sum_{v\in V} f(v)^2}.$$

For $f : V \to \mathbb{R}$ with $\sum_{v\in V} f(v) = 0$, Fact 2.10.4 asserts that

$$\sum_{v\in V} f(v)^2 = \frac{1}{2n}\sum_{u,v\in V}(f(u)-f(v))^2. \tag{3.14}$$

Substituting (3.14) into the denominator of the Rayleigh quotient, then dropping the condition that $f \perp \mathbb{1}$ once we notice that RHS of (3.13) is translation invariant, the proof is complete. □

Next, they relax the minimization problem on RHS of (3.13). Their main lemma, which we present shortly, shows that shortest path metrics are not too far away from ℓ_2^2 metrics, in terms of *average* distortion. This helps transition from the minimization problem involving ℓ_2^2 metrics to a minimization problem involving shortest path metrics. The proof relies on results in metric decomposition which are not our main focus here, so we omit the proof. For details, refer to [BLR10, Section 4].

Before stating the lemma, some new notations are in order. We consider vertex-weighted shortest path metrics here, meaning that the cost of a path is the sum of costs of vertices it passes through. For any graph $G = (V, E)$, and for any $s : V \to \mathbb{R}_{\geq 0}$, the vertex-weighted shortest path metric d_s induced by s is defined as

$$d_s(u, v) := \min_{p:u\text{-}v \text{ path}} \sum_{w\in V(p)} s(w), \tag{3.15}$$

where $V(p)$ is the set of vertices in the path p (including the endpoints u and v).

Lemma 3.5.7 (Average Distortion [BLR10, Theorem 4.4])**.** *For any graph $G = (V, E)$, there exists $\alpha(G) > 0$ such that the following holds: for any shortest path metric d_s, there is a function $f : V \to \mathbb{R}$, such that if we write $d_f(u, v) := |f(u) - f(v)|$ then $d_f(u, v) \le d_s(u, v)$ for all $u, v \in V$ and*

$$\sum_{u,v \in V} d_s(u, v)^2 \lesssim \alpha(G)^2 \cdot \sum_{u,v \in V} d_f(u, v)^2.$$

Furthermore, $\alpha(G)$ is the largest $\alpha > 0$ such that, for any vertex-weighted shortest path metric d_s and $r > 0$, there exists an $(r, \alpha, 1/2)$-padded decomposition of the metric space (V, d_s). It satisfies the following bounds:

- *([Bar96]) For any graph G, $\alpha(G) = O(\log n)$;*
- *([LS10]) If G has genus $g \ge 1$, then $\alpha(G) = O(\log g)$;*
- *([KPR93]) If G is K_h-minor free, then $\alpha(G) = O(h^2)$.*

Using this lemma and after some work, they upper bound the value of $\lambda_2'(G)$ using the value of a concave maximization program, relating to the spread of shortest path metrics:

$$\begin{aligned}
\frac{1}{2n}\lambda_2'(G) &= \min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E} d_f(u,v)^2}{\sum_{u,v\in V} d_f(u,v)^2} \\
&\lesssim \alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\ge 0}} \frac{\sum_{uv\in E} d_s(u,v)^2}{\sum_{u,v\in V} d_s(u,v)^2} \\
&= \alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\ge 0}} \frac{\sum_{uv\in E} (s(u)+s(v))^2}{\sum_{u,v\in V} d_s(u,v)^2} \\
&\le 2\alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\ge 0}} \frac{\sum_{uv\in E} (s(u)^2+s(v)^2)}{\sum_{u,v\in V} d_s(u,v)^2} \\
&\le 2\Delta \cdot \alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\ge 0}} \frac{\sum_{v\in V} s(v)^2}{\sum_{u,v\in V} d_s(u,v)^2} \\
&\overset{(*)}{\le} 2\Delta n^2 \cdot \alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\ge 0}} \frac{\sum_{v\in V} s(v)^2}{(\sum_{u,v\in V} d_s(u,v))^2} \\
&= 2\Delta n^2 \cdot \alpha(G)^2 \cdot \left[\max_{s:V\to\mathbb{R}_{\ge 0}} \frac{\sum_{u,v\in V} d_s(u,v)}{\sqrt{\sum_{v\in V} s(v)^2}} \right]^{-2},
\end{aligned}$$

where $(*)$ is by Cauchy-Schwarz inequality that $(\sum_{u,v\in V} d_s(u,v))^2 \le n^2 \sum_{u,v\in V} d_s(u,v)^2$, and the last equality is just rewriting the min program as a max program that is concave and more convenient to work with in latter steps. We summarize the results of this part in the following proposition.

Lemma 3.5.8 ([BLR10]). *Let $G = (V, E)$ be a graph with maximum degree Δ. Then,*

$$\lambda_2'(G) \lesssim \Delta n^3 \cdot \alpha(G)^2 \cdot \left[\max_{s:V \to \mathbb{R}_{\geq 0}} \frac{\sum_{u,v \in V} d_s(u,v)}{\sqrt{\sum_{v \in V} s(v)^2}} \right]^{-2},$$

where $\alpha(G)$ is the average distortion parameter in Lemma 3.5.7 and $d_s(u,v)$ is the shortest path metric defined in (3.15).

Metric Spread and Flow Congestion

Next, they relate the objective of the maximization program

$$\max_{s:V \to \mathbb{R}_{\geq 0}} \frac{\sum_{u,v \in V} d_s(u,v)}{\sqrt{\sum_{u \in V} s(u)^2}} =: \max_{s:V \to \mathbb{R}_{\geq 0}} \Lambda_s(G)$$

to a multicommodity flow problem where the goal is to minimize some measure of flow congestion. Refer to Section 2.9 for definitions about multicommodity flows.

Given a flow solution F, the congestion $c_F(u)$ at vertex $u \in V$ is the total amount of flow in F passing through u, and the vertex congestion of F considered here is the 2-norm of the amount of flow passing through each vertex:

$$\mathrm{con}(F) := \left(\sum_{v \in V} c_F(v)^2 \right)^{1/2}.$$

Intuitively, if the graph is well-connected, then both the spreading quantity $\Lambda_s(G)$ and the minimum congestion (for a suitable multicommodity flow problem) will be small. It turns out that a much stronger relation holds: maximum spreading is dual to minimum congestion.

Lemma 3.5.9 (Flow/Metric Duality, [BLR10, Theorem 2.2]). *For any graph $G = (V, E)$,*

$$\min_{F \in \mathcal{F}(G)} \mathrm{con}(F) = \max_{s:V \to \mathbb{R}_{\geq 0}} \Lambda_s(G),$$

where the minimum is taken over all flow solutions F for the multicommodity flow problem with uniform demand $D(u,v) = 1$ for all $u \neq v \in V$.

Proof. The proof is by writing out the Lagrangian of the minimum congestion program, simplifying it to obtain $\Lambda_s(G)$ as a dual program, and lastly finding a Slater point to establish strong duality.

Introduce primal variables $c(v)$ for the congestion at vertex v and $w(p)$ for the amount of flow sent along a path p. Use $\mathcal{P}$ to denote the set of all paths on G and $\mathcal{P}(u,v)$ to denote the set of all u-v paths on G. The minimum congestion program can then be written as

$$\begin{aligned}
\min_{c,w} \quad & \left(\sum_{v \in V} c(v)^2\right)^{1/2} \\
\text{subject to} \quad & c(v) = \sum_{p \ni v} w(p) \qquad \forall v \in V \\
& w(p) \geq 0 \qquad \forall p \in \mathcal{P} \\
& \sum_{p \in \mathcal{P}(u,v)} w(p) = 1 \qquad \forall u, v \in V, u \neq v.
\end{aligned}$$

Note that the objective function is convex. Using dual variables $s(v), \mu(p), \alpha(u,v)$ for the three constraints, we obtain the Lagrangian dual program as

$$\begin{aligned}
\max_{s,\mu,\alpha} \min_{c,w} \quad & \left(\sum_{v \in V} c(v)^2\right)^{1/2} - \sum_{v \in V} s(v) \left[c(v) - \sum_{p \ni v} w(p)\right] \\
& - \sum_{p \in \mathcal{P}} \mu(p) w(p) - \sum_{u \neq v} \alpha(u,v) \left[\sum_{p \in \mathcal{P}(u,v)} w(p) - 1\right] \\
\text{subject to} \quad & \mu(p) \geq 0 \qquad \forall p \in \mathcal{P}.
\end{aligned}$$

For fixed s, μ, α, we solve the inner minimization problem. We isolate the part relevant to c from the part relevant to w and minimize them separately. The part relevant to c is

$$\left(\sum_{v \in V} c(v)^2\right)^{1/2} - \sum_{v \in V} s(v) c(v).$$

First derivative test yields the local minimizer condition

$$c(v) = s(v) \cdot \left(\sum_{v \in V} c(v)^2\right)^{1/2} \qquad \forall v \in V,$$

and the objective becomes

$$\left(\sum_{v\in V} c(v)^2\right)^{1/2}\left(1-\sum_{v\in V} s(v)^2\right).$$

We see that the minimum value is 0 if $\sum_{v\in V} s(v)^2 \le 1$ and $-\infty$ otherwise. Therefore, we may remove this from the objective of the Lagrangian dual and instead add the constraints that $\sum_{v\in V} s(v)^2 \le 1$.

The part relevant to w is

$$\sum_{u,v\in V}\sum_{p\in\mathcal{P}(u,v)} w(p)\left[\sum_{v\in p} s(v)-\mu(p)-\mathbb{1}[u\neq v]\cdot\alpha(u,v)\right].$$

For the minimum value to not be $-\infty$, we need $\sum_{v\in p} s(v)-\mu(p)-\mathbb{1}[u\neq v]\cdot\alpha(u,v)=0$ for all $p\in\mathcal{P}(u,v)$, which is equivalent to $s(v)\ge 0$ for all $v\in V$ and

$$\alpha(u,v)\le\sum_{v\in p} s(v) \qquad \forall u\neq v, p\in\mathcal{P}(u,v).$$

Again, we add these as constraints and remove the w part from the objective of the Lagrangian. After these steps, the primal variables w and c are eliminated, μ becomes redundant, and we end up with the following maximization problem:

$$\begin{aligned}\max_{s\ge 0,\alpha} \quad & \sum_{u\neq v}\alpha(u,v)\\ \text{subject to} \quad & \alpha(u,v)\le\sum_{u'\in p} s(u') \qquad \forall u\neq v, p\in\mathcal{P}(u,v)\\ & \sum_{v\in V} s(v)^2\le 1.\end{aligned}$$

Clearly, the best choice of $\alpha(u,v)$ is $\alpha(u,v)=d_s(u,v)$ where $d_s(u,v)$ is the s-weighted shortest path length from u to v. Since $\Lambda_s(G)$ is homogeneous in s, we see that $\max_{s:V\to\mathbb{R}_{\ge 0}}\Lambda_s(G)$ is equivalent to the above program.

It remains to establish strong duality. It follows from the convexity of the primal objective and the existence of Slater point by taking $w(p)=1/|\mathcal{P}(u,v)|$ for any $p\in\mathcal{P}(u,v)$ and $c(v)=\sum_{p\ni v} w(p)$. □

Flow Congestion Lower Bound

Finally, a concrete lower bound on the minimum congestion is derived, which yields an upper bound on $\lambda'_2(G)$.

Lemma 3.5.10 (Congestion Lower Bound, [BLR10, Theorem 3.1, 3.11])**.** *Let $G = (V, E)$ be a graph and let F be a flow solution to the multicommodity flow problem on G with uniform demands $D(u, v) = 1$ for $u, v \in V$. Then,*

- *If G is of genus $g \geq 1$, and $n \gtrsim \sqrt{g}$, then $\mathrm{con}(F) \gtrsim n^2/\sqrt{g}$;*
- *If G is K_h-minor free for $h \geq 3$, and $n \gtrsim h\sqrt{\log h}$, then $\mathrm{con}(G) \gtrsim n^2/(h\sqrt{\log h})$.*

The idea of proof for the genus g case is that, if $\mathrm{con}(F)$ is too small, then we can round it to an integral flow solution[12] F' with $\mathrm{con}(F')$ small as well. As the demand graph is K_n, this turns out to induce an embedding of K_n in a genus g surface Σ with few pairs of crossing edges, contradicting known lower bounds on the minimum number of edge crossings when K_n is embedded in a genus g surface. The proof for the K_h-minor free case follows the same lines, but it proceeds by lower bounding the so-called intersection number instead of the crossing number. We will not delve into the details of the proof here, and remark that flow solutions can be rounded to integral flow solutions and relate to combinatorial parameters of the graph. The interested reader is referred to [BLR10] for details.

Putting It All Together

Now we are ready to prove the eigenvalue upper bounds in Theorem 3.5.5.

Proof of Theorem 3.5.5. Recall from Lemma 3.5.8 and Lemma 3.5.9 that

$$\lambda'_2(G) \lesssim \Delta n^3 \cdot \alpha(G)^2 \cdot \left(\min_{F \in \mathcal{F}(G)} \mathrm{con}(F) \right)^{-2}$$

where the minimum of F is taken over all flow solutions F on demand graph K_n. Applying Lemma 3.5.7 and Lemma 3.5.10, we have:

- When G is of genus g, $\alpha(G) \lesssim \log g$ and $\min_{F \in \mathcal{F}(G)} \mathrm{con}(F) \gtrsim n^2/\sqrt{g}$.
  Plugging these bounds in, we obtain $\lambda'_2(G) \leq O(g \log^2 g \cdot \Delta/n)$.

[12] An integral flow is a flow solution that has an integer amount of flow through each arc.

- When G is K_h-minor free, $\alpha(G) \lesssim h^2$ and $\min_{F \in \mathcal{F}(G)} \mathrm{con}(F) \gtrsim n^2/(h\sqrt{\log h})$.[13] Plugging these bounds in, we obtain $\lambda_2'(G) \leq O(h^6 \log h \cdot \Delta/n)$.

This completes the proof of Theorem 3.5.5. □

Remark 3.5.11 (Direct Upper Bounds on Vertex Expansion). *Using the same technique, [BLR10, Theorem 5.5] proves upper bounds directly on the vertex expansion of special classes of graphs: $O(1/\sqrt{n})$ for planar graphs, $O(\sqrt{g}\log g/\sqrt{n})$ for bounded genus graphs, and $O(h^3\sqrt{\log h}/\sqrt{n})$ for K_h-minor free graphs. While these are better upper bounds than those obtained via eigenvalue upper bounds and the "classical" vertex Cheeger inequality [Tan84, AM85, Alo86], no polynomial-time algorithm is known that finds a cut satisfying these bounds.*

3.5.6 Flow-Based Techniques for Bounding Higher Eigenvalues

Generalizing the work of Biswal, Lee and Rao [BLR10], Kelner, Lee, Price and Teng [KLPT11] derived an upper bound on higher eigenvalues of the unnormalized Laplacian for the same special classes of graphs.

Theorem 3.5.12 (Upper bound on $\lambda_k'(G)$, [KLPT11, Theorem 5.1]). *Let $G = (V, E)$ be a graph with maximum degree Δ, and let $1 \leq k \leq n$. Then,*

- *If G is planar, then $\lambda_k'(G) \leq O(\frac{\Delta k}{n})$.*
- *More generally, if G is of genus $g \geq 1$, then $\lambda_k'(G) \leq O(g \log^2 g \cdot \frac{\Delta k}{n})$.*
- *If G is K_h-minor free, then $\lambda_k'(G) \leq O(h^6 \log h \cdot \frac{\Delta k}{n})$.*

The high-level idea is to relate $\lambda_k'(G)$ to certain quantities relevant to shortest-path metrics on G, then control these quantities using the special properties of G. This result predates the higher-order Cheeger inequality [LOT12, LRTV12] (c.f. Section 3.1.3), and the main application was to provide a theoretical justification for heuristic algorithms for graph partitioning and clustering; see the references in [KLPT11]. Combined with [LOT12, LRTV12], Theorem 3.5.12 implies a spectral algorithm to retrieve k disjoint subsets $S_1, \ldots, S_k$ such that the k-way conductance $\phi_k(S_1, \ldots, S_k)$ is small.

We provide an outline of the proof by [KLPT11] in the rest of the subsection.

[13] When $h \geq \omega(\sqrt{\log n})$, we may use the better average distortion bound that $\alpha(G) \lesssim \log n$.

Rayleigh Quotient to Metric Parameters

The first component of the proof is to relate $\lambda'_k(G)$ to the following metric quantities:

Definition 3.5.13 (Metric Padding Parameter [KLPT11]). *Let (X, d_X) be a finite metric space. For any $x \in X$ and $r > 0$, let $B(x, r) := \{x' \in X : d_X(x, x') < r\}$ denote the open ball centered at x with radius r. For any partitioning P of X, using $P(x)$ to denote the partition that contains $x \in X$, define the padding parameter $\beta(P, \gamma)$ to be the infimal value of $\beta \geq 1$ such that*

$$|\{x \in X : B(x, \gamma/\beta) \subseteq P(x)\}| \geq \frac{|X|}{2}.$$

In other words, at least half of the points in X satisfy that all points (γ/β)-close to it are in the same partition. Further, let

$$\beta_\gamma(X, d_X) := \inf_P \beta(P, \gamma),$$

where the infimum is taken over all parititoning P of (X, d_X) where each partition has diameter at most γ.

Definition 3.5.14 (Metric Spreading Parameter [KLPT11]). *Let $G = (V, E)$ be a graph. Let $s : V \to \mathbb{R}_{\geq 0}$ be a weight function on the vertices, and d_s be the associated vertex-weighted shortest path metric on V (c.f. (3.15)). For $\varepsilon > 0$ and a collection Ψ of nonempty subsets of V, say that s is (Ψ, ε)-spreading if, for every $S \in \Psi$ one has*

$$\frac{1}{|S|^2} \sum_{u,v \in S} d_s(u, v) \geq \varepsilon \cdot \sqrt{\sum_{u \in V} s(u)^2}.$$

Write $\varepsilon_\Psi(G, s)$ to be the maximal value of ε such that s is (Ψ, ε)-spreading.

In [KLPT11], this definition is restricted to $\Psi = \Psi_r$ with Ψ_r being the collection of size-r subsets of V. We use (r, ε)-spreading in place of (Ψ_r, ε)-spreading and $\varepsilon_r(G, s)$ in place of $\varepsilon_{\Psi_r}(G, s)$ to refer to this restricted setting.

Given $G = (V, E)$, they use the metric padding parameter to find a well-padded partitioning of V. The metric spreading parameter plays the role of the spreading parameter in Definition 3.1.8 to upper bound the size of each partition. Using the same smooth localization procedure as seen in Lemma 3.1.14, they create from the partitioning k disjointly supported functions $f_1, \ldots, f_k : V \to \mathbb{R}$, each having small Rayleigh quotient. This is summarized in the following lemma.

Lemma 3.5.15 (Eigenvalues and Metric Parameters [KLPT11, Theorem 2.3])**.** *Let $G = (V, E)$ be a graph with maximum degree Δ. For any $1 \leq k \leq n$, the following holds. For any weight function $s : V \to \mathbb{R}_{\geq 0}$ satisfying*

$$\sum_{u \in V} s(u)^2 = 1,$$

we have

$$\lambda'_k(G) \leq \frac{256\Delta}{\varepsilon^2 n} \left(\beta_{\varepsilon/2}(V, d_s)\right)^2,$$

where $\varepsilon = \varepsilon_{\lfloor n/4k \rfloor}(G, s)$.

Proof. Let $\beta = \beta_{\varepsilon/2}(V, d_s)$. By Definition 3.5.13, there exists a partitioning P of V such that:

- Each partition $P(v)$ has diameter $\leq \varepsilon/2$;
- At least half of the vertices $v \in V$ satisfy $B(v, \varepsilon/(2\beta)) \subseteq P(v)$.

Let $S_1, \ldots, S_\ell$ be the partitions in P, and let

$$S'_i := \{u \in S_i : B(u, \varepsilon/(2\beta)) \subseteq S_i\}$$

be the "core" of S_i: vertices in S_i whose $\varepsilon/(2\beta)$-neighborhood is contained entirely in S_i. Then, by the second property of P we know that

$$|S'_1 \sqcup \cdots \sqcup S'_\ell| \geq n/2.$$

By Definition 3.5.14, every vertex subset S of size $r = \lfloor n/4k \rfloor$ satisfies

$$\frac{1}{|S|^2} \sum_{u,v \in S} d_s(u, v) \geq \varepsilon \cdot \sqrt{\sum_{u \in V} s(u)^2} = \varepsilon.$$

This implies in particular that the diameter of every subset of size r is at least ε, and so by contrapositive, $|S'_i| \leq |S_i| < r = \lfloor n/4k \rfloor$ for all $i \in [\ell]$. By possibly unioning some of the sets S_i, we can find disjoint subsets $T_1, \ldots, T_q$ and their "cores" $T'_1, \ldots, T'_q$ so that $B(u, \varepsilon/(2\beta)) \subseteq T_i$ for each $u \in T'_i$, and that each T'_i has size between $r/2$ and r. As $|\sqcup_i S'_i| \geq n/2$, this means that there are at least $q \geq (n/2)/r \geq 2k$ subsets.

Now we use these sets to define disjointly supported vectors with small Rayleigh quotient. Consider the following "smooth localization" $f_1, \ldots, f_q$ of the subsets:

$$f_i(u) := \max\left(0, \frac{\varepsilon}{2\beta} - d_s(u, T_i')\right).$$

First, note the following about the mass of each f_i:

$$\|f_i\|^2 = \sum_{u \in V} f_i(u)^2 \geq \left(\frac{\varepsilon}{2\beta}\right)^2 \cdot |T_i'| \geq \frac{\varepsilon^2 n}{32\beta^2 k}.$$

Next, observing that each f_i is supported on T_i, the total energy with respect to the unnormalized Laplacian is

$$\begin{aligned}
\sum_{i \in [q]} \mathcal{E}(f_i) &= \sum_{i \in [q]} \sum_{uv \in E} (f_i(u) - f_i(v))^2 \\
&\leq \sum_{i \in [q]} \sum_{u \in T_i} \sum_{v : uv \in E} (f_i(u) - f_i(v))^2 \\
&\leq \sum_{i \in [q]} \sum_{u \in T_i} \sum_{v : uv \in E} d_s(u, v)^2 \\
&\leq 2 \sum_{i \in [q]} \sum_{u \in T_i} \sum_{v : uv \in E} (s(u)^2 + s(v)^2) \\
&\leq 4\Delta \sum_{u \in V} s(u)^2 = 4\Delta.
\end{aligned}$$

The second inequality is because f_i is 1-Lipschitz with respect to d_s, the third inequality is because $d_s(u,v)^2 = (s(u)+s(v))^2 \leq 2(s(u)^2+s(v)^2)$, and the final inequality is because each $s(u)^2$ appears at most $2 \cdot \deg(u) \leq 2\Delta$ times in the sum. Since the number of functions is at least $2k$, there exists k such f_i's with disjoint support, such that $\mathcal{E}(f_i) \leq 4\Delta/k$. Therefore,

$$\frac{\mathcal{E}(f_i)}{\|f_i\|^2} \leq \frac{128\Delta}{\varepsilon^2 n} \beta^2$$

for the chosen f_i's, and using Proposition 2.5.4[14] this implies the desired conclusion. □

In light of Lemma 3.5.15, the desired eigenvalue upper bounds will follow from upper bounds on the metric padding parameter and lower bounds on the metric spreading parameter.

[14] While Proposition 2.5.4 is stated for the normalized Laplacian, the same proof can be applied to obtain an analogous statement for the unnormalized Laplacian and the corresponding Rayleigh quotient.

Bounds on Metric Padding

The second component of the proof is to upper bound the metric padding parameter in Definition 3.5.13 for shortest path metrics of special classes of graphs. The results are readily available in the literature, and we summarize the bounds here.

Lemma 3.5.16 (Upper Bounding Metric Padding Parameter [KPR93, FT03, LS10, KLPT11])**.** *Let $G = (V, E)$ be a graph, and let d_s be any shortest path metric corresponding to a vertex weight function $s : V \to \mathbb{R}_{\geq 0}$. Let $\gamma > 0$.*

- *[KPR93, FT03] If G is planar, then $\beta_\gamma(V, d_s) \leq O(1)$.*
- *[LS10] If G has genus $g \geq 1$, then $\beta_\gamma(V, d_s) \leq O(\log g)$.*
- *[KPR93, FT03] If G excludes K_h as a minor, then $\beta_\gamma(V, d_s) \leq O(h^2)$.*

(Note that the big-O bound does not depend on γ.)

Bounds on Metric Spreading

The third and final component of the proof is to lower bound the metric spreading parameter in Definition 3.5.14. This is the main novelty of their work. The first step is to consider the dual of the metric spreading parameter, which turns out to be a congestion minimization program on a new multicommodity flow problem called subset flows. The second step is to employ combinatorial and analytic tools to lower bound the minimum congestion.

Lemma 3.5.17 (Flow/Metric Duality [KLPT11])**.** *Let $G = (V, E)$ be a graph and let Ψ be a collection of nonempty subsets of V. Then, the metric spreading maximization problem is strongly dual to the "subset flow" minimum congestion problem:*

$$\max_{s:V\to\mathbb{R}_{\geq 0}} \varepsilon_\Psi(G, s) = \min_{F\in\mathcal{F}^\Psi(G)} \mathrm{con}(F),$$

where $\mathcal{F}^\Psi(G)$ is the set of all multicommodity flows on G whose demand graph satisfies

$$D_h(u, v) = \sum_{A\in\Psi:\{u,v\}\subseteq A} \frac{h(A)}{|A|^2} \qquad \forall u, v \in V$$

for some distribution h on Ψ, i.e. $h : \Psi \to \mathbb{R}_{\geq 0}$ with $\sum_{A\in\Psi} h(A) = 1$.

The interpretation of subset flows is that the demand graph is a weighted sum of "base" demand graphs on subsets, where on each subset S the base demand graph has $1/|S|^2$ demand between each pair of vertices in S. Note that in [KLPT11], only $\Psi = \Psi_r$ was considered, so the admissible subsets are only those of size r for some fixed r.

Proof. The proof is again by standard Lagrangian duality. Rewrite the metric spreading maximization problem as follows:

$$\begin{aligned}
\max_{\varepsilon,\delta,s} \quad & \varepsilon \\
\text{subject to} \quad & \frac{1}{|A|^2} \sum_{u,v \in A} \delta(u,v) \geq \varepsilon \qquad \forall A \in \Psi \\
& \delta(u,v) \leq \sum_{u' \in p} s(u') \qquad \forall u,v \in V \ \forall p \in \mathcal{P}(u,v) \\
& s(u) \geq 0 \qquad \forall u \in V \\
& \sum_{u \in V} s(u)^2 = 1.
\end{aligned}$$

Here $\mathcal{P}(u,v)$ denotes the set of all paths on G from u to v and $\delta(u,v)$ represents the length of the shortest path from u to v. Introduce dual variables $h(A)$ for the first constraint, $\alpha(p)$ for the second constraint, and μ for the final constraint. We obtain the Lagrangian dual program as

$$\begin{aligned}
\min_{h,\alpha,\mu} \max_{\varepsilon,\delta,s} \quad & \varepsilon + \sum_{A \in \Psi} h(A) \left(\frac{1}{|A|^2} \sum_{u,v \in A} \delta(u,v) - \varepsilon \right) \\
& + \sum_{u,v} \sum_{p \in \mathcal{P}(u,v)} \alpha(p) \left(\sum_{u' \in p} s(u') - \delta(u,v) \right) + \mu \left(1 - \sum_{u \in V} s(u)^2 \right) \\
\text{subject to} \quad & h(A) \geq 0 \qquad \forall A \in \Psi \\
& \alpha(p) \geq 0 \qquad \forall p \in \mathcal{P} \\
& s(u) \geq 0 \qquad \forall u \in V.
\end{aligned}$$

We first solve the inner maximization problem to eliminate the primal variables, and then interpret the dual variables as subset flow parameters. For the inner maximization problem, the part involving ε is

$$\varepsilon \left(1 - \sum_{A \in \Psi} h(A) \right),$$

as ε is unconstrained, for the maximum value to not be ∞, we need $\sum_{A\in\Psi} h(A) = 1$, in which case the maximum is 0. The part involving $\delta(u,v)$ is

$$\delta(u,v)\left[\sum_{A\in\Psi:\{u,v\}\subseteq A}\frac{h(A)}{|A|^2} - \sum_{p\in\mathcal{P}(u,v)}\alpha(p)\right].$$

Again, as $\delta(u,v)$ is unconstrained, for the maximum value to not be ∞, we need

$$\sum_{p\in\mathcal{P}(u,v)}\alpha(p) = \sum_{A\in\Psi:\{u,v\}\subseteq A}\frac{h(A)}{|A|^2} =: D_h(u,v),$$

in which case the maximum is 0. Finally, the part involving $s(u)$ is

$$\left(\sum_{p\in\mathcal{P}:u\in p}\alpha(p)\right)\cdot s(u) - \mu\cdot s(u)^2.$$

Write

$$C(u) := \sum_{p\in P:u\in p}\alpha(p).$$

When $\mu < 0$ this is unbounded, and otherwise first derivative test gives the optimizer

$$s(u) = \frac{C(u)}{2\mu}.$$ [15]

Simplifying, the dual program becomes

$$\begin{aligned}
\min_{h,\alpha,\mu} \quad & \sum_{u\in V} C(u)\left(\frac{C(u)}{2\mu}\right) + \mu\left(1 - \sum_{u\in V}\left(\frac{C(u)}{2\mu}\right)^2\right)\\
\text{subject to} \quad & h(A)\geq 0 \qquad \forall A\in\Psi\\
& \sum_{A\in\Psi} h(A) = 1\\
& \alpha(p)\geq 0 \qquad \forall p\in\mathcal{P}\\
& \sum_{p\in\mathcal{P}(u,v)}\alpha(p) = D_h(u,v) \qquad \forall u,v\in V.
\end{aligned}$$

[15] We treat $0/0 = 0$ here.

The objective is

$$\mu + \frac{1}{4\mu}\sum_{u\in V} C(u)^2$$

and by first derivative test the minimizer is $\mu = \sqrt{\sum_{u\in V} C(u)^2}/2$, attaining a minimum value of 2μ.

By now the subset flow interpretation should be clear: $h : \Psi \to \mathbb{R}_{\geq 0}$ is a distribution on Ψ, $D_h(u, v)$ is the corresponding demand between u and v in the subset flow problem, $\alpha(p)$ is the amount of flow along path p in the flow solution, which we denote by F. Then, the objective is exactly $\mathrm{con}(F)$, and the above constraints can be condensed into $F \in \mathcal{F}^\Psi(G)$. Therefore, the Lagrangian dual of the metric spreading maximization problem is

$$\min_{F\in\mathcal{F}^\Psi(G)} \mathrm{con}(F).$$

It remains to establish strong duality. Clearly, the primal maximization problem is concave. We can find a Slater point as follows: take $s(u) = 1/\sqrt{n}$ for all $u \in V$, $\delta(u, v) = 1/\sqrt{n}$ for all $u, v \in V$, and $\varepsilon = 1/(10\sqrt{n})$. This is feasible and all inequality constraints are strict. This completes the proof. □

Lemma 3.5.18 (Congestion Lower Bound [KLPT11, Corollary 3.6, Theorem 4.1])**.** *Let $G = (V, E)$, and let Ψ be a given collection of nonempty subsets of V. Let $h : \Psi \to \mathbb{R}_{\geq 0}$ be a distribution on Ψ, and let $\mathcal{F}^\Psi(G)$ be the set of all multicommodity flows on G whose demand graph satisfies*

$$D_h(u, v) = \sum_{A\in\Psi:\{u,v\}\subseteq A} \frac{h(A)}{|A|^2} \qquad \forall u, v \in V.$$

Write $M^\Psi := \sum_{A\in\Psi} h(A)/|A|^2$. Then, for any flow solution $F \in \mathcal{F}^\Psi(G)$,

- *(if G is planar) $\mathrm{con}(F) \gtrsim n^{-1/2}(M^\Psi)^{-1/4}$ if $M^\Psi < o(1)$,*
- *(if G has genus $g \geq 1$) $\mathrm{con}(F) \gtrsim (gn)^{-1/2}(M^\Psi)^{-1/4}$ if $gM^\Psi < o(1)$,*
- *(if G is K_h-minor free) $\mathrm{con}(F) \gtrsim (h^2 \log h \cdot n)^{-1/2}(M^\Psi)^{-1/4}$ if $(h^2 \log h)M^\Psi < o(1)$.*

Proof. We use the notion of "(c, a)-congested" from [KLPT11]. The exact definition do not matter; we only require the following facts:

- [KLPT11, Theorem 4.1] Let $h : \Psi \to \mathbb{R}_{\geq 0}$ be a distribution on Ψ, and let $\bar{\mathcal{F}}^\Psi(G)$ be the set of multicommodity flows on G whose demand graph satisfies

$$\bar{D}_h(u,v) = \sum_{A \in \Psi : \{u,v\} \subseteq A} h(A) \qquad \forall u, v \in V.$$

Then, for any $F \in \bar{\mathcal{F}}^\Psi(G)$,

$$\mathrm{con}(F)^2 \gtrsim \frac{1}{cn} \left(\sum_{S \in \Psi} h(S)|S|^2 \right)^{5/2} - c_0 \frac{a}{n} \left(\sum_{S \in \Psi} h(S)|S|^2 \right) \tag{3.16}$$

for some universal constant $c_0 > 0$.[16]

- [KLPT11, Corollary 3.6] If G is planar, then G is $(O(1), O(1))$-congested. If G has bounded genus $g \geq 1$, then G is $(O(g), O(\sqrt{g}))$-congested. If G is K_h-minor free, then G is $(O(h^2 \log h), O(h\sqrt{\log h}))$-congested.

To bridge the gap between $\mathcal{F}^\Psi(G)$ and $\bar{\mathcal{F}}^\Psi(G)$, define the following distribution

$$\bar{h}(A) := \frac{h(A)/|A|^2}{\sum_{S \in \Psi} h(S)/|S|^2} = \frac{h(A)}{|A|^2 M^\Psi}.$$

Then, any flow solution to $\mathcal{F}^\Psi(G)$ corresponding to distribution h is M^Ψ times a flow solution to $\bar{\mathcal{F}}^\Psi(G)$ corresponding to distribution $\bar{h}$, and so its congestion is also M^Ψ times that of the corresponding $\bar{\mathcal{F}}^\Psi(G)$ flow. We also note that

$$\sum_{S \in \Psi} \bar{h}(S)|S|^2 = \sum_{S \in \Psi} \frac{h(S)}{M^\Psi} = \frac{1}{M^\Psi}.$$

Now, we derive a congestion lower bound for each graph class.

- If G is planar, then G is $(O(1), O(1))$-congested. Taking $c = O(1)$ and $a = O(1)$ in (3.16), we have

$$\min_{F \in \mathcal{F}^\Psi(G)} \mathrm{con}(F) = M^\Psi \cdot \min_{F \in \bar{\mathcal{F}}^\Psi(G)} \mathrm{con}(F) \gtrsim \sqrt{\frac{(M^\Psi)^{-1/2}}{O(1) \cdot n} - \frac{O(1)}{n} M^\Psi},$$

which is at least $\Omega(n^{-1/2}(M^\Psi)^{-1/4})$ if $M^\Psi < o(1)$.

[16] To be exact, [KLPT11, Theorem 4.1] lower bounds a related quantity called "intersection number". The intersection number is at most the congestion measure, and so the theorem implies (3.16).

- If G has genus $g \geq 1$, then G is $(O(g), O(\sqrt{g}))$-congested. Taking $c = O(g)$ and $a = O(\sqrt{g})$ in (3.16), we have

$$\min_{F \in \mathcal{F}^{\Psi}(G)} \operatorname{con}(F) = M^{\Psi} \cdot \min_{F \in \bar{\mathcal{F}}^{\Psi}(G)} \operatorname{con}(F) \gtrsim \sqrt{\frac{(M^{\Psi})^{-1/2}}{O(g) \cdot n} - \frac{O(\sqrt{g})}{n} M^{\Psi}},$$

which is at least $\Omega((gn)^{-1/2}(M^{\Psi})^{-1/4})$ if $gM^{\Psi} < o(1)$.

- If G is K_h-minor free, then G is $(O(h^2 \log h), O(h\sqrt{\log h}))$-congested. Taking $c = O(h^2 \log h)$ and $a = O(h\sqrt{\log h})$ in (3.16), we have

$$\min_{F \in \mathcal{F}^{\Psi}(G)} \operatorname{con}(F) = M^{\Psi} \cdot \min_{F \in \bar{\mathcal{F}}^{\Psi}(G)} \operatorname{con}(F) \gtrsim \sqrt{\frac{(M^{\Psi})^{-1/2}}{O(h^2 \log h) \cdot n} - \frac{O(h\sqrt{\log h})}{n} M^{\Psi}},$$

which is at least $\Omega((h^2 \log h \cdot n)^{-1/2}(M^{\Psi})^{-1/4})$ if $(h^2 \log h)M^{\Psi} < o(1)$.

This completes the proof. □

Putting It All Together

Now we are ready to prove the eigenvalue upper bound in Theorem 3.5.12.

Proof of Theorem 3.5.12. Start with the eigenvalue upper bound in Lemma 3.5.15 that

$$\lambda'_k(G) \lesssim \frac{\Delta}{\varepsilon^2 n} \left(\beta_{\varepsilon/2}(V, d_s)\right)^2, \tag{3.17}$$

where $\varepsilon = \varepsilon_r(G, s)$ with $r = \lfloor n/4k \rfloor$.

- If G is planar, there are two cases to consider per Lemma 3.5.18. If $M^{\Psi_r} < o(1)$ where Ψ_r is the collection of all size-r vertex subsets, then by Lemma 3.5.17 and Lemma 3.5.18 it is possible to take $\varepsilon = \Theta(n^{-1/2}(M^{\Psi_r})^{-1/4})$. In this case, $M^{\Psi_r} = 1/r^2$, and $\varepsilon = \Theta(\sqrt{r/n}) = \Theta(1/\sqrt{k})$. Now by Lemma 3.5.16, $\beta_{\varepsilon/2}(V, d_s) = O(1)$. Plugging all the values in (3.17) we get

$$\lambda'_k(G) \lesssim \frac{\Delta k}{n}.$$

If $M^{\Psi_r} \geq \Omega(1)$, that means $r \leq O(1)$ and so $k \geq \Omega(n)$. Then, by the trivial upper bound that $\lambda'_k(G) \leq \lambda'_n(G) \leq 2\Delta$[17], we have

$$\lambda'_k(G) \leq 2\Delta \lesssim \frac{\Delta k}{n}.$$

- If G is of genus $g \geq 1$, there are again two cases to consider per Lemma 3.5.18. If $gM^{\Psi_r} < o(1)$, then we can apply the lemma and the metric/flow duality in Lemma 3.5.17 and take $\varepsilon = \Theta((gn)^{-1/2}(M^{\Psi_r})^{-1/4}) = \Theta(1/\sqrt{gk})$. Now by Lemma 3.5.16, $\beta_{\varepsilon/2}(V, d_s) = O(\log g)$. Plugging all the values in (3.17) we get

$$\lambda'_k(G) \lesssim \frac{\Delta k}{n} \cdot g \log^2 g.$$

If $gM^{\Psi_r} \geq \Omega(1)$, then $g \geq \Omega(r^2) = \Omega((n/k)^2)$. Then, by the trivial upper bound that $\lambda'_k(G) \leq \lambda'_n(G) \leq 2\Delta$, we have

$$\lambda'_k(G) \leq 2\Delta \lesssim \Delta \cdot \left(\frac{k}{n} \cdot g\right) \lesssim \frac{\Delta k}{n} \cdot g \log^2 g.$$

- If G is K_h-minor free, there are again two cases to consider per Lemma 3.5.18. If $(h^2 \log h)M^{\Psi_r} \leq O(1)$, then we can apply the lemma and the metric/flow duality in Lemma 3.5.17 and take $\varepsilon = \Theta((h^2 \log h \cdot n)^{-1/2}(M^{\Psi_r})^{-1/4}) = \Theta(1/\sqrt{(h^2 \log h) \cdot k})$. Now by Lemma 3.5.16, $\beta_{\varepsilon/2}(V, d_s) = O(h^2)$. Plugging all the values in (3.17) we get

$$\lambda'_k(G) \lesssim \frac{\Delta k}{n} \cdot h^6 \log h.$$

If $(h^2 \log h)M^{\Psi_r} \geq \Omega(1)$, then $(h^2 \log h) \geq \Omega(r^2) = \Omega((n/k)^2)$. Then, a similar argument as in the bounded genus case yields

$$\lambda'_k(G) \leq 2\Delta \lesssim \Delta \cdot \left(\frac{k}{n} \cdot h^2 \log h\right) \lesssim \frac{\Delta k}{n} \cdot h^6 \log h.$$

This concludes the proof. □

[17] This follows from a Rayleigh quotient characterization of $\lambda'_k(G)$ similar to Proposition 2.5.3, and that

$$R(f) = \frac{\sum_{uv \in E}(f(u) - f(v))^2}{\sum_{v \in V} f(v)^2} \leq 2\frac{\sum_{uv \in E}(f(u)^2 + f(v)^2)}{\sum_{v \in V} f(v)^2} \leq 2\Delta.$$

for all $f : V \to \mathbb{R}$.

3.6 Tighter Relaxations of Expansions

3.6.1 Adding Triangle Inequalities

One drawback of Cheeger's inequality is that the quality of the produced cut suffers from a square-root loss, since the guarantee of the Cheeger cut S is only that $\phi(S) \le O(\sqrt{\lambda_2(G)})$. This is especially problematic when the graph has small conductance. For instance, when G is the graph of two copies of $K_{n/2}$ connected by a perfect matching, $\phi(G) = \Theta(1/n)$ and $\lambda_2(G) = \Theta(1/n)$[18], so the Cheeger cut guarantee is only $\phi(S) \le O(1/\sqrt{n})$. This is a fundamental drawback of the spectral approach, and so a different approach is needed to obtain better low-conductance cuts.

The first breakthrough towards obtaining better low-conductance cuts is by Leighton and Rao [LR99]. They proved a generalization of the max-flow min-cut theorem that, for *multicommodity* flow problems on undirected graphs with product demands $D(u,v) = \pi(u)\pi(v)$, the gap between max-flow and min-cut is $O(\log n)$. The max-flow is then formulated as an LP, which can be solved and rounded to a cut in polynomial time. For any vertex weights π, the cut produced is guaranteed to have π-weighted edge expansion at most $O(\log n)$ times the minimum. This in particular implies $O(\log n)$ approximation for edge expansion and edge conductance. Refer to their paper for details.

The second breakthrough is by Arora, Rao, and Vazirani [ARV09] and gives the best worst-case approximation guarantee of $O(\sqrt{\log n})$ for edge expansion and (via a reduction; see Proposition 2.3.8) edge conductance of undirected graphs. They proved that the following Goemans-Linial SDP relaxation for undirected edge expansion ϕ' has an integrality gap of $O(\sqrt{\log n})$:

$$
\begin{aligned}
\min_{f:V\to\mathbb{R}^n} \quad & \frac{1}{n}\sum_{uv\in E} \|f(u)-f(v)\|^2 \\
\text{subject to} \quad & \sum_{u,v\in V} \|f(u)-f(v)\|^2 = n^2 \\
& \|f(u)-f(v)\|^2 + \|f(v)-f(u')\|^2 \ge \|f(u)-f(u')\|^2 \qquad \forall u,v,u'\in V.
\end{aligned}
\tag{3.18}
$$

This is indeed an SDP because both the objective and the constraints are linear in the Gram matrix $X(u,v) = \langle f(u), f(v)\rangle$. Note that this formulation without the ℓ_2^2 triangle inequalities in the last constraint is equivalent to the second smallest eigenvalue of the normalized Laplacian matrix when the graph is regular (see e.g. [Tre16]).

[18] One can see this by writing $G = K_{n/2} \times K_2$ and using the known facts about the spectrum of complete graphs and the spectrum of Cartesian products of graphs. One good reference is [BH11, Section 1.4].

A major contribution in [ARV09] is a structure theorem on vectors satisfying the ℓ_2^2 triangle inequalities. It asserts that, given a "well-spread" set of vectors satisfying the ℓ_2^2 triangle inequalities, there are two large subsets L and R, such that all vectors in L are far away from all vectors in R.

Definition 3.6.1 (Well-Spread Vectors)**.** *Let $k \in \mathbb{N}$ and let $f : V \to \mathbb{R}^k$ be a set of vectors that satisfy $\sum_{u,v \in V} \|f(u) - f(v)\|^2 = n^2$. Let $B(u,r) := \{v \in V : \|f(v) - f(u)\| < r\}$ denote the open ℓ_2-ball of radius r centered at $f(u)$. We say that $\{f(u)\}_{u \in V}$ is well-spread if $|B(u, \frac{1}{\sqrt{10}})| \leq \frac{n}{10}$ for all $u \in V$.*

Theorem 3.6.2 (ℓ_2^2 Structure Theorem [ARV09, Theorem 1])**.** *Let $\{f(u)\}_{u \in V}$ be a set of vectors[19] that satisfy the ℓ_2^2 triangle inequalities and with $\sum_{u,v \in V} \|f(u) - f(v)\|^2 = n^2$. If $\{f(u)\}_{u \in V}$ is well-spread, then there exist two sets $L, R \subseteq V$ such that $|L|, |R| \geq \Omega(n)$ and*

$$d(L,R) := \min_{u \in L, v \in R} \|f(u) - f(v)\|^2 \gtrsim 1/\sqrt{\log n}.$$

Moreover, there is a randomized polynomial-time algorithm that finds such sets with high probability.

The proof consists of novel geometric arguments involving measure concentration and chaining, which we refer the reader to [ARV09] for details. Using Theorem 3.6.2, they derived an $O(\sqrt{\log n})$ approximation algorithm for edge expansion.

Theorem 3.6.3 ($O(\sqrt{\log n})$ Approximation of Edge Expansion [ARV09])**.** *Let $G = (V, E)$ be an undirected graph and let $\lambda_2^{\Delta}(G)$ be the program in (3.18). Let $\varphi(G)$ be the edge expansion of G as defined in Section 2.3.1. Then,*

$$\frac{\varphi(G)}{\sqrt{\log n}} \lesssim \lambda_2^{\Delta}(G) \leq \varphi(G).$$

Proof. We first prove the easy direction that $\lambda_2^{\Delta}(G) \leq \varphi(G)$. Given a subset $S \subseteq V$ such that $0 < |S| \leq n/2$ and $\varphi(S) = \varphi(G)$, consider the following vector solution to $\lambda_2^{\Delta}(G)$:

$$f(u) := \begin{cases} \vec{0}, & \text{if } u \in S, \\ \left(\frac{n}{\sqrt{2 \cdot |S| \cdot |S^c|}}, 0, \ldots, 0 \right)^T, & \text{otherwise.} \end{cases}$$

[19] In [ARV09], the vectors $f(u)$ are assumed to be of unit length. We note that the structure theorem holds without this assumption as well; see for example [Rot16] for a writeup.

It is routine to check that this is feasible, and the objective value is

$$\frac{1}{n}\sum_{uv\in E}\|f(u)-f(v)\|^2 = |E(S,S^c)|\cdot\frac{n}{2\cdot|S|\cdot|S^c|}\leq \varphi(S),$$

since $|S^c|\geq n/2$.

Now we prove the hard direction that $\lambda_2^\Delta(G)\gtrsim \varphi(G)/\sqrt{\log n}$. The rounding technique first appeared in [LR99]. Let $f: V\to\mathbb{R}^n$ be an optimal solution to $\lambda_2^\Delta(G)$. There are two cases to consider, depending on whether f is well-spread or not per Definition 3.6.1.

- Suppose that f is well-spread. Then, by Theorem 3.6.2, there exist two sets $L,R\subseteq V$ such that $|L|,|R|\geq\Omega(n)$ and

  $$d(L,R) := \min_{u\in L, v\in R}\|f(u)-f(v)\|^2 \gtrsim 1/\sqrt{\log n}.$$

  Write D for $d(L,R)$. Consider the following threshold set determined by the distance of a vertex from L:

  $$S_t := \{u\in V : d(u,L) > t\},$$

  where $0\leq t < D$. We see that $L\subseteq S_t^c$ and $R\subseteq S_t$, so that

  $$\int_0^D \min(|S_t|,|S_t^c|)\,dt \geq D\min(|L|,|R|)\gtrsim\frac{n}{\sqrt{\log n}},$$

  and we also have

  $$\int_0^D |E(S_t,S_t^c)|\,dt \leq \sum_{uv\in E}|d(u,L)-d(v,L)| \overset{(*)}{\leq} \sum_{uv\in E} d(u,v) = n\lambda_2^\Delta(G).$$

  where $(*)$ uses that $d(u,v) := \|f(u)-f(v)\|^2$ satisfies triangle inequality. By the usual averaging argument, there exists a $t\in[0,D)$ such that by taking $S=S_t$ or $S=S_t^c$ we have $0<|S|\leq n/2$ and $\varphi(S)\lesssim\lambda_2^\Delta(G)\cdot\sqrt{\log n}$.

- Suppose that f is not well-spread, so that it has a large core $C := B(u_0,\frac{1}{\sqrt{10}})$ with $|C| > n/10$ for some $u_0\in V$. That means $d(u,u_0)\leq 1/10$ for all $u\in C$. Consider the following threshold set determined by the distance of a vertex from C:

  $$S_t := \{u\in S : d(u,C) > t\},$$

where $t \in \mathbb{R}_{\geq 0}$. We see that $C \subseteq S_t^c$ so that $|S_t^c| \geq |S_t|/9$, and so

$$\int_0^\infty \min(|S_t|, |S_t^c|)\, dt \geq \frac{1}{9} \int_0^\infty |S_t|\, dt = \frac{1}{9} \sum_{u \in V} d(u, C).$$

Recall that u_0 is the center of C, so that $d(u, C) \geq d(u, u_0) - 1/10$ for all $u \in V$. To lower bound $\sum_{u \in V} d(u, u_0)$ we use

$$2n \sum_{u \in V} d(u, u_0) = \sum_{u,v \in V} (d(u, u_0) + d(v, u_0)) \geq \sum_{u,v \in V} d(u, v) = n^2,$$

which gives finally that

$$\sum_{u \in V} d(u, C) \geq \sum_{u \in V} \left(d(u, u_0) - \frac{1}{10} \right) \geq n \left(\frac{1}{2} - \frac{1}{10} \right) \geq \frac{n}{3},$$

and so the "average" denominator satisfies

$$\int_0^\infty \min(|S_t|, |S_t^c|)\, dt \gtrsim n.$$

The "average" numerator bound is the same as before:

$$\int_0^\infty |E(S_t, S_t^c)|\, dt = \sum_{uv \in E} |d(u, C) - d(v, C)| \leq \sum_{uv \in E} d(u, v) = n \lambda_2^\Delta(G),$$

so in this case we can in fact find $S = S_t$ or $S = S_t^c$ such that $0 < |S| \leq n/2$ and $\varphi(S) \lesssim \lambda_2^\Delta(G)$.

Combining the two cases finishes the proof. □

3.6.2 Fast Algorithms

The approximation algorithm in [ARV09] is only guaranteed to run in polynomial time, while the Cheeger rounding algorithm for edge conductance runs in near-linear time. Subsequently, [AHK05, AK07, She09] bridged the gap by designing fast primal-dual algorithms to solve the SDP in (3.18). The theoretically fastest $O(\sqrt{\log n})$ approximation algorithm so far for edge expansion is by Sherman [She09], which requires solving $n^{o(1)}$ maximum flow problems, and by a recent breakthrough in maximum flow solvers [CKL$^+$22] the overall runtime of Sherman's algorithm is almost linear in the input size.

There are also fast $O(\log n)$-approximation algorithms for edge expansion (with arbitrary vertex weights) using the cut-matching game framework [KRV09, OSVV08]. These algorithms only require solving a polylogarithmic number of maximum s-t flows and run in almost-linear time [CKL$^+$22].

3.6.3 Tighter Programs for More General Settings

The ARV structure theorem of Theorem 3.6.2 is a general result about metric properties of vectors satisfying the ℓ_2^2 triangle inequalities, and has little to do with the underlying problem being the sparsest cut. By designing appropriate SDP relaxations of other expansion quantities that incorporate the ℓ_2^2 triangle inequalities, subsequent work apply the structure theorem to prove $O(\sqrt{\log n})$ integrality gaps, thus providing a randomized polynomial-time algorithm for approximating these expansion quantities to within $O(\sqrt{\log n})$ of the optimal.

Directed Edge Expansion

Agarwal, Charikar, Macharychev and Macharychev [ACMM05] designed an SDP for approximating directed edge expansion using a directed semi-metric. Their idea was to introduce an extra vector $f(0)$ to the embedding, and to define the semi-metric[20] as

$$\vec{d}(u,v) := \|f(u)-f(v)\|^2 - \|f(u)-f(0)\|^2 + \|f(v)-f(0)\|^2 \geq 0. \tag{3.19}$$

For a given directed graph $G=(V,E,w)$, their program $\mathsf{sdp}^{\Delta}(G)$ is formulated as follows:

$$\begin{aligned} \min_{f:V\cup\{0\}\to\mathbb{R}^n} \quad & \frac{1}{n}\sum_{uv\in E}\left(\|f(u)-f(v)\|^2 - \|f(u)-f(0)\|^2 + \|f(v)-f(0)\|^2\right) \\ \text{subject to} \quad & \|f(u)-f(v)\|^2 + \|f(v)-f(u')\|^2 \geq \|f(u)-f(u')\|^2 \quad \forall u,v,u' \in V\cup\{0\} \\ & \sum_{u,v\in V}\|f(u)-f(v)\|^2 = n^2. \end{aligned} \tag{3.20}$$

Note that taking $u'=0$ in the triangle inequality implies that $\vec{d}(u,v)\geq 0$. This is an SDP because the objective is a linear function in the Gram matrix $X(u,v)=\langle f(u),f(v)\rangle$ and the constraints are linear equalities and inequalities in X as well.

Their main result is to prove that $\mathsf{sdp}^{\Delta}(G)$ is an SDP relaxation of directed edge expansion $\varphi(G)$, and that the integrality gap is $O(\sqrt{\log n})$.

Theorem 3.6.4 ($O(\sqrt{\log n})$ Approximation of Directed Edge Expansion [ACMM05]). *Let $G=(V,E)$ be a directed graph and let $\mathsf{sdp}^{\Delta}(G)$ be the program in (3.20). Let $\vec{\varphi}(G)$ be the directed edge expansion of G as defined in Section 2.3.2. Then,*

$$\frac{\vec{\varphi}(G)}{\sqrt{\log n}} \lesssim \mathsf{sdp}^{\Delta}(G) \leq 2\vec{\varphi}(G).$$

[20] Note that $\vec{d}(u,v)=\vec{d}(v,u)$ may not hold.

Proof. We first prove the easy direction that $\mathsf{sdp}^{\Delta}(G) \leq \vec{\varphi}(G)$. Given a set $S \subseteq V$ such that $|\delta^+(S)|/\min(|S|,|S^c|) = \vec{\varphi}(G)$, consider the following vector solution $f : V \cup \{0\} \to \mathbb{R}^{n+1}$ to $\mathsf{sdp}^{\Delta}(G)$:

$$f(u) := \begin{cases} \vec{0}, & \text{if } u \in S \cup \{0\}, \\ \left(\frac{n}{\sqrt{2\cdot|S|\cdot|S^c|}}, 0, \ldots, 0\right)^T, & \text{otherwise.} \end{cases}$$

It is easy to check that f is feasible. To compute its objective value, note that for $uv \in E$, if $f(u) = f(v)$ then $\vec{d}(u,v) = 0$ ($\vec{d}$ as defined in (3.19)), and if $f(u) \neq f(v)$ then either $(u \in S, v \in S^c)$ or $(u \in S^c, v \in S)$. In the second case we again have $\vec{d}(u,v) = 0$, and in the first case $\vec{d}(u,v) = n^2/(|S|\cdot|S^c|)$. Therefore, the objective is

$$|\delta^+(S)| \cdot \frac{n^2}{|S|\cdot|S^c|} \leq 2\vec{\varphi}(S).$$

Next, we prove the hard direction that $\mathsf{sdp}^{\Delta}(G) \gtrsim \vec{\varphi}(G)/\sqrt{\log n}$. Let $f : V \cup \{0\} \to \mathbb{R}^{n+1}$ be an optimal solution to $\mathsf{sdp}^{\Delta}(G)$. Per the ARV structure theorem in Theorem 3.6.2, there are two cases to consider.

- Suppose that $\{f(u)\}_{u\in V}$ is well-spread. Then, by Theorem 3.6.2, there exists two sets $L, R \subseteq V$ such that $|L|, |R| \geq \Omega(n)$ and

$$d(L,R) := \min_{u\in L, v\in R} \|f(u) - f(v)\|^2 \gtrsim 1/\sqrt{\log n}.$$

  To relate to the directed semi-metric $\vec{d}$, we would like to impose order on the term $\|f(u) - f(0)\|^2$. Let $r \in \mathbb{R}_{\geq 0}$ be a median of $d(u,0)$ for $u \in L$ such that if

$$L^+ := \{u \in L : d(u,0) \geq r\} \qquad \text{and} \qquad L^- := \{u \in L : d(u,0) \leq r\},$$

  then $|L^+|, |L^-| \geq |L|/2$. Define R^+ and R^- similarly using the same threshold r.

  If $|R^+| \geq |R|/2$, take $L^* := L^-$ and $R^* := R^+$. Otherwise, take $L^* := L^+$ and $R^* := R^-$. By construction, we still have $|L^*|, |R^*| \geq \Omega(n)$ and $d(L^*, R^*) \geq \Omega(1/\sqrt{\log n})$. Moreover, either $d(u,0) \leq d(v,0)$ for all $u \in L^*, v \in R^*$, or $d(u,0) \geq d(v,0)$ for all $u \in L^*, v \in R^*$. Swapping L^* and R^* if necessary, assume that the former is the case, so that

$$\vec{d}(u,v) = d(u,v) - d(u,0) + d(v,0) \geq d(u,v) \gtrsim \frac{1}{\sqrt{\log n}}$$

for all $u \in L^*$ and $v \in R^*$. Consider the following threshold set determined by the directed distance from L^* to a vertex:

$$S_t := \{v \in V : \vec{d}(L^*, v) > t\}$$

Letting $D := \vec{d}(L^*, R^*)$, and considering $t \in [0, D)$, we see that $L^* \subseteq S_t^c$ and $R^* \subseteq S_t$, so that the "average" denominator satisfies

$$\int_0^D \min(|S_t|, |S_t^c|)\, dt \geq D \cdot \min(|L^*|, |R^*|) \gtrsim n/\sqrt{\log n}.$$

The "average" numerator satisfies

$$\begin{aligned}
\int_0^D \min(|\delta^+(S_t)|, |\delta^+(S_t^c)|)\, dt &\leq \int_0^D |\delta^+(S_t^c)|\, dt \\
&\leq \sum_{uv \in E} \max(0, \vec{d}(L^*, v) - \vec{d}(L^*, u)) \\
&\overset{(*)}{\leq} \sum_{uv \in E} \vec{d}(u, v) = n \cdot \mathsf{sdp}^{\Delta}(G).
\end{aligned}$$

The step $(*)$ requires explanation. Suppose $\vec{d}(L^*, v) > \vec{d}(L^*, u)$ and suppose $a \in L^*$ is such that $\vec{d}(a, u) = \vec{d}(L^*, u)$. Then

$$\vec{d}(L^*, v) - \vec{d}(L^*, u) \leq \vec{d}(a, v) - \vec{d}(a, u) = d(a, v) - d(a, u) + d(v, 0) - d(u, 0) \leq \vec{d}(u, v).$$

Since $\vec{d}(u, v) \geq 0$, we have shown that

$$\max(0, \vec{d}(L^*, v) - \vec{d}(L^*, u)) \leq \vec{d}(u, v) \tag{3.21}$$

for all $uv \in E$, and so $(*)$ goes through. Therefore, we can pick $S = S_t$ for some $t \in [0, D)$ so that $\vec{\varphi}(S) \lesssim \mathsf{sdp}^{\Delta}(G) \cdot \sqrt{\log n}$.

- Suppose that $\{f(u)\}_{u \in V}$ is not well-spread, so that there exists a large core $C := B(u_0, 1/\sqrt{10})$ (B is defined using ℓ_2 distance) with $|C| > n/10$ for some $u_0 \in V$. That means $d(u, u_0) \leq 1/10$ for all $u \in C$. Note that

$$\begin{aligned}
n \sum_{u \in V} \left(\vec{d}(u, u_0) + \vec{d}(u_0, u)\right) = 2n \sum_{u \in V} d(u, u_0) &= \sum_{u,v \in V} (d(u, u_0) + d(u_0, v)) \\
&\geq \sum_{u,v \in V} d(u, v) = n^2,
\end{aligned}$$

so that either $\sum_{u\in V} \vec{d}(u_0,u) \geq n/2$ or $\sum_{u\in V} \vec{d}(u,u_0) \geq n/2$.

Suppose for now that the former is true. Consider the following threshold sets determined by the directed distance from C to a vertex:

$$S_t := \{u \in V : \vec{d}(C,u) > t\}, \qquad t \in \mathbb{R}_{\geq 0}$$

Since $|C| > n/10$ and $C \subseteq S_t^c$, we have $|S_t^c| \geq |S_t|/9$ and so $\min(|S_t|,|S_t^c|) \geq |S_t|/9$ for all $t \in \mathbb{R}_{\geq 0}$. Then, the "average" denominator is

$$\begin{aligned}
\int_0^\infty \min(|S_t|,|S_t^c|)\,dt \geq \frac{1}{9}\int_0^\infty |S_t|\,dt &= \frac{1}{9}\sum_{u\in V} \vec{d}(C,u) \\
&\overset{(*)}{\geq} \frac{1}{9}\sum_{u\in V}\left(\vec{d}(u_0,u) - 2\max_{c\in C} d(u_0,c)\right) \\
&\geq \frac{1}{9}\left(\frac{n}{2}-\frac{n}{5}\right) = \frac{n}{30}.
\end{aligned}$$

The step $(*)$ goes through because for any $c \in C$,

$$\begin{aligned}
\vec{d}(c,u) &= d(c,u) - d(c,0) + d(u,0) \\
&\geq (d(u_0,u) - d(u_0,c)) - (d(u_0,0) + d(u_0,c)) + d(u,0) \\
&= \vec{d}(u_0,u) - 2d(u_0,c) \geq \vec{d}(u_0,u) - 2\max_{c'\in C} d(u_0,c').
\end{aligned}$$

The "average" numerator is

$$\begin{aligned}
\int_0^\infty \min(|\delta^+(S_t)|,|\delta^+(S_t^c)|)\,dt \leq \int_0^\infty |\delta^+(S_t^c)|\,dt &= \sum_{uv\in E} \max(0, \vec{d}(C,v) - \vec{d}(C,u)) \\
&\leq \sum_{uv\in E} \vec{d}(u,v) = n \cdot \mathsf{sdp}^\Delta(G),
\end{aligned}$$

where the last inequality follows again from (3.21). This establishes the existence of $S = S_t$ for some $t \in \mathbb{R}_{\geq 0}$, such that $\vec{\varphi}(S) \lesssim \mathsf{sdp}^\Delta(G)$.

If $\sum_{u\in V} \vec{d}(u,u_0) \geq n/2$, we define the threshold sets as

$$S_t := \{u \in V : \vec{d}(u,C) > t\},$$

and the proof proceeds similarly.

Combining the two cases finishes the proof. □

Vertex Expansion

Feige, Hajiaghayi, and Lee [FHL08] designed an $O(\sqrt{\log n})$ approximation algorithm for the minimum ratio vertex cut problem, which is equivalent to the minimum vertex expansion problem.[21] Their approach is again to design an SDP relaxation of the problem, then show that the integrality gap is $O(\sqrt{\log n})$.

For simplicity, we present the setting under uniform vertex weights. Let $G = (V, E)$ be a graph. A vertex separator of G divides the graph into three disjoint vertex subsets A, B, and S, where S is the separator and A and B are the separated. So, the requirements are that $V = S \sqcup A \sqcup B$ and $E(A, B) = \emptyset$. The sparsity of a vertex separator (A, B, S) is defined as

$$\alpha(A, B, S) := \frac{|S|}{|A \cup S| \cdot |B \cup S|},$$

and $\alpha(G)$ is the minimum sparsity over all vertex separators of G.

Their vector program relaxation for minimum sparsity vertex separator is as follows:

$$\begin{aligned}
\min_{a(u),b(u),z:V\to\mathbb{R}^{2n}} \quad & \frac{1}{K}\sum_{u\in V}(1 - \|a(u)\|^2 - \|b(u)\|^2) \\
\text{subject to} \quad & \|a(u)\|^2 + \|b(u)\|^2 \leq 1 \qquad \forall u \in V \\
& \langle a(u), b(u)\rangle = 0 \qquad \forall u \in V \\
& \langle a(u), b(v)\rangle = \langle a(v), b(u)\rangle = 0 \qquad \forall uv \in E \\
& \|z\|^2 = 1 \\
& \langle z, a(u)\rangle = \|a(u)\|^2, \ \langle z, b(u)\rangle = \|b(u)\|^2 \qquad \forall u \in V \\
& \sum_{u,v\in V} \|a(u) - a(v)\|^2 = 2K \\
& \|a(u) - a(v)\|^2 + \|a(v) - a(u')\|^2 \geq \|a(u) - a(u')\|^2 \qquad \forall u, v, u' \in V
\end{aligned}$$

where K is a parameter to be tuned. The motivation for this program is the integer program where $a(u), b(u), s(u) \in \{0, 1\}$ indicate membership in sets A, B, S respectively.

The proof that the integrality gap is $O(\sqrt{\log n})$ is by applying the ℓ_2^2 structure theorem [ARV09] to the vectors $a(u)$. Refer to their paper for more details.

[21] It is essential that $\min(1, \cdot)$ is part of the definition of vertex expansion $\psi(G)$ in Section 2.3.1.

Directed Hypergraph Expansion

Chan and Sun [CS18] gave an approximation algorithm for directed hypergraph expansion (see Section 2.2 and Section 2.3.3 for the relevant definitions) via a reduction to directed edge expansion [ACMM05].

Definition 3.6.5 (Derived Graph of Directed Hypergraphs [CS18, Fact 1.1])**.** *Let $H = (V, E, w)$ be a directed hypergraph over vertex measure $\pi : V \to \mathbb{R}^+$. The derived graph $G_H = (V', E', w')$ over vertex measure $\pi' : V' \to \mathbb{R}^+$ is defined as follows:*

- $V' := V \cup \{u_e^- : e \in E\} \cup \{u_e^+ : e \in E\}$
- $E' := \{(v, u_e^-) : v \in e^-, e \in E\} \cup \{(u_e^-, u_e^+) : e \in E\} \cup \{(u_e^+, v') : v' \in e^+, e \in E\}$
- $w'(v, u_e^-) = w'(u_e^+, v') = \infty$ *and* $w'(u_e^-, u_e^+) = w(e)$ *for all* $e \in E$, $(v, v') \in e^- \times e^+$
- $\pi'(u) = \pi(u)$ *for all* $u \in V$, *and* $\pi'(u_e^-) = \pi'(u_e^+) = 0$ *for all* $e \in E$.

Refer to Figure 3.4 for an illustration.

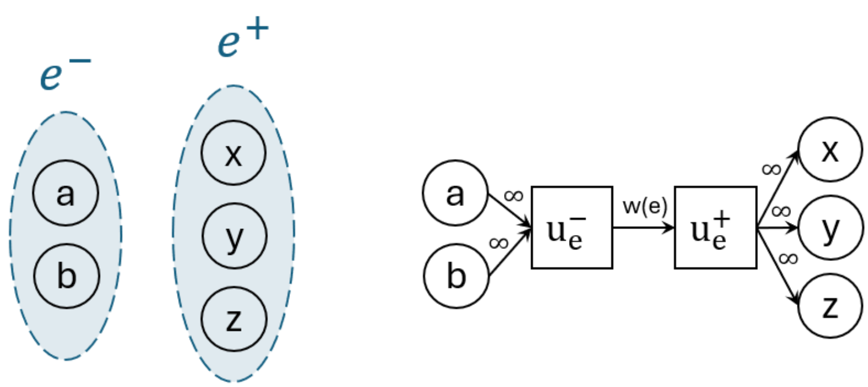


Figure 3.4: The construction described in Definition 3.6.5. On the left is a directed hyperedge in H with weight $w(e)$ and on the right is the component in the derived graph G_H corresponding to e, where the number above an arc indicates its weight. The new nodes u_e^- and u_e^+ have vertex measure zero and all other nodes in G_H inherit their original measures in H.

From [CS18, Fact 1], there is a correspondence between subsets $S \subseteq V$ and $S' \subseteq V'$ so that $\vec{\phi}_\pi(S) = \Theta\left(\vec{\phi}_{\pi'}(S')\right)$. Therefore, if we perform a black-box reduction from $\vec{\phi}_\pi(H)$ to

$\vec{\phi}_{\pi'}(G_H)$ and apply Theorem 3.6.4, we obtain an $O(\sqrt{\log |V'|})$ approximation algorithm for directed hypergraph expansion. However, the approximation guarantee using this approach degrades to $O(\sqrt{\log(n+m)})$ (since $|V'| = \Theta(n+m)$), which is worse when $m = \omega(\text{poly}(n))$.

Summary

While these results yield the best-known approximation ratios in their respective settings, the design of the SDP relaxations appears to be rather disparate. It is desirable to use one single recipe to derive SDP relaxations with $O(\sqrt{\log n})$ integrality gaps for generalized expansion quantities.

3.6.4 Orthogonal Separators and Higher Expansions

The study of k-way conductance $\phi_k(G)$ in undirected graphs is motivated by the higher-order Cheeger inequality [LOT12, LRTV12] connecting $\phi_k(G)$ to the k-th smallest eigenvalue $\lambda_k(G)$. Louis and Makarychev [LM14a] studied the related but different problem of sparsest k-way partitioning $\Phi_k(G)$ and designed an $O(\sqrt{\log n \log k})$-approximation algorithm that runs in randomized polynomial time. To obtain their result, they designed an SDP relaxation of $\Phi_k(G)$ which, again, incorporated the ℓ_2^2 triangle inequalities. To obtain disjoint subsets with small conductance from the vector solution to the SDP relaxation, they used the technique of orthogonal separators, originally devised to study Unique Games [CMM06] and later applied to small-set expansion and other graph partitioning problems [BFK+14]. We first introduce the problem of sparsest k-partitioning and their SDP relaxation of the problem. Next, we define orthogonal separators, then state and prove their approximation guarantee.

Definition 3.6.6 (Sparsest k-Partitioning)**.** *Let $G = (V, E)$ be a graph, and let $2 \le k \le n$. The sparsest k-partitioning of G is defined as*

$$\Phi_k(G) := \min_{S_1 \sqcup \cdots \sqcup S_k = V} \Phi_k(S_1, \ldots, S_k) \quad \textit{where} \quad \Phi_k(S_1, \ldots, S_k) := \max_{i \in [k]} \frac{|E(S_i, S_i^c)|}{\text{vol}(S_i)}.$$

Comparing the definition of $\Phi_k(G)$ with the definition of k-way conductance $\phi_k(G)$ in Definition 3.1.5, the only difference is the collection of sets over which the minimum is taken. Since every k-set partitioning is a collection of k disjoint subsets, we necessarily have $\phi_k(G) \le \Phi_k(G)$. The gap can, however, be large.

Definition 3.6.7 (SDP Relaxation of $\phi_k(G)$ [LM14a]). *Let $G = (V, E)$ be a graph. The SDP relaxation of $\phi_k(G)$ with ℓ_2^2 triangle inequalities, denoted $\mathsf{sdp}_k^{\Delta}(G)$, is defined as follows:*

$$\begin{aligned}
\min_{f:V\to\mathbb{R}^n} \quad & \frac{1}{k}\sum_{uv\in E} \|f(u)-f(v)\|^2 \\
\text{subject to} \quad & \sum_{u\in V} \deg(u)\,\|f(u)\|^2 = k \\
& \sum_{v\in V} \deg(v)\langle f(u), f(v)\rangle = 1 \qquad \forall u \in V \\
& \|f(u)-f(v)\|^2 + \|f(v)-f(u')\|^2 \geq \|f(u)-f(u')\|^2 \qquad u, v, u' \in V \\
& 0 \leq \langle f(u), f(v)\rangle \leq \|f(u)\|^2 \qquad \forall u, v \in V.
\end{aligned}$$

The second constraint $\sum_{v\in V} \deg(v)\langle f(u), f(v)\rangle = 1$ acts as a "spreading" constraint, which asserts that the vectors $f(v)$ cannot be too concentrated in one direction. The last constraint is equivalent to the ℓ_2^2 triangle inequality applied to $f(u)$, $f(v)$ and $\vec{0}$.

Note that this is an SDP because the objective function and all the constraints are linear in the Gram matrix $X(u, v) = \langle f(u), f(v)\rangle$. Note also that by changing $\deg(u)$ to an arbitrary vertex weight function $\pi(u)$, we can use $\mathsf{sdp}_k^{\Delta}(G)$ to approximate the following π-weighted sparsest k-partitioning:

$$(\Phi_\pi)_k(S_1, \ldots, S_k) := \max_{i\in[k]} \frac{|E(S_i, S_i^c)|}{\pi(S_i)} \qquad \text{and} \qquad (\Phi_\pi)_k(G) := \min_{S_1\sqcup\cdots\sqcup S_k = V} (\Phi_\pi)_k(S_1, \ldots, S_k).$$

First, they have to show that $\mathsf{sdp}_k^{\Delta}(G)$ is indeed a relaxation of $\Phi_k(G)$.

Proposition 3.6.8 (Easy Direction [LM14a]). *Let $G = (V, E)$ be an undirected graph. Then,*

$$\mathsf{sdp}_k^{\Delta}(G) \leq \Phi_k(G).$$

Proof. Suppose there is a partitioning $S_1, \ldots, S_k$ such that $\Phi_k(S_1, \ldots, S_k) = \Phi_k(G)$. Define the following vector solution $f : V \to \mathbb{R}^n$:

$$f_i(u) := \begin{cases} \frac{1}{\sqrt{\mathrm{vol}(S_i)}}, & \text{if } u \in S_i, \\ 0, & \text{otherwise.} \end{cases}$$

Here $f_i(u)$ denotes the i-th coordinate of $f(u)$. It is routine to check that this is a feasible solution (crucially we need $\cup_{i\in[k]} S_i = V$), and the objective value is

$$\begin{aligned}\frac{1}{k}\sum_{uv\in E}\|f(u)-f(v)\|^2 &= \frac{1}{k}\sum_{i\in[k]}\sum_{uv\in E}(f_i(u)-f_i(v))^2\\ &= \frac{1}{k}\sum_{i\in[k]}\frac{|E(S_i,S_i^c)|}{\mathrm{vol}(S_i)} \le \max_{i\in[k]}\frac{|E(S_i,S_i^c)|}{\mathrm{vol}(S_i)} = \Phi_k(G).\end{aligned}$$

This completes the proof. □

As mentioned before, their rounding algorithm utilizes orthogonal separators. Informally, orthogonal separators is a way to sample vertex subsets $S \subseteq V$, such that two vertices are unlikely to be both included in S if their embedding vectors are close to orthogonal, and two vertices are unlikely to be separated by S (i.e. one included, one excluded) if their embedding vectors are close to one another. The formal definition is below.

Definition 3.6.9 (Orthogonal Separators [CMM06, BFK+14, LM14a])**.** *Let $f : X \to \mathbb{R}^d$ be a collection of vectors that satisfy the ℓ_2^2 triangle inequalities. We say that a distribution over subsets of X is an s-orthogonal separator of X with distortion D, probability scale $\alpha > 0$, and separation threshold $\beta < 1$, if the following conditions hold for $S \subseteq X$ chosen according to this distribution:*

1. *For all $u \in X$, $\Pr[u \in S] = \alpha\|f(u)\|^2$.*
2. *For all $u, v \in X$ with $\langle f(u), f(v)\rangle \le \beta \max\{\|f(u)\|^2, \|f(v)\|^2\}$,*
$$\Pr[u, v \in S] \le \frac{\alpha \min\{\|f(u)\|^2, \|f(v)\|^2\}}{s}.$$
3. *For all $u, v \in X$,*
$$\Pr[\mathbb{1}_S(u) \neq \mathbb{1}_S(v)] \le \alpha D\|f(u)-f(v)\|^2,$$
where $\mathbb{1}_S$ is the indicator function of S.

The following theorem asserts the existence of orthogonal separators with small distortion.

Theorem 3.6.10 (Existence of Orthogonal Separators [CMM06, BFK+14, LM14a])**.** *There exists a randomized polynomial-time algorithm that, given $f : X \to \mathbb{R}^d$ satisfying the ℓ_2^2 triangle inequalities, and parameters s and $\beta < 1$, generates an s-orthogonal separator with distortion $D = O_\beta(\sqrt{\log n \log s})$ and scale $\alpha \ge 1/\operatorname{poly}(n)$.*

They also need to define a normalization of the solution vectors $f(u)$. Define $f(0) := \vec{0}$.

Proposition 3.6.11 (Normalization [CMM06])**.** *Let $f : V \cup \{0\} \to \mathbb{R}^n$ be vectors satisfying the ℓ_2^2 triangle inequality. Then, there is a mapping $f(u) \mapsto \bar{f}(u)$ to an inner product space, such that:*

- $\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 + \left\|\bar{f}(v) - \bar{f}(u')\right\|^2 \geq \left\|\bar{f}(u) - \bar{f}(u')\right\|^2$ *for all $u, v, u' \in V$;*
- *For all $u, v \in V \cup \{0\}$,*
$$\langle \bar{f}(u), \bar{f}(v)\rangle = \frac{\langle f(u), f(v)\rangle}{\max(\|f(u)\|^2, \|f(v)\|^2)};$$
- $\left\|\bar{f}(u)\right\|^2 = 1$ *if $f(u) \neq \vec{0}$;*
- *If $\langle f(u), f(v)\rangle = 0$, then $\langle \bar{f}(u), \bar{f}(v)\rangle = 0$;*
- *For all nonzero vectors $f(u)$ and $f(v)$,*
$$\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 \leq \frac{2\left\|f(u) - f(v)\right\|^2}{\max(\|f(u)\|^2, \|f(v)\|^2)}.$$

The rounding algorithm, as detailed in Algorithm 2, is by generating many (random) orthogonal separators with appropriate parameters, then extract k disjoint sparse cuts from them. A postprocessing procedure is required to turn this into a partitioning. We define the mass of vertex subsets $S \subseteq V$ as

$$\mu(S) := \sum_{u \in S} \deg(u) \left\|f(u)\right\|^2.$$

Note that $\mu(V) = k$.

The "hard direction" of the main theorem amounts to showing that the $\ell = \lfloor(1-\varepsilon)k\rfloor$ sets generated by Algorithm 2 all have small edge conductance. To this end, they prove the following guarantees about the algorithm.

Lemma 3.6.12 (Key Properties of Algorithm 2 [LM14a])**.** *Let $G = (V, E)$ be an undirected graph. Then, the following guarantees about Algorithm 2 hold.*

(a) For every vertex $u \in V$ and $i \in \{1, \ldots, T\}$, we have $\Pr[u \in S_i'] \geq \alpha/2$.

Algorithm 2 Rounding Algorithm using Orthogonal Separators

Input: Graph $G = (V, E)$, solution $f : V \to \mathbb{R}^n$ to $\mathsf{sdp}_k^{\Delta}(G)$, parameters k, ε
Output: $\ell := \lfloor (1-\varepsilon)k \rfloor$ disjoint subsets $S_1, \ldots, S_\ell \subseteq V$

1: Compute the normalization $\bar{f}$ from f per Proposition 3.6.11
2: Sample $T = 2n/\alpha$ independent $(12k/\varepsilon)$-orthogonal separators $S_1, \ldots, S_T$ for vectors $\bar{f}(u)$ and some choice of $\alpha \geq 1/\operatorname{poly}(n)$, with separation threshold $\beta = 1 - \varepsilon/4$ and distortion $D = O_\beta(\sqrt{\log n \log k})$
3: For each $i \in [T]$, define $S_i' := S_i$ if $\mu(S_i) \leq 1 + \varepsilon/2$ and $S_i' := \emptyset$ otherwise
4: For each $i \in [T]$, let $S_i'' := S_i' \setminus (\cup_{j<i} S_j')$
5: For each $i \in [T]$, let $P_i := \{u \in S_i'' : \|f(u)\|^2 \geq t_i\}$, where t_i is chosen to minimize $\phi(P_i)$
6: **return** the ℓ sets from P_i with the smallest edge conductance $\phi(P_i)$

(b) All sets S_i'' are disjoint, and

$$\Pr[\mu(\cup S_i'') = k] \geq 1 - ne^{-n}.$$

(c) For a set $S \subseteq V$, define

$$\nu(S) := \sum_{\substack{uv \in E \\ u \in S, v \notin S}} \|f(u)\|^2 + \sum_{\substack{uv \in E \\ u,v \in S}} \left| \|f(u)\|^2 - \|f(v)\|^2 \right|.$$

Then,

$$\mathbb{E}\left[\sum_{i \in [T]} \nu(S_i'')\right] \lesssim kD \cdot \mathsf{sdp}_k^{\Delta}(G).$$

Proof. To prove (a), note that we apply orthogonal separators to the normalized vectors $\bar{f}(u)$. Due to the second constraint of $\mathsf{sdp}_k^{\Delta}(G)$ that

$$\sum_{v \in V} \deg(v) \langle f(u), f(v) \rangle = 1$$

for all $u \in V$, we have $f(u) \neq \vec{0}$ for all $u \in V$, and so $\Pr[u \in S_i] = \alpha \left\| \bar{f}(u) \right\|^2 = \alpha$ by Proposition 3.6.11. Thus, it suffices to prove that

$$\Pr[\mu(S_i) \leq 1 + \varepsilon/2 \,|\, u \in S_i] \geq 1/2$$

for all $i \in [T]$. Let $B := \{v \in V : \langle \bar{f}(u), \bar{f}(v) \rangle \le \beta\}$. Whenever $v \in B$, we have

$$\Pr[v \in S_i \,|\, u \in S_i] = \frac{\Pr[u, v \in S_i]}{\Pr[u \in S_i]} \le \frac{\varepsilon}{12k}$$

by the second property of orthogonal separators. Since $\mu(B) \le \mu(V) = k$, by linearity of expectation and Markov's inequality it follows that

$$\Pr\left[\mu(S_i \cap B) \le \frac{\varepsilon}{6} \,|\, u \in S_i\right] \ge \frac{1}{2}.$$

Next, use the spreading constraint to show that most of the vectors are in B. Indeed,

$$\begin{aligned} 1 &= \sum_{v \in V} \deg(v) \langle f(v), f(u) \rangle \\ &= \sum_{v \in V} \deg(v) \langle \bar{f}(v), \bar{f}(u) \rangle \cdot \max(\|f(v)\|^2, \|f(u)\|^2) \qquad \text{(by Proposition 3.6.11)} \\ &\ge \sum_{v \notin B} \left[\deg(v) \|f(v)\|^2 \cdot \left(1 - \frac{\varepsilon}{4}\right)\right], \end{aligned}$$

which implies that $\mu(V \setminus B) \le 1 + \varepsilon/3$. To summarize, with $S_i = (S_i \cap B) \cup (S_i \setminus B)$,

$$\Pr\left[\mu(S_i) \le 1 + \frac{\varepsilon}{2} \,|\, u \in S_i\right] \ge \Pr\left[\mu(S_i \cap B) \le \frac{\varepsilon}{6} \,|\, u \in S_i\right] \ge \frac{1}{2}.$$

This completes the proof of (a). (b) immediately follows since $\cup S''_i = \cup S'_i$ and the probability that a vertex $u \in V$ is not in any S'_i is at most

$$\left(1 - \frac{\alpha}{2}\right)^T = \left(1 - \frac{\alpha}{2}\right)^{\frac{2n}{\alpha}} \le e^{-n}.$$

A union bound yields the result.

To prove (c), consider the following partitioning of the vertex set

$$V = S''_1 \sqcup S''_2 \sqcup \cdots \sqcup S''_T \sqcup (V \setminus (\cup S''_i)).$$

If an edge $uv \in E$ is cut by the partitioning, then the contribution to $\sum_{i \in [T]} \nu(S''_i)$ is at most $\|f(u)\|^2 + \|f(v)\|^2$. Otherwise, the contribution is at most $|\, \|f(u)\|^2 - \|f(v)\|^2 \,|$. Let E_{cut} be the set of all edges cut by the partitioning. Then,

$$\mathbb{E}\left[\sum_{i \in [T]} \nu(S''_i)\right] \le \mathbb{E}\left[\sum_{uv \in E_{cut}} \left(\|f(u)\|^2 + \|f(v)\|^2\right)\right] + \sum_{uv \in E} |\, \|f(u)\|^2 - \|f(v)\|^2 \,|. \quad (3.22)$$

By ℓ_2^2 triangle inequality, the latter is at most $\sum_{uv\in E} \|f(u)-f(v)\|^2 = k\cdot \mathsf{sdp}_k^{\Delta}(G)$. To bound the former, we say that $uv \in E_{cut}$ is cut by S_i' if i is the smallest index that uv is cut by S_i', i.e. $\mathbb{1}_{S_i'}(u) \neq \mathbb{1}_{S_i'}(v)$. Then, for any edge $uv \in E$,

$$\begin{aligned}\Pr[uv \in E_{cut}] = \sum_{i=1}^{T} \Pr\left[uv \text{ is cut by } S_i'\right] &= \sum_{i=1}^{T} \Pr\left[u,v \notin \cup_{j<i} S_j' \text{ and } \mathbb{1}_{S_i'}(u) \neq \mathbb{1}_{S_i'}(v)\right] \\ &= \sum_{i=1}^{T} \Pr\left[u,v \notin \cup_{j<i} S_j'\right] \cdot \Pr\left[\mathbb{1}_{S_i'}(u) \neq \mathbb{1}_{S_i'}(v)\right].\end{aligned}$$

By (a), the first term in the i-th summand is bounded above by

$$\left(1-\frac{\alpha}{2}\right)^{i-1},$$

and by Proposition 3.6.11 and properties of orthogonal separators, the second term in the i-th summand has the uniform upper bound of

$$\Pr\left[\mathbb{1}_{S_i'}(u) \neq \mathbb{1}_{S_i'}(v)\right] \leq \alpha D \left\|\bar{f}(u)-\bar{f}(v)\right\|^2 \leq \frac{2\alpha D \|f(u)-f(v)\|^2}{\max\{\|f(u)\|^2, \|f(v)\|^2\}}.$$

Summing up we then get

$$\Pr[uv \in E_{cut}] \leq \left(\sum_{i=1}^{T}\left(1-\frac{\alpha}{2}\right)^{i-1}\right) \cdot \frac{2\alpha D \|f(u)-f(v)\|^2}{\max\{\|f(u)\|^2, \|f(v)\|^2\}} \lesssim \frac{D \|f(u)-f(v)\|^2}{\max\{\|f(u)\|^2, \|f(v)\|^2\}}.$$

Continuing from (3.22),

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in[T]} \nu(S_i'')\right] &\leq \mathbb{E}\left[\sum_{uv\in E_{cut}} \left(\|f(u)\|^2 + \|f(v)\|^2\right)\right] + k\cdot \mathsf{sdp}_k^{\Delta}(G) \\ &\lesssim \sum_{uv\in E}\left[\frac{D\|f(u)-f(v)\|^2}{\max\{\|f(u)\|^2, \|f(v)\|^2\}} \cdot \left(\|f(u)\|^2 + \|f(v)\|^2\right)\right] + k\cdot \mathsf{sdp}_k^{\Delta}(G) \\ &\lesssim D\cdot \sum_{uv\in E} \|f(u)-f(v)\|^2 + k\cdot \mathsf{sdp}_k^{\Delta}(G) \lesssim kD\cdot \mathsf{sdp}_k^{\Delta}(G),\end{aligned}$$

completing the proof of (c). □

Theorem 3.6.13 (Hard Direction [LM14a]). *Let $G = (V, E)$ be an undirected graph. Let $2 \le k \le n$ and $\varepsilon \in (0, 1)$. Then,*

$$\phi_{\lfloor (1-\varepsilon)k \rfloor}(G) \le O_\varepsilon(\sqrt{\log n \log k} \cdot \mathsf{sdp}_k^\Delta(G))$$

Moreover, there is a randomized polynomial-time algorithm that produces $\ell = \lfloor (1-\varepsilon)k \rfloor$ sets $P_1, \dots, P_\ell$ such that $\phi_\ell(P_1, \dots, P_\ell)$ satisfies the same upper bound.

Proof. We run Algorithm 2. Since Lemma 3.6.12(b) asserts that the probability that $\cup S_i'' \neq V$ is exponentially low, and that S_i'' are pairwise disjoint by construction, we may assume that $\{S_i''\}$ is indeed a partitioning of V. Let

$$Z := \frac{\sum_{i \in [T]} \nu(S_i'')}{\sum_{i \in [T]} \mu(S_i'')} = \frac{1}{k} \sum_{i \in [T]} \nu(S_i'').$$

The correct way to view $\nu(S_i'')$ is the expected edge boundary size under a certain threshold rounding scheme, and so the following definition

$$\mathcal{I} := \left\{ i \in [T] : S_i'' \neq \emptyset \text{ and } \frac{\nu(S_i'')}{\mu(S_i'')} \le \frac{2Z}{\varepsilon} \right\}$$

chooses precisely those sets that will round to a small conductance set. Let $S_{\mathcal{I}}'' := \cup_{i \in \mathcal{I}} S_i''$. The total mass of sets S_i'' for i *outside* $\mathcal{I}$ is upper bounded by

$$\sum_{i \notin \mathcal{I}} \mu(S_i'') < \frac{\varepsilon}{2Z} \sum_{i \notin \mathcal{I}} \nu(S_i'') \le \frac{\varepsilon}{2Z} \cdot kZ,$$

and so

$$\mu(S_{\mathcal{I}}'') \ge k - \frac{k\varepsilon}{2} = k\left(1 - \frac{\varepsilon}{2}\right).$$

Since $\mu(S_i'') \le (1 + \varepsilon/2)$ for all $i \in [T]$, the set $\mathcal{I}$ has at least $\ell = \lfloor (1-\varepsilon)k \rfloor$ elements.

Now, for the threshold rounding, for any $i \in [T]$, let $M_i := \max\{\|f(u)\|^2 : u \in S_i''\}$ and define threshold sets

$$U_t := \{u \in S_i'' : \|f(u)\|^2 > t\}$$

for $t \in [0, M_i]$. The "average" denominator is

$$\int_0^{M_i} \mathrm{vol}(U_t)\, dt = \sum_{u \in S_i''} \deg(u) \|f(u)\|^2 = \mu(S_i'').$$

Let $uv \in E$. We investigate when uv is cut by U_t.

- If $u, v \notin S_i''$, then $uv \notin \delta(U_t)$ for any $t \in [0, M_i]$.
- If $u \in S_i''$ and $v \notin S_i''$, then $uv \in \delta(U_t)$ if and only if $t < \|f(u)\|^2$.
- If $u, v \in S_i''$, then $uv \in \delta(U_t)$ if and only if

$$\min(\|f(u)\|^2, \|f(v)\|^2) \leq t < \max(\|f(u)\|^2, \|f(v)\|^2).$$

Therefore, the "average" numerator is

$$\int_0^{M_i} |\delta(U_t)|\, dt = \sum_{\substack{uv \in E \\ u \in S_i'', v \notin S_i''}} \|f(u)\|^2 + \sum_{\substack{uv \in E \\ u, v \in S_i''}} \left| \|f(u)\|^2 - \|f(v)\|^2 \right| = \nu(S_i'').$$

Therefore, for $i \in \mathcal{I}$, P_i in step 5 of Algorithm 2 satisfies

$$\phi(P_i) \leq \frac{\nu(S_i'')}{\mu(S_i'')} \leq \frac{2Z}{\varepsilon}.$$

Since $\mathbb{E}[Z] \leq O(D \cdot \mathsf{sdp}_k^{\Delta}(G))$ by Lemma 3.6.12(c), with probability at least $1/2$ picking the ℓ sets $P_{i(1)}, \dots, P_{i(\ell)}$ with the smallest conductance we have (note that β is a function of ε)

$$\phi_\ell(P_{i(1)}, \dots, P_{i(\ell)}) \leq O_\beta(\varepsilon^{-1}\sqrt{\log n \log(12k/\varepsilon)} \cdot \mathsf{sdp}_k^{\Delta}(G)) = O_\varepsilon(\sqrt{\log n \log k} \cdot \mathsf{sdp}_k^{\Delta}(G)),$$

as required. Finally, a careful but straightforward inspection of Algorithm 2 verifies that the algorithm indeed runs in randomized polynomial time. □

The final step is to obtain a partitioning of V from the disjoint subsets.

Corollary 3.6.14 ([LM14a])**.** *Let $G = (V, E)$ be an undirected graph. Let $2 \leq k \leq n$ and $\varepsilon \in (0, 1)$. Then,*

$$\Phi_{\lfloor (1-\varepsilon)k \rfloor}(G) \leq O_\varepsilon(\sqrt{\log n \log k} \cdot \mathsf{sdp}_k^{\Delta}(G)) \leq O_\varepsilon(\sqrt{\log n \log k} \cdot \Phi_k(G)).$$

Moreover, there is a randomized polynomial-time algorithm that produces $\ell = \lfloor (1-\varepsilon)k \rfloor$ sets $P_1, \dots, P_\ell$ such that $P_1 \sqcup \dots \sqcup P_\ell = V$ and $\Phi_\ell(P_1, \dots, P_\ell)$ satisfies the same upper bound.

Proof. Run Algorithm 2 with $\varepsilon' = \varepsilon/2$ and Theorem 3.6.13 guarantees $\ell' := \lfloor (1-\varepsilon')k \rfloor$ subsets $P'_1, \ldots, P'_{\ell'}$ such that

$$\max_{i\in[\ell']} \phi(P'_i) \le O_\varepsilon(\sqrt{\log n \log k} \cdot \mathsf{sdp}_k^\Delta(G)) := W.$$

Take the $\ell - 1$ subsets $P_1, \ldots, P_{\ell-1}$ with the smallest volumes and let P_ℓ be the set containing all the remaining vertices. We upper bound $\phi(P_\ell)$ as follows:

$$\begin{aligned}\phi(P_\ell) &= \frac{|E(P_\ell, P_\ell^c)|}{\mathrm{vol}(P_\ell)} \\ &\le \frac{\sum_{i\in[\ell-1]} |E(P_i, P_i^c)|}{\frac{\ell'-\ell+1}{\ell'}\mathrm{vol}(V)} \\ &\le \frac{W \cdot \sum_{i\in[\ell-1]} \mathrm{vol}(P_i)}{\varepsilon \cdot \mathrm{vol}(V)} \le \frac{W}{\varepsilon}.\end{aligned}$$

This together with Proposition 3.6.8 completes the proof. □

Hypergraph Orthogonal Separators

Louis and Makarychev [LM14b] defined a variant of orthogonal separators, which they call hypergraph orthogonal separators, to study approximation algorithms for small-set hypergraph expansion and (via a reduction; see Section 2.3.4) small-set vertex expansion. Once the existence of hypergraph orthogonal separators with the desired parameters is established, the rounding algorithm and its analysis are very similar to those in [LM14a]. Therefore, we present the definition of hypergraph orthogonal separators, state the existance claim and the main approximation result, and leave the interested reader to [LM14b] for details.

Definition 3.6.15 (Hypergraph Orthogonal Separators [LM14b])**.** *Let $f : X \to \mathbb{R}^d$ be a collection of vectors that satisfy the ℓ_2^2 triangle inequalities. We say that a distribution over subsets of X is a hypergraph s-orthogonal separator of X with distortion D, probability scale $\alpha > 0$, and separation threshold $\beta < 1$, if the following conditions hold for $S \subseteq X$ chosen according to this distribution:*

1. *For all $u \in X$,* $\Pr[u \in S] = \alpha \|f(u)\|^2$.
2. *For all $u, v \in X$ with $\|f(u) - f(v)\|^2 \ge \beta \min\{\|f(u)\|^2, \|f(v)\|^2\}$,*

$$\Pr[u, v \in S] \le \frac{\alpha \min\{\|f(u)\|^2, \|f(v)\|^2\}}{s}.$$

3. *For all* $e \subseteq X$,

$$\Pr[e \text{ is cut by } S] \leq \alpha D \max_{u,v \in e} \|f(u) - f(v)\|^2 .$$

Compared to the definition of orthogonal separators in Definition 3.6.9, there is a slight change in the second condition, and the third condition bounds the probability of cutting a hyperedge instead of an edge, with the upper bound coming from the objective of the SDP relaxation for hypergraph small-set expansion, which we shall give shortly. The existence of hypergraph orthogonal separators with bounded distortion is established in [LM14b].

Theorem 3.6.16 (Existence of Hypergraph Orthogonal Separators [LM14b])**.** *There exists a randomized polynomial-time algorithm that, given* $f : X \to \mathbb{R}^d$ *satisfying the* ℓ_2^2 *triangle inequalities, and parameters* s *and* $\beta < 1$, *generates a hypergraph* s*-orthogonal separator with distortion* $D = O(\beta^{-1} \cdot s \log s \log\log s \cdot \sqrt{\log |X|})$ *and scale* $\alpha \geq \max(1/|X|, 1/s)$.

Now, we define the problem of hypergraph small-set expansion and the SDP relaxation proposed in [LM14b].

Definition 3.6.17 (Hypergraph Small-Set Expansion [LM14b])**.** *Let* $H = (V, E)$ *be a hypergraph, and* $\eta \in (0, 1/2]$ *be a parameter. The* η*-small set expansion (or* η*-SSE) of* $S \subseteq V$ *and of* H *is defined as*

$$\phi_\eta(G) := \min_{0 < \mathrm{vol}(S) \leq \eta \cdot \mathrm{vol}(V)} \phi(S),$$

where $\phi(S) := |\delta(S)|/\mathrm{vol}(S)$ *is the usual conductance*[22] *of* S *and* $\delta(S) := \{e \in E : e \cap S \neq \emptyset \text{ and } e \cap S^c \neq \emptyset\}$ *is the set of hyperedges cut by* S.

Definition 3.6.18 (SDP Relaxation of Hypergraph Small-Set Expansion [LM14b])**.** *Let* $H = (V, E)$ *be a hypergraph. The SDP relaxation of* $\phi_\eta(H)$ *with* ℓ_2^2 *triangle inequalities, denoted* $\mathsf{sdp}_\eta^\Delta(H)$, *is defined as follows:*

$$\begin{aligned}
\min_{f: V \to \mathbb{R}^n} \quad & \sum_{e \in E} \max_{u,v \in e} \|f(u) - f(v)\|^2 \\
\text{subject to} \quad & \sum_{u \in V} \deg(u) \|f(u)\|^2 = 1 \\
& \sum_{v \in V} \deg(v) \langle f(u), f(v) \rangle \leq \eta n \cdot \|f(u)\|^2 \qquad \forall u \in V \\
& \|f(u) - f(v)\|^2 + \|f(v) - f(u')\|^2 \geq \|f(u) - f(u')\|^2 \qquad u, v, u' \in V \\
& 0 \leq \langle f(u), f(v) \rangle \leq \|f(u)\|^2 \qquad \forall u, v \in V.
\end{aligned}$$

[22] Here we define hypergraph SSE using degree-weighted vertex measure for consistency with the rest of the subsection. The definition can be extended to arbitrary vertex measures and the same approximation result applies by tweaking the definition of the SDP relaxation.

Theorem 3.6.19 (Approximation of Hypergraph SSE [LM14b, Theorem 1.3])**.** *There is a randomized polynomial-time approximation algorithm for the hypergraph small-set expansion problem that, given a hypergraph* $H = (V, E)$ *and parameters* $\varepsilon \in (0, 1)$ *and* $\eta \in (0, 1/2]$*, finds a set* $S \subseteq V$ *of size at most* $(1 + \varepsilon)\eta \cdot n$ *such that* $\phi(S)$ *is at most*

$$O_\varepsilon\left(\eta^{-1}\log(\eta^{-1})\log\log(\eta^{-1}) \cdot \sqrt{\log n}\right)$$

times the optimal. In particular, when η *and* ε *are fixed, the algorithm gives an* $O(\sqrt{\log n})$*-approximation.*

Remark 3.6.20 (Implication for Small-Set Vertex Expansion)**.** *It is possible to obtain a result analogous to Theorem 3.6.19 for small-set vertex expansion, via a reduction to small-set hypergraph expansion. The reduction is by combining Proposition 2.3.6 and Proposition 2.3.7 and running the algorithm for hypergraph SSE on the constructed hypergraph, which can be transformed into a small cut in the original graph with small vertex expansion. Refer to [LM14b] for more details.*

Summary

To summarize, orthogonal separators is a fitting tool for rounding vector programs whose solution vectors satisfy the ℓ_2^2 triangle inequalities as well as some kind of "spreading" constraints. Such constraints are present, for example, in multi-way partitioning problems and small-set expansion problems, and the rounding algorithms employing orthogonal separators attain good approximation ratios, matching the "gold standard" of $O(\sqrt{\log n})$ in the special cases of graph/hypergraph edge expansions.

Chapter 4

Cheeger Inequalities for Vertex Expansion using Reweighted Eigenvalues

In this chapter, we derive an "optimal" Cheeger's inequality for vertex expansion, connecting the vertex expansion $\psi(G)$ of a graph $G = (V, E)$ and the maximum reweighted second smallest eigenvalue $\lambda_2^*(G)$ of the Laplacian matrix:

$$\frac{\psi^2(G)}{\log \Delta} \lesssim \lambda_2^*(G) \lesssim \psi(G).$$

This improves on an earlier result by Olesker-Taylor and Zanetti [OZ22] and also generalizes to the weighted vertex expansion, answering an open question by them.

Building on this connection, we then develop a new spectral theory for vertex expansion. We discover that several interesting generalizations of Cheeger inequalities relating edge conductances and eigenvalues have a close analog in relating vertex expansions and reweighted eigenvalues. These include:

- An analog of Trevisan's result that relates the bipartite vertex expansion $\psi_B(G)$ of a graph and the maximum reweighted lower spectral gap $\zeta^*(G)$ of the adjacency matrix. This implies the first approximation algorithm for bipartite vertex expansion.
- An analog of higher-order Cheeger's inequalities that relates the k-way vertex expansion $\psi_k(G)$ of a graph and the maximum reweighted k-th smallest eigenvalue $\lambda_k^*(G)$

of the Laplacian matrix. This implies the first approximation algorithm for k-way vertex expansion.

- An analog of improved Cheeger's inequality that relates the vertex expansion $\psi(G)$ and the reweighted eigenvalues $\lambda_2^*(G)$ and $\lambda_k^*(G)$. This provides an improved bound for $\psi(G)$ using $\lambda_2^*(G)$, when the k-way vertex expansion $\psi_k(G)$ is large for a small k.

Finally, inspired by this connection, we present negative evidence to the 0/1-polytope edge expansion conjecture by Mihail and Vazirani. We construct 0/1-polytopes whose graphs have very poor vertex expansion. This implies that the fastest mixing time to the uniform distribution on the vertices of these 0/1-polytopes is almost linear in the graph size. This does not provide a counterexample to the conjecture, but this is in contrast with known positive results which proved poly-logarithmic mixing time to the uniform distribution on the vertices of subclasses of 0/1-polytopes.

4.1 Our Results

We now present our results formally. The reader is advised to refer to Section 2.2 and Section 2.3 for various notations and definitions.

4.1.1 Optimal Cheeger Inequality for Vertex Expansion

Our first main result is an "optimal" Cheeger inequality that relates the maximum reweighted second smallest eigenvalue $\lambda_2^*(G)$ and vertex expansion $\psi(G)$. This is the cumulation of a line of work [BDX04, Roc05, OZ22] that studies the fastest mixing time problem and its relation with vertex expansion, as we have thoroughly reviewed in Section 3.2.2. We recall the definition of $\lambda_2^*(G)$ and of weighted vertex expansion here for convenience.

Definition 4.1.1 (Maximum Reweighted Spectral Gap [BDX04] (restatement of Definition 1.1.1)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted spectral gap is defined as*

$$\begin{aligned} \lambda_2^*(G) \;:=\; & \max_{P \geq 0} \quad 1 - \alpha_2(P) & \\ \text{subject to} \quad & P(u,v) = P(v,u) = 0 & \forall uv \notin E \\ & \sum_{v \in V} P(u,v) = 1 & \forall u \in V \\ & \pi(u)P(u,v) = \pi(v)P(v,u) & \forall uv \in E. \end{aligned}$$

The graph is assumed to have a self-loop on each vertex, to ensure that the optimization problem for $\lambda_2^*(G)$ is always feasible. In the context of Markov chains, this corresponds to allowing a nonzero holding probability on each vertex.

Definition 4.1.2 (Weighted Vertex Expansion)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the vertex expansion of G is defined as*

$$\psi(G) := \min\Big\{1, \min_{\emptyset \neq S \subseteq V} \{\psi(S) \mid 0 < \pi(S) \leq \frac{1}{2}\}\Big\},$$

where $\psi(S) := \pi(\partial S)/\pi(S)$ and ∂S is the vertex boundary of S as defined in Section 2.3.1.

Olesker-Taylor and Zanetti proved the Cheeger-type inequality in Theorem 3.2.4 that

$$\frac{\psi(G)^2}{\log n} \lesssim \lambda_2^*(G) \lesssim \psi(G).$$

when π is the uniform distribution. They posed the problem of reducing the $\log n$ factor in Theorem 3.2.4 to $\log \Delta$, and also the problem of generalizing their result to weighted vertex expansion. Our first result provides a positive resolution of these two problems.

Theorem 4.1.3 (Cheeger Inequality for Weighted Vertex Expansion)**.** *For any undirected graph $G = (V, E)$ with maximum degree Δ and any probability distribution π on V,*

$$\frac{\psi(G)^2}{\log \Delta} \lesssim \lambda_2^*(G) \lesssim \psi(G). \tag{4.1}$$

In terms of the fastest mixing time $\tau^*_{\text{mix}}(G)$ to the stationary distribution,

$$\frac{1}{\psi(G)} \lesssim \tau^*_{\text{mix}}(G) \lesssim \frac{\log \Delta \cdot \log \pi_{\min}^{-1}}{\psi^2(G)}.$$

In Section 4.8, we show that the $\log \Delta$ factor in Theorem 4.1.3 is optimal, by exhibiting graphs G with $\lambda_2^*(G) \asymp \frac{\psi(G)^2}{\log \Delta}$. Note that the tightness result does not rely on the small-set expansion hypothesis (c.f. Section 3.1.5).

The improvement to $\log \Delta$ shows that $\lambda_2^*(G)$ is a strictly tighter parameter to relate to vertex expansion than $\lambda_2'(G)$. The latter was studied in the early works of Tanner [Tan84], Alon and Milman [AM85], and Alon [Alo86], where the Cheeger-type inequality relating $\lambda_2'(G)$ and $\psi(G)$ has a factor of Δ between the upper and lower bounds (see Section 3.2.1).

We note that Louis, Raghavendra and Vempala [LRV13] gave an SDP approximation algorithm for vertex expansion with the same approximation guarantee (see Section 3.2.4), but their SDP is different from and stronger than that in Definition 1.1.1 (as proven in Lemma 4.3.2), and so it does not have the natural interpretation as the reweighted second eigenvalue and does not imply the result on fastest mixing time.

The current best known approximation algorithm for vertex expansion $\psi(G)$ is an $O(\sqrt{\log n})$ SDP-based approximation algorithm by Feige, Hajiaghayi and Lee [FHL08]. This is an extension of the $O(\sqrt{\log n})$ SDP-based approximation algorithm for edge conductance $\phi(G)$ by Arora, Rao, and Vazirani [ARV09]. The SDP formulation of [ARV09] is known to be strictly more powerful than the spectral formulation by the second eigenvalue. Refer to Section 3.6 for details.

Even though $\lambda_2^*(G)$, sdp_∞ in [LRV13], and the SDP in [FHL08] are all semidefinite programming relaxations for $\psi(G)$ and satisfy similar inequalities, we note that the approach of using reweighted eigenvalues has some additional features. One important feature is that $\lambda_2^*(G)$ is closely related to fastest mixing time. This allows one to develop a spectral theory for vertex expansion that relates (i) vertex expansion, (ii) reweighted eigenvalues and (iii) fastest mixing time, which parallels the classical spectral graph theory that relates (i) edge conductance, (ii) eigenvalues and (iii) mixing time.

More recently, Louis [Lou15] and Chan, Louis, Tang, Zhang [CLTZ18] developed a spectral theory for hypergraphs (see Section 3.4.1). Through a reduction from vertex expansion to hypergraph edge expansion (see Fact 3.2.19), they developed a spectral theory for vertex expansion as well. This theory also relates (i) expansion, (ii) eigenvalues and (iii) mixing time, and so their work is closest to the current work. Compared to their approach, we note that the current approach using reweighted eigenvalues is more direct and effective for vertex expansion. The reduction in Fact 3.2.19 from vertex expansion $\psi(S)$ for a graph G with maximum degree $\Delta_{\max}$ and minimum degree $\Delta_{\min}$ to hypergraph edge conductance $\phi_H(S)$ only satisfies

$$\Delta_{\min} \cdot \phi_H(S) \leq \psi(S) \lesssim \Delta_{\max} \cdot \phi_H(S),$$

and so the approximation ratio depends on the ratio between the maximum degree and the minimum degree. In contrast, the current approach using reweighted eigenvalues does not have this dependency. Also, the definitions of the hypergraph diffusion process and its eigenvalues are quite technically involved and require considerable effort to make rigorous [CTWZ19]. We believe that the definitions of reweighted eigenvalues are more intuitive and more closely related to ordinary eigenvalues.

Remark 4.1.4 (Reweighting Conjectures in Approximation Algorithms). *Besides the fastest mixing time problem, we note that these "reweighting problems" relating vertex expansion and reweighted eigenvalues are also well motivated in the study of approximation algorithms. One example is a conjecture of Arora and Ge [AG11, Conjecture 12], which roughly states that, if a graph G has almost perfect vertex expansion for every set, then there exists a reweighted doubly stochastic matrix P of the adjacency matrix of G so that P has few eigenvalues less than $-\frac{1}{16}$. They proved that if the conjecture was true, then there is an improved subexponential time algorithm for coloring* 3*-colorable graphs. Another example is a conjecture of Steurer [Ste10, Conjecture 9.2], which is also known to be related to a reweighting problem between vertex expansion and the graph spectrum, that if true would imply an improved subexponential time approximation algorithm for the sparsest cut problem.*

The proof of Theorem 4.1.3 is based on the techniques in [LRV13, BHT00], which we have discussed in detail in Section 3.2.4.

4.1.2 Maximum Reweighted Lower Spectral Gap and Bipartite Vertex Expansion

Trevisan [Tre09] proved that the lower spectral gap $1 + \alpha_n(G)$ of the normalized adjacency matrix of $G = (V, E)$ is small if and only if there is a subset $S \subseteq V$ which is an almost bipartite component in G with small edge conductance $\phi(S)$. See Section 3.1.2 for a review of their work.

We define the analogous notions for vertex expansion and for reweighted lower spectral gap.

Definition 4.1.5 (Bipartite Vertex Expansion). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the bipartite vertex expansion of G is defined as*

$$\psi_B(G) := \min \Big\{ 1, \min_{\emptyset \neq S \subseteq V} \{ \psi(S) \mid G[S] \text{ is an induced bipartite graph} \} \Big\}.$$

Definition 4.1.6 (Maximum Reweighted Lower Spectral Gap). *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted lower spectral* *gap is defined as*

$$\begin{aligned}
\zeta^*(G) \;:=\; \max_{P\geq 0} \quad & \lambda_{\min}(D_P + P) \\
\text{subject to} \quad & P(u,v) = P(v,u) = 0 && \forall uv \notin E \\
& \sum_{v\in V} P(u,v) \leq 1 && \forall u \in V \\
& \pi(u)P(u,v) = \pi(v)P(v,u) && \forall uv \in E.
\end{aligned}$$

where D_P is the diagonal matrix of row sums of P such that $D_P(u,u) = \sum_{v\in V} P(u,v)$ for $u \in V$. We note that this program is slightly different from that in Definition 1.1.1, and the main reason is that self-loops should not be allowed in this problem. We will explain more about this in Section 4.4.

We prove an analog of Trevisan's result that the maximum reweighted lower spectral gap is small if and only if there is an induced bipartite subgraph on S with small vertex expansion $\psi(S)$.

Theorem 4.1.7 (Cheeger Inequality for Bipartite Vertex Expansion)**.** *For any undirected graph $G = (V,E)$ with maximum degree Δ and any probability distribution π on V,*

$$\frac{\psi_B(G)^2}{\log \Delta} \lesssim \zeta^*(G) \lesssim \psi_B(G).$$

This is the first approximation algorithm for bipartite vertex expansion to our knowledge. Finding a two-colorable set with small vertex expansion is one of the three ways in Blum's coloring tools [Blu94] to make progress in designing approximation algorithms for coloring 3-colorable graphs. Indeed, it is in this context that Arora and Ge [AG11] made the reweighting conjecture mentioned in the introduction. Theorem 4.1.7 does not imply anything new about approximating graph coloring, but we hope that it is a step towards answering Arora and Ge's conjecture.

4.1.3 Higher-Order Cheeger Inequality for Vertex Expansion

Lee, Oveis Gharan and Trevisan [LOT12] and Louis, Raghavendra, Tetali and Vempala [LRTV12] proved the higher-order Cheeger inequality, which state that the k-th smallest eigenvalue $\lambda_k(G)$ of the normalized Laplacian matrix of $G = (V,E)$ is small if and only if the k-way edge conductance $\phi_k(G)$ is small. More precisely, they proved that $\lambda_k \lesssim \phi_k \lesssim k^2\sqrt{\lambda_k}$ and $\phi_{\frac{k}{2}} \lesssim \sqrt{\lambda_k \log k}$. See Section 3.1.3 for a review of their work.

We consider the analogous notion of k-way vertex expansion.

Definition 4.1.8 (k-Way Vertex Expansion)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the k-way vertex expansion of G is defined as*

$$\psi_k(G) := \min\left\{1, \min_{S_1,\ldots,S_k \subseteq V} \max_{1\le i\le k} \psi(S_i)\right\},$$

where the minimum is taken over pairwise disjoint subsets $S_1, \ldots, S_k$ of V.

Definition 4.1.9 (Maximum Reweighted k-th Smallest Eigenvalue)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted k-th smallest eigenvalue of the normalized Laplacian matrix of G is defined as $\lambda_k^*(G) := \max_{P\ge 0} \lambda_k(I - P)$, where P is subject to the same constraints stated in Definition 1.1.1.*

We prove an analog of higher-order Cheeger inequalities that the maximum reweighted k-th smallest eigenvalue is small if and only if the k-way vertex expansion is small. As in previous work [LOT12, LRTV12], there is a better approximation guarantee if we consider only $\frac{k}{2}$-way vertex expansion.

Theorem 4.1.10 (Higher-Order Cheeger Inequality for Vertex Expansion)**.** *For any undirected graph $G = (V, E)$ with maximum degree Δ and any probability distribution π on V,*

$$\lambda_k^*(G) \lesssim \psi_k(G) \lesssim k^{\frac{9}{2}} \log k \sqrt{\log\Delta \cdot \lambda_k^*(G)} \quad \text{and} \quad \psi_{\frac{k}{2}}(G) \lesssim \sqrt{k} \log k \sqrt{\log\Delta \cdot \lambda_k^*(G)}.$$

Chan, Louis, Tang and Zhang [CLTZ18] developed a spectral theory for hypergraphs and proved a higher-order Cheeger inequality for hypergraph (edge) expansion. Through a reduction from vertex expansion to hypergraph expansion (see Fact 3.2.19), they proved that $\psi_{\frac{k}{2}}(G) \lesssim k^{\frac{5}{2}} \log k \log\log k \cdot \log\Delta \cdot \sqrt{\xi_k}$ for graphs with constant ratio between the maximum degree and the minimum degree, where $\xi_k \lesssim \psi_k(G)$ is a relaxation for k-way vertex expansion. Compared to their result, Theorem 4.1.10 does not require the assumption about the maximum degree and the minimum degree of G, and has a better approximation ratio for $\frac{k}{2}$-way vertex expansion. Furthermore, Theorem 4.1.10 provides the first true approximation algorithm for k-way vertex expansion $\psi_k(G)$ to our knowledge.

4.1.4 Improved Cheeger Inequality for Vertex Expansion

Kwok, Lau, Lee, Oveis Gharan, and Trevisan [KLL+13] proved an improved Cheeger inequality that $\phi(G) \lesssim k\lambda_2(G)/\sqrt{\lambda_k(G)}$ for any $k \ge 2$. This shows that $\lambda_2(G)$ is a tighter approximation to $\phi(G)$ when $\lambda_k(G)$ is large for a small k. The result provides an explanation for the good empirical performance of the spectral partitioning algorithm. See Section 3.1.4 for a review of their work.

We prove an analogous result that if the $\lambda_k^*(G)$ is large for a small k, then $\lambda_2^*(G)$ is a tighter approximation to the vertex expansion $\psi(G)$. The following result is close to the tight result in [KLL+13] for edge conductance as we will elaborate in Remark 4.6.4.

Theorem 4.1.11 (Improved Cheeger Inequality for Vertex Expansion)**.** *For any undirected graph $G = (V, E)$ with maximum degree Δ, and for any probability distribution π on V and any k such that $2 \le k \le \frac{n}{2}$,*

$$\lambda_2^*(G) \lesssim \psi(G) \lesssim \frac{k \cdot \lambda_2^*(G) \cdot \log \Delta}{\sqrt{\lambda_k^*(G)}}.$$ [1]

We remark that the reweightings used in $\lambda_2^*(G)$ and $\lambda_k^*(G)$ may be different. Through Theorem 4.1.10, we obtain the following corollary that only depends on the graph structure: If the k-way vertex expansion $\psi_k(G)$ is large for a small k, then $\lambda_2^*(G)$ is a tighter approximation to $\psi(G)$.

4.1.5 Vertex Expansion of 0/1-Polytopes

Mihail and Vazirani (see [FM92]) conjectured that the graph $G = (V, E)$ (i.e. 1-skeleton) of any 0/1-polytope is an edge expander, such that $|\delta(S)|/|S| \ge 1$ for every subset $S \subseteq V$ with $|S| \le |V|/2$, where $\delta(S)$ denotes the set of edges between S and $V \backslash S$. This conjecture would imply fast mixing time of random walks to the stationary distribution, with applications in designing fast sampling algorithms for many classes of combinatorial objects. The conjecture is proved to be correct in several cases [FM92, Kai04, ALOV19], most notably the recent resolution of the matroid expansion conjecture [ALOV19] by Anari, Liu, Oveis Gharan and Vinzant.

In all these positive results, the Markov chain can be set up so that the stationary distribution is the uniform distribution, with the mixing time to the stationary distribution poly-logarithmic in the graph size. Then the fast sampling algorithms can also be used to obtain an approximate counting algorithm on the number of vertices in the given 0/1-polytope, with poly-logarithmic runtime in the graph size. Therefore, sampling from the uniform distribution is usually the setting of interest.

[1] This improves the original result in [KLT22], which had $k^{\frac{3}{2}}$ in the upper bound instead of k.

Inspired by the connection between fastest mixing time and vertex expansion, we consider a variant of Mihail and Vazirani's conjecture: Is the graph of every 0/1-polytope a vertex expander? Perhaps surprisingly, we show that there are 0/1-polytopes whose graphs are very poor vertex expanders.

Theorem 4.1.12 (0/1-Polytopes with Poor Vertex Expansion)**.** *Let π be the uniform distribution. For any $k > 2$ and any $n > 2k$ sufficiently large, there is a 0/1-polytope $Q = Q_{n,k} \subseteq \{0,1\}^n$ with $O(n^k)$ vertices and*

$$\psi(Q) \lesssim \frac{(4k)^k}{n^{k-2}}.$$

Theorem 4.1.12 and Theorem 3.2.4 together imply that even the fastest mixing time of the reversible random walks on some 0/1-polytopes is almost linear in the graph size.

Corollary 4.1.13 (Torpid Mixing to Uniform Distribution)**.** *For any constant $k > 2$, there exists a 0/1-polytope Q such that any reversible Markov chain on its graph $G_Q = (V, E)$ with stationary distribution $\vec{1}/|V|$ has mixing time $\Omega\big(|V|^{1-\frac{2}{k}}\big)$.*

While Theorem 4.1.12 does not provide a counterexample to the conjecture of Mihail and Vazirani, it shows that even if the conjecture is true, there are 0/1-polytopes for which random walks cannot be used for efficient uniform sampling and for efficient approximate counting.

Remark 4.1.14. *After posting the first version of the paper [KLT22] on arXiv, we found out that Gillmann [Gil07, Chapter 3.2] has already constructed examples of 0/1-polytopes whose graphs have poor vertex expansion. The polytopes $Q \subseteq \{0,1\}^n$ constructed have $2^{(h(c)+o(1))n}$ vertices and satisfy*

$$\psi(Q) \lesssim 2^{-(h(c)-2c)n},$$

where $h(x) := -x \log x - (1-x)\log(1-x)$ is the binary entropy function and $c := 1/5$ (correspondingly $h(c) = 0.7219...$). Applying Theorem 3.2.4, this would imply a fastest mixing time bound of $\Omega(|V|^{0.4459...})$. By choosing smaller values of c, an almost linear fastest mixing time bound can be obtained as in Corollary 4.1.13.

4.2 Our Techniques

From a technical perspective, the advantage of relating reweighted eigenvalues to vertex expansions is that many ideas relating eigenvalues to edge conductances can be carried over to the new setting. So, many steps in our proofs are natural extensions of previous arguments, and we focus our discussion here on the new elements.

Vertex Expansion: The proof of Theorem 3.2.4 by Olesker-Taylor and Zanetti is based on the dual characterization of reweighted second eigenvalue in Proposition 3.2.3, due to Roch [Roc05], and it has two main steps. In the first step, they used the Johnson-Lindenstrass lemma to project the SDP solution into a $O(\log n)$-dimensional solution, and then further reduce it to a 1-dimensional "spectral" solution by taking the best coordinate. This is the step where the $\log n$ factor is lost. In the second step, they introduced an interesting new concept called the "matching conductance", and used some combinatorial arguments about greedy matchings for the analysis of Cheeger rounding on Roch's dual program.

In our proof of Theorem 4.1.3, we also use Roch's dual characterization and follow the same two steps. In the first step, we use the Gaussian projection method in [LRV13] (see Section 3.2.4) to reduce the SDP solution to a 1-dimensional solution directly, and adapt their analysis to show that only a factor of $\log \Delta$ is lost. In the second step, we bypass the concept of matching conductance and do a more traditional analysis of Cheeger rounding as in Bobkov, Houdré and Tetali [BHT00] (see Section 3.2.4). It turns out that this analysis works smoothly for weighted vertex conductance, while the approach using matching conductance faced some difficulties as described in [OZ22] (c.f. Section 3.2.2). A new element in our proof is the introduction of an intermediate dual program using graph orientation, which is important in the analysis of both steps.

Bipartite Vertex Expansion: The proof of Theorem 4.1.7 for bipartite vertex expansion follows closely the proof of Theorem 4.1.3 and Trevisan's result [Tre09] (see Section 3.1.2), once the correct formulation in Definition 4.1.6 is found.

Multiway Vertex Expansion: For the proof of higher-order Cheeger inequality for vertex expansion in Theorem 4.1.10, one technical issue is that we do not know of a convex relaxation for the maximum reweighted k-th smallest eigenvalue in Definition 4.1.9. Instead, we define a related quantity $\sigma_k^*(G)$ called the maximum reweighted sum of the k smallest eigenvalues in Definition 4.5.1, which can be written as a semidefinite program. We show in Proposition 4.5.2 that this quantity has a nice dual characterization that satisfies the sub-isotropy condition. This allows us to adapt the techniques in [LOT12] to decompose the SDP solution into k disjointly supported SDP solutions with small objective values, so that we can apply Theorem 4.1.3 to find k disjoint sets with small vertex expansion. A review of the background from [LOT12] needed for the proof is available in Section 3.1.3.

Improved Cheeger Inequality: The proof of improved Cheeger inquality for vertex expansion is similar to that in [KLL+13], which has two main steps. The first step is to prove that if the 1-dimensional solution to Roch's dual program is close to a k-step function, then Cheeger rounding performs well. The second step is to prove that if the 1-dimensional solution to Roch's dual program is far from a k-step function, then we can construct an SDP solution to σ_k^* with small objective value, which proves that λ_k^* is small. Therefore, if λ_k^* is large, then the 1-dimensional solution must be close to a k-step function, and hence Cheeger rounding performs well. One interesting aspect in this proof is to relate the performance of a rounding algorithm of one SDP (in this case $\lambda_2^*(G)$) to the objective value of another SDP (in this case $\sigma_k^*(G)$). A review of the background from [KLL+13] is available in Section 3.1.4.

Vertex Expansion of 0/1-Polytopes: The examples in Theorem 4.1.12 for 0/1-polytope is obtained by a simple probabilistic construction. The graph of a 0/1-polytope is defined by the set of points chosen in $\{0,1\}^n$. Let L be the set of points with k ones, and let R be the set of points with $(n-k)$ ones. We prove that if we choose a random set M of points with $n/2$ ones and set $|M| \asymp 4^k n^2$, then with high probability there are no edges between L and R in the resulting polytope, and so M is a small vertex separator of L and R where each has $\binom{n}{k}$ points. The proof is by elementary geometric arguments about the edges of a polytope, and a simple result bounding the number of linear threshold functions in the boolean hypercube $\{0,1\}^n$.

Remark 4.2.1 (Concurrent Work)**.** *Jain, Pham, and Vuong [JPV22] independently published a proof of Theorem 4.1.3 for the uniform distribution case. Their approach is based on a better analysis of dimension reduction for maximum matching, which is quite different from our approach as we bypassed the concept of matching conductance in [OZ22]; see Section 3.2.3 for a review. In Chapter 5, we will adopt their approach in analyzing dimension reduction.*

4.3 Optimal Cheeger Inequality for Vertex Expansion

The goal of this section is to prove Theorem 4.1.3. Our starting point is the dual program by [Roc05], which we copy here for convenience.

Proposition 4.3.1 (Dual Programs for Fastest Mixing [Roc05] (restatement of Proposition 3.2.3 and Definition 3.2.5))**.** *Given an undirected graph $G=(V,E)$ and a probability distribution π on V, the following semidefinite program is dual to the primal program in*

Definition 1.1.1 with strong duality $\lambda_2^*(G) = \gamma(G)$ where

$$\begin{aligned} \gamma(G) := \min_{f:V\to\mathbb{R}^n,\ g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} \pi(v) g(v) \\ \text{subject to} \quad & \sum_{v\in V} \pi(v) \|f(v)\|^2 = 1 \\ & \sum_{v\in V} \pi(v) f(v) = \vec{0} \\ & g(u) + g(v) \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E. \end{aligned}$$

$\gamma^{(1)}(G)$ as defined in Definition 3.2.5 is the 1-dimensional program of $\gamma(G)$ with $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

The proof is by combining the proofs in [OZ22, LRV13, BHT00] with a graph orientation idea. We follow the same two-step plan as in [OZ22]. We will prove in Proposition 4.3.6 in Section 4.3.2 that $\gamma^{(1)}(G) \lesssim \gamma(G) \cdot \log \Delta$ for any probability distribution π. Note that this already improves Theorem 3.2.4 to the optimal bound, when π is the uniform distribution. Then, we will prove in Theorem 4.3.7 in Section 4.3.3 that $\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi(G)$ for any probability distribution π on V, which generalizes Theorem 3.2.7. As in [OZ22], combining Proposition 4.3.1 and Proposition 4.3.6 and Theorem 4.3.7 gives Theorem 4.1.3. We recommend reading the relevant subsections in Section 3.2 before continuing with this section.

4.3.1 Dual Program on Graph Orientation

To extend the techniques in [LRV13, BHT00] to prove the two steps, we will introduce a "directed" program $\vec{\gamma}(G)$ to bring $\gamma(G)$ closer to $\mathsf{sdp}_\infty(G)$ in Definition 3.2.16.

Observe that the two programs $\gamma(G)$ and $\mathsf{sdp}_\infty(G)$ have very similar form. The only difference is that the last constraint in $\gamma(G)$ only requires that $g(u)+g(v) \geq \|f(u) - f(v)\|^2$ for $uv \in E$, while the last constraint in Definition 3.2.16 has a stronger requirement that $\min\{g(u), g(v)\} \geq \|f(u) - f(v)\|^2$ for $uv \in E$. So $\mathsf{sdp}_\infty(G)$ is a stronger relaxation than $\gamma(G) = \lambda_2^*(G)$.

Lemma 4.3.2. *For any undirected graph $G = (V, E)$ and any distribution π on V,*

$$\lambda_2^*(G) \leq \mathsf{sdp}_\infty(G).$$

For our analysis of $\lambda_2^*(G)$, we consider the following "directed" program $\vec{\gamma}(G)$ where the last constraint is $\max\{g(u), g(v)\} \geq \|f(u) - f(v)\|^2$ for $uv \in E$. We also state the corresponding 1-dimensional version as in Proposition 4.3.1 in the following definition.

Definition 4.3.3 (Directed Dual Programs for $\gamma(G)$)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V,*

$$\begin{aligned}
\vec{\gamma}(G) := \min_{f:V\to\mathbb{R}^n,\ g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} \pi(v) g(v) \\
\text{subject to} \quad & \sum_{v\in V} \pi(v) \|f(v)\|^2 = 1 \\
& \sum_{v\in V} \pi(v) f(v) = \vec{0} \\
& \max\{g(u), g(v)\} \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E.
\end{aligned}$$

$\vec{\gamma}^{(1)}(G)$ is defined as the 1-dimensional program of $\vec{\gamma}(G)$ where $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

Note that $\vec{\gamma}(G)$ is not a semidefinite program because of the max constraint, but $\gamma(G)$ and $\vec{\gamma}(G)$ are closely related and $\vec{\gamma}(G)$ is only used in the analysis as a proxy for $\gamma(G)$.

Lemma 4.3.4. *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\gamma(G) \leq \vec{\gamma}(G) \leq 2\gamma(G) \quad \text{and} \quad \gamma^{(1)}(G) \leq \vec{\gamma}^{(1)}(G) \leq 2\gamma^{(1)}(G).$$

Proof. As $g \geq 0$, any feasible solution f, g to $\vec{\gamma}(G)$ is a feasible solution to $\gamma(G)$ and so the first inequalities follow. On the other hand, for any feasible solution f, g to $\gamma(G)$, note that $f, 2g$ is a feasible solution to $\vec{\gamma}(G)$ and so the second inequalities follow. □

The reason that we call $\vec{\gamma}(G)$ the "directed" program is as follows. For each edge $uv \in E$, the constraint in $\mathsf{sdp}_\infty(G)$ requires both $g(u)$ and $g(v)$ to be at least $\|f(u) - f(v)\|^2$, while the constraint in $\vec{\gamma}(G)$ only requires at least one of $g(u)$ or $g(v)$ to be at least $\|f(u) - f(v)\|^2$. We think of $\vec{\gamma}(G)$ as assigning a direction to each edge and requiring that $g(v) \geq \|f(u) - f(v)\|^2$ for each directed edge $u \to v$. Then, we can rewrite the programs $\vec{\gamma}(G)$ and $\vec{\gamma}^{(1)}(G)$ by eliminating the variables $g(v)$ for $v \in V$, by minimizing over all possible orientations of the edge set E.

Lemma 4.3.5 (Directed Dual Programs Using Orientation for $\gamma(G)$)**.** *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of the undirected edges in E. Then*

$$\begin{aligned}\vec{\gamma}(G) = \min_{f:V\to\mathbb{R}^n} \min_{\overrightarrow{E}} \quad & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} \|f(u) - f(v)\|^2 \\ \text{subject to} \quad & \sum_{v\in V} \pi(v) \|f(v)\|^2 = 1 \\ & \sum_{v\in V} \pi(v) f(v) = \vec{0}.\end{aligned}$$

Similarly, $\vec{\gamma}^{(1)}(G)$ can be written in the same form with $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

Proof. In one direction, given an orientation $\overrightarrow{E}$, define $g(v) := \max_{u:uv\in\overrightarrow{E}} \|f(u) - f(v)\|^2$, so that f, g is a feasible solution to $\vec{\gamma}(G)$ as stated in Definition 4.3.3 with the same objective value.

In the other direction, given a solution f, g in Definition 4.3.3, we can define an orientation $\overrightarrow{E}$ of E so that each directed edge uv satisfies $g(v) \geq \|f(u) - f(v)\|^2$. Note that $g(v) \geq \max_{u:uv\in\overrightarrow{E}} \|f(u) - f(v)\|^2$, and setting it to be an equality would satisfy all the constraints and not increase the objective value as $g \geq 0$. □

This formulation will be useful in both the Gaussian projection step for Proposition 4.3.6 and the threshold rounding step for Theorem 4.3.7.

4.3.2 Gaussian Projection

The following proposition is an improvement of Proposition 3.2.6 in [OZ22]. The formulation in Lemma 4.3.5 allows us to use the expected maximum of Gaussian random variables in Fact 2.10.6 to analyze the projection as was done in [LRV13].

Proposition 4.3.6 (Gaussian Projection for $\gamma(G)$)**.** *For any undirected graph $G = (V, E)$ with maximum degree Δ and any probability distribution π on V,*

$$\gamma(G) \leq \gamma^{(1)}(G) \lesssim \gamma(G) \cdot \log \Delta.$$

Proof. We will prove that $\vec{\gamma}(G) \leq \vec{\gamma}^{(1)}(G) \lesssim \vec{\gamma}(G) \cdot \log \Delta$, and the proposition will follow from Lemma 4.3.4. The first inequality is immediate as $\vec{\gamma}^{(1)}(G)$ is a restriction of $\vec{\gamma}(G)$, so we focus on proving the second inequality.

The proof is almost identical to the proof of Theorem 3.2.17 that

$$\mathsf{sdp}_\infty(G) \le \lambda_\infty(G) \lesssim \mathsf{sdp}_\infty(G) \cdot \log \Delta.$$

Let $f : V \to \mathbb{R}^n$ and $\overrightarrow{E}$ be a solution to $\vec{\gamma}(G)$ as stated in Lemma 4.3.5. As in [LRV13], we construct a 1-dimensional solution $y \in \mathbb{R}^n$ to $\vec{\gamma}^{(1)}(G)$ by setting $y(v) = \langle f(v), z \rangle$, where $h \sim N(0,1)^n$ is a Gaussian random vector with independent entries.

First, consider the expected objective value of y to $\vec{\gamma}^{(1)}(G)$. For each max term in the summand,

$$\mathbb{E}\left[\max_{u:u\to v} \left(y(u) - y(v)\right)^2\right] = \mathbb{E}\left[\max_{u:u\to v} \left\langle f(u) - f(v), z \right\rangle^2\right] \le 2 \max_{u:u\to v} \|f(u) - f(v)\|^2 \cdot \log \Delta,$$

where the last inequality is by applying Fact 2.10.6 on the centered Gaussian random variable $\langle f(u) - f(v), z\rangle$ with variance $\|f(u) - f(v)\|^2$ for each of the at most Δ terms. By linearity of expectation, the expected objective value of $\vec{\gamma}^{(1)}(G)$ is

$$\mathbb{E}\left[\sum_{v\in V} \pi(v) \max_{u:u\to v} \left(y(u) - y(v)\right)^2\right] \le 2\log\Delta \cdot \sum_{v \in V} \pi(v) \max_{u:u\to v} \|f(u) - f(v)\|^2 = 2\vec{\gamma}(G)\cdot \log\Delta.$$

Therefore, by Markov's inequality,

$$\Pr\left[\sum_{v\in V} \pi(v) \max_{u:u\to v} \left(y(u) - y(v)\right)^2 \ge 48 \log\Delta \cdot \vec{\gamma}(G)\right] \le \frac{1}{24}.$$

Next, by applying Fact 2.10.7 with $Y_v = \sqrt{\pi(v)} \cdot y(v)$, it follows that

$$\mathbb{E}\left[\sum_{v\in V} \pi(v) y(v)^2\right] = \sum_{v\in V} \pi(v) \|f(v)\|^2 = 1 \quad\implies\quad \Pr\left[\sum_{v\in V}\pi(v)y(v)^2 \ge \frac{1}{2}\right] \ge \frac{1}{12}.$$

Finally, since $\sum_{v\in V} \pi(v) f(v) = \vec{0}$, it holds that

$$\sum_{v\in V} \pi(v) y(v) = \sum_{v\in V} \pi(v) \langle f(v), h\rangle = \left\langle \sum_{v\in V} \pi(v) f(v), h\right\rangle = 0.$$

Therefore, with probability at least $\frac{1}{24}$, all of these events hold simultaneously. The second event

$$\sum_{v\in V} \pi(v) y(v)^2 \ge \frac{1}{2}$$

means we can rescale y by a factor of at most $\sqrt{2}$, so that the constraint $\sum_{v\in V} \pi(v) y(v)^2 = 1$ is satisfied and the objective value is at most $96 \log\Delta \cdot \vec{\gamma}(G)$. Hence we conclude that $\vec{\gamma}^{(1)}(G) \lesssim \vec{\gamma}(G) \cdot \log \Delta$. □

4.3.3 Cheeger Rounding for Vertex Expansion

We generalize Theorem 3.2.7 to weighted vertex expansion. Our proof does not use the concept of matching conductance in [OZ22]; rather, it is based on a more traditional analysis as in [BHT00] using the directed program $\vec{\gamma}^{(1)}(G)$ in Lemma 4.3.5.

Theorem 4.3.7 (Cheeger Inequality for Weighted Vertex Expansion)**.** *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\psi(G)^2 \lesssim \gamma^{(1)}(G) \lesssim \psi(G).$$

The organization is as follows. We will prove the easy direction Lemma 4.3.8 in Section A.1. For the hard direction, we will work on $\vec{\gamma}^{(1)}(G)$ instead. We will follow the two-step proof template for Theorem 3.1.1.[2] First, we relate $\vec{\gamma}^{(1)}(G)$ with an ℓ_1 version of the program, denoted $\vec{\eta}(G)$. We will show that

$$\vec{\eta}(G)^2 \lesssim \vec{\gamma}^{(1)}(G).$$

Then in the second step, we define a modified vertex boundary condition for directed graphs and use it for the analysis of the threshold rounding. In the end, to find a set with small vertex expansion in the underlying undirected graph, we clean up the solution obtained from threshold rounding.

Below is the precise statement of the easy direction. The proof is deferred to Section A.1.

Lemma 4.3.8 (Easy Direction)**.** *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\gamma^{(1)}(G) \le 2\psi(G).$$

We now turn to proving the hard direction.

Step 1 (ℓ_2^2 to ℓ_1). In the first step, we would like to show that $\vec{\eta}(G)^2 \lesssim \vec{\gamma}^{(1)}(G)$, where $\vec{\gamma}^{(1)}(G)$ is defined in Lemma 4.3.5 (with $f : V \to \mathbb{R}$) and $\vec{\eta}(G)$ is defined as

$$\begin{aligned} \vec{\eta}(G) := \min_{h:V\to\mathbb{R}} \min_{\overrightarrow{E}} \quad & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} |h(u) - h(v)| \\ \text{subject to} \quad & \sum_{v\in V} \pi(v)|h(v)| = 1 \\ & \max\big(\pi(\{v \in V : h(v) < 0\}), \pi(\{v \in V : h(v) > 0\})\big) \le \frac{1}{2}. \end{aligned}$$

[2] Note that this is different from the proof flow in [KLT22].

The last constraint is equivalent to requiring that 0 be a π-weighted median of h. It is the appropriate "balance" constraint for ℓ_1 programs and ensures that the set S produced by threshold rounding in the second step satisfies $\pi(S) \leq 1/2$, without needing any truncation of the solution h.

Let $(f, \overrightarrow{E})$ be a feasible solution to $\vec{\gamma}^{(1)}(G)$ and let $c \in \mathbb{R}$ be a π-weighted median of f, and define $h : V \to \mathbb{R}$ as in (3.1) in Theorem 3.1.1. Clearly, 0 is a π-weighted median of h, and comparing the constraints we have

$$\sum_{v \in V} \pi(v)|h(v)| = \sum_{v \in V} \pi(v)(f(v) - c)^2 \geq \sum_{v \in V} \pi(v)f(v)^2 = 1, \tag{4.2}$$

where the inequality follows from $\sum_{v \in V} \pi(v)f(v) = 0$.

To compare the objective values, note that (3.2) holds here as well, and so

$$\begin{aligned}
& \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} |h(u) - h(v)| \\
\leq\ & \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} |f(u) - f(v)| \left(|f(u) - c| + |f(v) - c|\right) \qquad \text{(by (3.2))} \\
\leq\ & \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} |f(u) - f(v)| \left(2|f(v) - c| + |f(u) - f(v)|\right) \\
\overset{(*)}{\leq}\ & \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} (f(u) - f(v))^2 + 2\sqrt{\sum_{v \in V} \pi(v)(f(v) - c)^2 \cdot \sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} (f(u) - f(v))^2} \\
=\ & \vec{\gamma}^{(1)}(G) + 2\sqrt{\sum_{v \in V} \pi(v)(f(v) - c)^2 \cdot \vec{\gamma}^{(1)}(G)},
\end{aligned}$$

where the step $(*)$ uses the Cauchy-Schwarz inequality. Combining this with (4.2),

$$\begin{aligned}
\frac{\sum_{v \in V} \pi(v) \max_{u:uv \in \overrightarrow{E}} |h(u) - h(v)|}{\sum_{v \in V} \pi(v)|h(v)|} &\leq \vec{\gamma}^{(1)}(G) + \frac{2\sqrt{\sum_{v \in V} \pi(v)(f(v) - c)^2 \cdot \vec{\gamma}^{(1)}(G)}}{\sum_{v \in V} \pi(v)(f(v) - c)^2} \\
&= \vec{\gamma}^{(1)}(G) + 2\sqrt{\frac{\vec{\gamma}^{(1)}(G)}{\sum_{v \in V} \pi(v)(f(v) - c)^2}} \\
&\leq \vec{\gamma}^{(1)}(G) + 2\sqrt{\vec{\gamma}^{(1)}(G)} \lesssim \sqrt{\vec{\gamma}^{(1)}(G)},
\end{aligned}$$

where the last asymptotic inequality is because

$$\vec{\gamma}^{(1)}(G) \leq 2\gamma^{(1)}(G) \leq 4\psi(G) \leq 4$$

by Lemma 4.3.4, Lemma 4.3.8, and the definition of $\psi(G)$. Therefore, we can scale down h so that $\sum_{v\in V}\pi(v)|h(v)| = 1$ and take the same $\overrightarrow{E}$, and $(h, \overrightarrow{E})$ is a feasible solution to $\vec{\eta}(G)$, thus certifying that

$$\vec{\eta}(G) \leq \sum_{v\in V}\pi(v)\max_{u:uv\in\overrightarrow{E}}|h(u)-h(v)| \lesssim \sqrt{\vec{\gamma}^{(1)}(G)},$$

as we have set out to prove.

Step 2 (threshold rounding). For the threshold rounding, we first define the appropriate "vertex cover" boundary $\vec{\partial}_\tau S$ for the analysis of the directed program $\vec{\eta}(G)$. Note that, unlike ∂S, $\vec{\partial}_\tau S$ may contain vertices in S. A good interpretation is to think of $\vec{\partial}_\tau S$ as a vertex cover of the edge boundary $\delta(S)$ in the undirected sense.

Definition 4.3.9 (Directed Vertex Cover Boundary and Expansion)**.** *Let $G = (V, \overrightarrow{E})$ be a directed graph, and π be a distribution on V. For $S \subseteq V$, define the directed vertex cover boundary and the directed vertex cover expansion as*

$$\vec{\partial}_\tau S := \left\{v\in S \mid \exists u\notin S \text{ with } uv\in\overrightarrow{E}\right\} \cup \left\{v\notin S \mid \exists u\in S \text{ with } uv\in\overrightarrow{E}\right\}$$

and $\vec{\psi}_\tau(S) := \pi(\vec{\partial}_\tau S)/\pi(S)$.

The main step is to prove that threshold rounding will find a set S with small directed vertex cover expansion $\vec{\psi}_\tau(S)$.

Proposition 4.3.10 (Threshold Rounding for $\vec{\eta}(G)$)**.** *Let $G = (V, E)$ be an undirected graph and π be a distribution on V. Given a solution $(h, \overrightarrow{E})$ to $\vec{\eta}(G)$ with*

$$\frac{\sum_{v\in V}\pi(v)\max_{u:uv\in\overrightarrow{E}}|h(u)-h(v)|}{\sum_{v\in V}\pi(v)|h(v)|} = \eta_h,$$

threshold rounding on h yields a set $S \subseteq V$ with $0 < \pi(S) \leq 1/2$ and $\vec{\psi}_\tau(S) \lesssim \eta_h$.

Proof. Let $t \in \mathbb{R}$ be a parameter, and define $S_t \subseteq V$ as in (3.3) in Theorem 3.1.1. Note that, since 0 is a π-weighted median of h, $\pi(S_t)$ is at most $1/2$ for any $t \in \mathbb{R}$. The "average" π-weight of S_t is

$$\int_{-\infty}^{\infty}\pi(S_t)\,dt = \sum_{v\in V}\pi(v)\int_{-\infty}^{\infty}\mathbb{1}[v\in S_t]\,dt = \sum_{v\in V}\pi(v)|h(v)|.$$

Since

$$\int_{-\infty}^{\infty} \pi(\vec{\partial}_\tau S_t)\, dt = \sum_{v\in V} \pi(v) \int_{-\infty}^{\infty} \mathbb{1}[v \in \vec{\partial}_\tau S_t]\, dt,$$

to analyze the "average" size of $\vec{\partial}_\tau S_t$ we investigate when a given vertex v is included in $\vec{\partial}_\tau S_t$.

Verify that a vertex v is in $\vec{\partial}_\tau S_t$ if and only if $\min\{h(u) \mid uv \in \overrightarrow{E}\} \le t \le \max\{h(u) \mid uv \in \overrightarrow{E}\}$, where we recall the assumption that every vertex has a self loop, and so $vv \in \overrightarrow{E}$ and thus $\min\{h(u) \mid uv \in \overrightarrow{E}\} \le h(v) \le \max\{h(u) \mid uv \in \overrightarrow{E}\}$. Therefore, the "average" size of $\vec{\partial}_\tau S_t$ is

$$\begin{aligned}
\int_{-\infty}^{\infty} \pi(\vec{\partial}_\tau S_t)\, dt &= \sum_{v\in V} \pi(v) \int_{-\infty}^{\infty} \mathbb{1}[v \in \vec{\partial}_\tau S_t]\, dt \\
&= \sum_{v\in V} \pi(v) \left(\max_{u:uv\in \overrightarrow{E}} h(u) - \min_{u:uv\in \overrightarrow{E}} h(u) \right) \\
&\le 2\sum_{v\in V} \pi(v) \max_{u:uv\in \overrightarrow{E}} |h(u) - h(v)|.
\end{aligned}$$

Note that $\pi(\vec{\partial}_\tau S) = 0$ whenever $\pi(S) = 0$. Therefore, $S = S_t$ for some $t \in \mathbb{R}$ satisfies $0 < \pi(S) \le 1/2$ and

$$\vec{\psi}_\tau(S) \le \frac{\int_{-\infty}^{\infty} \pi(\vec{\partial}_\tau S_t)\, dt}{\int_{-\infty}^{\infty} \pi(S_t)\, dt} \le \frac{2\sum_{v\in V} \pi(v) \max_{u:uv\in \overrightarrow{E}} |h(u) - h(v)|}{\sum_{v\in V} \pi(v)|h(v)|} = 2\eta_h.$$

This completes the proof. □

Finally, given a set S with small directed vertex cover expansion $\vec{\psi}_\tau(S)$, we show how to find a set $S' \subseteq S$ with small vertex expansion $\psi(S')$. This step is similar to the step in [OZ22] from matching conductance to vertex expansion (c.f. Proposition 3.2.10).

Lemma 4.3.11 (Postprocessing for Vertex Expansion)**.** *Let $G = (V, \overrightarrow{E})$ be a directed graph. Given a set S with $\vec{\psi}_\tau(S) < 1/2$, there is a set $S' \subseteq S$ with $\psi(S') \le 2\vec{\psi}_\tau(S)$ in the underlying undirected graph of G.*

Proof. From Definition 4.3.9, the observation is that all undirected edges in $\delta(S)$ are incident to at least one vertex in $\vec{\partial}_\tau S$. Define $S' := S - \vec{\partial}_\tau S$. Then observe that $\partial S' \subseteq \vec{\partial}_\tau S$, as there are no incoming edges to S' from $V - (S' \cup \vec{\partial}_\tau S)$ and all outgoing edges from S' go to $\vec{\partial}_\tau(S)$. This implies that

$$\pi(\partial S') \leq \pi\left(\vec{\partial}_\tau S\right) = \vec{\psi}_\tau(S) \cdot \pi(S) \leq 2\vec{\psi}_\tau(S) \cdot \pi(S'),$$

where the last inequality uses the assumption that $\vec{\psi}_\tau(S) = \pi\big(\vec{\partial}_\tau S\big)/\pi(S) < 1/2$ and so $\pi(S') \geq \pi(S) - \pi\big(\vec{\partial}_\tau S\big) \geq \pi(S)/2$. We conclude that $\psi(S') \leq 2\vec{\psi}_\tau(S)$. $\square$

We put together the steps and complete the proof of Theorem 4.3.7 in Section A.1.

4.4 Cheeger Inequality for Bipartite Vertex Expansion

The goal of this section is to prove Theorem 4.1.7, which relates the maximum reweighted lower spectral gap $\zeta^*(G)$ in Definition 4.1.6 and the bipartite vertex expansion $\psi_B(G)$ in Definition 4.1.5. The proof follows closely the proof of Theorem 4.1.3 in the previous section, so some steps will be stated without proofs, and the focus will be on the threshold rounding step.

4.4.1 Primal and Dual Programs

The primal program $\zeta^*(G)$ in Definition 4.1.6 has the following dual program which is similar to $\gamma(G)$ in Proposition 4.3.1.

Proposition 4.4.1 (Dual Program for Lower Spectral Gap [Roc05])**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the following semidefinite program is dual to the primal program in Definition 4.1.6 with strong duality $\zeta^*(G) = \nu(G)$ where*

$$\begin{aligned}\nu(G) := \min_{f:V\to\mathbb{R}^n,\ g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} \pi(v) g(v) \\ \text{subject to} \quad & \sum_{v\in V} \pi(v) \|f(v)\|^2 = 1 \\ & g(u) + g(v) \geq \|f(u) + f(v)\|^2 \qquad \forall uv \in E.\end{aligned}$$

There are two differences between $\gamma(G)$ in Proposition 4.3.1 and $\nu(G)$ in Proposition 4.4.1. One is that the constraint $g(u) + g(v) \geq \|f(u) - f(v)\|^2$ in $\gamma(G)$ is replaced by the constraint $g(u) + g(v) \geq \|f(u) + f(v)\|^2$ in $\nu(G)$, which are handled in a similar way. The other is that the constraint of $\sum_{v \in V} \pi(v) f(v) = \vec{0}$ in $\gamma(G)$ is not present in $\nu(G)$.

The nice form of the dual program $\nu(G)$ is the main reason behind the definition of the primal program $\zeta^*(G)$. By the variational characterization of eigenvalues, the quadratic form of $D_P + P$, and the π-reversibility of P,

$$\lambda_{\min}(D_P + P) = \min_{x:V \to \mathbb{R}} \frac{x^T(D_P + P)x}{x^T x} = \min_{x:V \to \mathbb{R}} \frac{\sum_{uv \in E} \pi(u)P(u,v)(x(u) + x(v))^2}{\sum_{u \in V} \pi(u)x(u)^2},$$

and this is the intermediate form we need to derive the dual; see [Roc05].

Later, as in Section 4.3.1, we will define a directed dual program $\vec{\nu}(G)$, and the dual constraint $g \geq 0$ crucially enables us to relate the two programs. The dual constraint $g \geq 0$ comes from the primal constraint $\sum_{v \in V} P(u,v) \leq 1$, whereas if we use $\sum_{v \in V} P(u,v) = 1$ then g will be unconstrained.

For $\lambda_2^*(G)$, we sidestep the issue by adding self loops to each vertex of G. The non-negativity of g in the dual program $\gamma(G)$ follows indirectly from $g(u) + g(u) \geq \|f(u) - f(u)\|^2$, as now $(u,u) \in E$. Moreover, adding self loops does not change the vertex expansion of G. Therefore, $\lambda_2^*(G)$ can take the more natural form where P does correspond to a transition matrix. However, we cannot do the same for $\zeta^*(G)$, because the additional constraint on g becomes $g(u) + g(u) \geq \|f(u) + f(u)\|^2$ which changes the objective value, and also that adding self loops takes away the bipartiteness of subgraphs. This is why we have to explicitly impose the non-negativity of g here.

4.4.2 Proof of Theorem 4.1.7

We use the same two-step plan as in Section 4.3. First, we project the solution to the dual program in Proposition 4.4.1 into a 1-dimensional solution to the following program.

Definition 4.4.2 (One-Dimensional Dual Program for Lower Spectral Gap). *Given an undirected graph $G = (V,E)$ and a probability distribution π on V, $\nu^{(1)}(G)$ is defined as*

$$\begin{aligned} \min_{f:V \to \mathbb{R},\ g:V \to \mathbb{R}_{\geq 0}} \quad & \sum_{v \in V} \pi(v) g(v) \\ \text{subject to} \quad & \sum_{v \in V} \pi(v) \|f(v)\|^2 = 1 \\ & g(u) + g(v) \geq \|f(u) + f(v)\|^2 \qquad \forall uv \in E. \end{aligned}$$

As in Proposition 4.3.6, we use the Gaussian projection method in [LRV13] to prove the following guarantee.

Proposition 4.4.3 (Gaussian Projection for $\nu(G)$)**.** *For any undirected graph $G = (V, E)$ with maximum degree Δ and any probability distribution π on V,*

$$\nu(G) \leq \nu^{(1)}(G) \lesssim \nu(G) \cdot \log \Delta.$$

In the second step, we prove a Cheeger-type inequality relating $\psi_B(G)$ and $\nu(G)$.

Theorem 4.4.4. *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\psi_B(G)^2 \lesssim \nu^{(1)}(G) \lesssim \psi_B(G).$$

Combining Proposition 4.4.1 and Proposition 4.4.3 and Theorem 4.4.4 gives

$$\psi_B(G)^2 \lesssim \nu^{(1)}(G) \lesssim \nu(G) \cdot \log \Delta = \zeta^*(G) \cdot \log \Delta \quad \text{and} \quad \zeta^*(G) = \nu(G) \leq \nu^{(1)}(G) \lesssim \psi_B(G),$$

proving Theorem 4.1.7. We will prove Proposition 4.4.3 and Theorem 4.4.4 in the following subsections.

4.4.3 Dual Program on Graph Orientation

As in Section 4.3.1, we introduce a directed program for the analysis of both steps.

Definition 4.4.5 (Directed Dual Programs for $\nu(G)$)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V,*

$$\begin{aligned} \vec{\nu}(G) := \min_{f: V \to \mathbb{R}^n,\ g: V \to \mathbb{R}_{\geq 0}} \quad & \sum_{v \in V} \pi(v) g(v) \\ \text{subject to} \quad & \sum_{v \in V} \pi(v) \left\| f(v) \right\|^2 = 1 \\ & \max\{g(u), g(v)\} \geq \left\| f(u) + f(v) \right\|^2 \qquad \forall uv \in E. \end{aligned}$$

$\vec{\nu}^{(1)}(G)$ is defined as the 1-dimensional program of $\vec{\nu}(G)$ where $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

As in Lemma 4.3.4, we show that $\nu(G)$ and $\vec{\nu}(G)$ are closely related. The proof is the same as in Lemma 4.3.4 and is omitted, but note that $g \geq 0$ is needed.

Lemma 4.4.6. *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\nu(G) \leq \vec{\nu}(G) \leq 2\nu(G) \quad \text{and} \quad \nu^{(1)}(G) \leq \vec{\nu}^{(1)}(G) \leq 2\nu^{(1)}(G).$$

As in Lemma 4.3.5, we use an orientation of the edges to eliminate the variables $g(v)$ for $v \in V$ in $\vec{\nu}(G)$. The proof is the same as in Lemma 4.3.5 and is omitted, but note that $g \geq 0$ is needed.

Lemma 4.4.7 (Directed Dual Programs Using Orientation for $\nu(G)$)**.** *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of the undirected edges in E. Then*

$$\begin{aligned} \vec{\nu}(G) = \min_{f:V\to\mathbb{R}^n} \min_{\overrightarrow{E}} \quad & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} \|f(u) + f(v)\|^2 \\ \text{subject to} \quad & \sum_{v\in V} \pi(v) \|f(v)\|^2 = 1. \end{aligned}$$

Similarly, $\vec{\nu}^{(1)}(G)$ can be written in the same form with $f : V \to \mathbb{R}$ instead of $f : V \to \mathbb{R}^n$.

Once we have this formulation using orientation, we can use the same proof as in Proposition 4.3.6 to show that $\vec{\nu}(G) \leq \vec{\nu}^{(1)}(G) \lesssim \log \Delta \cdot \vec{\nu}(G)$, and thus Proposition 4.4.3 follows from Lemma 4.4.6 and we omit the proof. It remains to prove Theorem 4.4.4, which will be done in the next subsection.

4.4.4 Cheeger Rounding for Bipartite Vertex Expansion

The goal of this subsection is to prove Theorem 4.4.4. We will prove the following easy direction in Section A.2.

Lemma 4.4.8 (Easy Direction)**.** *For any undirected graph $G = (V, E)$ and any probability distribution π on V,*

$$\nu^{(1)}(G) \leq 2\psi_B(G).$$

For the hard direction, we will work with $\vec{\nu}^{(1)}(G)$ instead. We will follow the two-step proof template for Theorem 3.1.1.[3] First, we relate $\vec{\nu}^{(1)}(G)$ with an ℓ_1 version of the program $\vec{\xi}(G)$ that

$$\vec{\xi}(G)^2 \lesssim \vec{\nu}^{(1)}(G).$$

[3] Again, note that this is different from the proof flow in [KLT22].

Then, in the second step, we define a modified bipartite vertex expansion condition for directed graphs and use it for the analysis of the threshold rounding.

We now begin the proof of the hard direction.

Step 1 (ℓ_2^2 to ℓ_1). In the first step, we would like to show that $\vec{\xi}(G)^2 \lesssim \vec{\nu}^{(1)}(G)$, where $\vec{\nu}^{(1)}(G)$ is defined in Lemma 4.4.7 (with $f : V \to \mathbb{R}$) and $\vec{\xi}(G)$ is defined as

$$\begin{aligned} \vec{\xi}(G) := \min_{h:V\to\mathbb{R}} \min_{\overrightarrow{E}} \quad & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} |h(u)+h(v)| \\ \text{subject to} \quad & \sum_{v\in V} \pi(v)|h(v)| = 1. \end{aligned}$$

Let $(f, \overrightarrow{E})$ be a feasible solution to $\vec{\nu}^{(1)}(G)$. Define $h : V \to \mathbb{R}$ as in the proof of Trevisan's result in Theorem 3.1.4, i.e.

$$h(u) := \begin{cases} f(u)^2 & \text{if } f(u) \geq 0 \\ -f(u)^2 & \text{if } f(u) < 0. \end{cases}$$

Clearly, $\sum_{v\in V} \pi(v)|h(v)| = \sum_{v\in V} \pi(v) f(v)^2 = 1$, and by the inequality (3.4) we have

$$\begin{aligned} & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} |h(u)+h(v)| \\ \leq\ & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} |f(u)+f(v)| \left(|f(u)|+|f(v)|\right) \\ \leq\ & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} |f(u)+f(v)| \left(2|f(v)|+|f(u)+f(v)|\right) \\ \overset{(*)}{\leq}\ & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} (f(u)+f(v))^2 + 2\sqrt{\sum_{v\in V} \pi(v) f(v)^2 \cdot \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} (f(u)+f(v))^2} \\ =\ & \vec{\nu}^{(1)}(G) + 2\sqrt{\vec{\nu}^{(1)}(G)} \lesssim \sqrt{\vec{\nu}^{(1)}(G)}, \end{aligned}$$

where the step $(*)$ uses the Cauchy-Schwarz inequality and the last asymptotic inequality is because

$$\vec{\nu}^{(1)}(G) \leq 2\nu^{(1)}(G) \leq 4\psi_B(G) \leq 4$$

by Lemma 4.4.6, Lemma 4.4.8, and the definition of $\psi_B(G)$. Therefore, $(h, \overrightarrow{E})$ is a feasible solution to $\vec{\xi}^{(1)}(G)$, and certifies that

$$\vec{\xi}^{(1)}(G) \leq \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} |h(u)+h(v)| \lesssim \sqrt{\vec{\nu}^{(1)}(G)}.$$

Step 2 (threshold rounding). Let S_1, S_2 be two disjoint subsets of V. In the edge conductance setting, rephrasing using our terminology, Trevisan [Tre09] defined the "bipartite edge boundary" $\delta(S_1, S_2)$ as $E(S_1) \cup E(S_2) \cup \delta(S_1 \cup S_2)$ where $E(S_i)$ is the set of induced edges in S_i for $i \in \{1, 2\}$, and the "bipartite edge conductance" $\phi(S_1, S_2)$ as $|\delta(S_1, S_2)|/\mathrm{vol}(S_1 \cup S_2)$ (see Section 3.1.2). We define the appropriate "bipartite vertex cover boundary" $\vec{\partial}_\tau(S_1, S_2)$ for vertex expansion and for directed graphs in the following definition. As in Definition 4.3.9, note that $\vec{\partial}_\tau(S_1, S_2)$ could contain vertices in $V \setminus (S_1 \cup S_2)$. Again, a good intuition is to think of $\vec{\partial}_\tau(S_1, S_2)$ as a vertex cover of the edges in the bipartite edge boundary $\delta(S_1, S_2)$ in the undirected sense.

Definition 4.4.9 (Directed Bipartite Vertex Cover Boundary and Expansion)**.** *Let $G = (V, \overrightarrow{E})$ be a directed graph. Let S_1, S_2 be two disjoint subsets of V. The directed bipartite vertex cover boundary of (S_1, S_2) is defined as*

$$\begin{aligned}\vec{\partial}_\tau(S_1, S_2) \quad := \quad & \left\{v \in S_1 \mid \exists u \notin S_2 \text{ with } uv \in \overrightarrow{E}\right\} \cup \\ & \left\{v \in S_2 \mid \exists u \notin S_1 \text{ with } uv \in \overrightarrow{E}\right\} \cup \\ & \left\{v \notin S_1 \cup S_2 \mid \exists u \in S_1 \cup S_2 \text{ with } uv \in \overrightarrow{E}\right\},\end{aligned}$$

and the directed bipartite vertex cover expansion as

$$\vec{\psi}_\tau(S_1, S_2) := \frac{\pi\big(\vec{\partial}_\tau(S_1, S_2)\big)}{\pi(S_1 \cup S_2)}.$$

An example of directed bipartite vertex cover expansion is provided in Figure 4.1.

We prove that threshold rounding as in Section 3.1.2, when applied on $\vec{\xi}^{(1)}(G)$, will give a set with small directed bipartite vertex cover expansion.

Proposition 4.4.10 (Threshold Rounding for $\vec{\xi}(G)$)**.** *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Given a solution $(h, \overrightarrow{E})$ with*

$$\sum_{v \in V} \pi(v) \max_{u : uv \in \overrightarrow{E}} |h(u) + h(v)| = \xi_h \quad \textit{and} \quad \sum_{v \in V} \pi(v)|h(v)| = 1,$$

threshold rounding on h yields two disjoint subsets $S_1, S_2 \subseteq V$ with $\vec{\psi}_\tau(S_1, S_2) \lesssim \xi_h$.

Proof. Let $t \in \mathbb{R}_{\geq 0}$ be a parameter. Define $S_t := \{v \in V \mid h(v) > t\}$ and $S_{-t} := \{v \in V \mid h(v) < -t\}$. Then, the "average" denominator value is

$$\int_0^\infty \pi(S_t \cup S_{-t})\, dt = \sum_{v \in V} \pi(v)|h(v)| = 1.$$

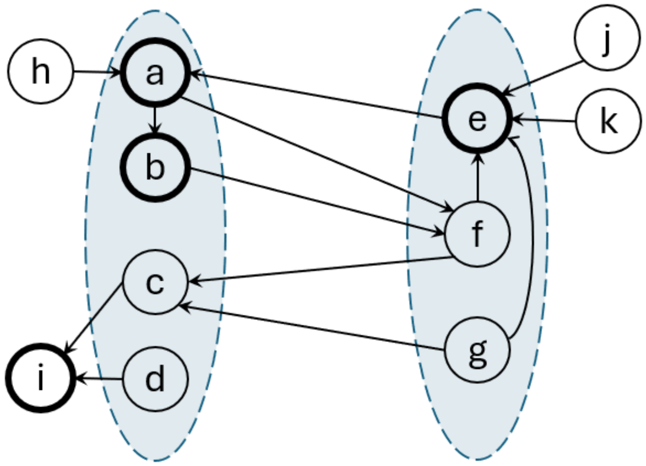


Figure 4.1: In the graph shown, the bipartitions are $S_1 = \{a, b, c, d\}$ and $S_2 = \{e, f, g\}$. The vertices a, b, e, i, with thick outlines, are the vertices in $\vec{\partial}_\tau(S_1, S_2)$.

For the numerator, we consider when a vertex v is in $\vec{\partial}_\tau(S_t, S_{-t})$. Assume without loss that $h(v) \geq 0$; the other case is symmetric. There are two scenarios where $v \in \vec{\partial}_\tau(S_t, S_{-t})$:

1. The first scenario is when $uv \in E[S_t]$ in the undirected sense for some directed edge $u \to v$. For a fixed edge $u \to v$, this happens when $t < \min(h(u), h(v)) \leq |h(u) + h(v)|/2$ (note that $h(u) \geq 0$). Therefore, the first scenario happens when

$$t < \max_{u:uv\in \vec{E}} \max\{0, \min(h(u), h(v))\} \leq \frac{1}{2} \max_{u:uv\in \vec{E}} |h(u) + h(v)|.$$

2. The second scenario is when $uv \in \delta(S_t \cup S_{-t})$ in the undirected sense for some directed edge $u \to v$. For a fixed edge $u \to v$, this happens when $|h(u)| \leq t < |h(v)|$ (so $u \notin S_t \cup S_{-t}$ and $v \in S_t \cup S_{-t}$), or when $|h(v)| \leq t < |h(u)|$ (so $u \in S_t \cup S_{-t}$ and $v \notin S_t \cup S_{-t}$). Therefore, the second scenario happens when

$$\min_{u:u\to v} |h(u)| \leq t < |h(v)| \qquad \text{or} \qquad |h(v)| \leq t < \max_{u:u\to v} |h(u)|.$$

Therefore, the "average" numerator value is

$$\begin{aligned}
& \int_0^\infty \pi\big(\vec{\partial}_\tau(S_t, S_{-t})\big)\, dt \\
= \; & \sum_{v \in V} \pi(v) \cdot \int_0^\infty \mathbb{1}\Big[v \in \vec{\partial}_\tau(S_t, S_{-t})\Big]\, dt \\
\leq \; & \sum_{v \in V} \pi(v) \Bigg[\frac{1}{2} \max_{u: u \to v} |h(u) + h(v)| + \max\left(0, |h(v)| - \min_{u: u \to v} |h(u)|\right) \\
& \qquad + \max\left(0, \max_{u: u \to v} |h(u)| - |h(v)|\right)\Bigg] \\
\leq \; & \sum_{v \in V} \pi(v) \left[\frac{1}{2} \max_{u: u \to v} |h(u) + h(v)| + \max_{u: u \to v} |h(u) + h(v)| + \max_{u: u \to v} |h(u) + h(v)|\right] \\
\leq \; & \frac{5}{2} \xi_h.
\end{aligned}$$

We conclude that there exists (S_t, S_{-t}) with $\vec{\psi}_\tau(S_t, S_{-t}) \lesssim \xi_h$. □

Finally, as in Lemma 4.3.11, given (S_1, S_2) with small directed bipartite vertex expansion, we show how to extract an induced bipartite graph with small vertex expansion.

Lemma 4.4.11 (Postprocessing for Bipartite Vertex Expansion)**.** *Let $G = (V, \overrightarrow{E})$ be a directed graph. Given two disjoint subsets $S_1, S_2 \subseteq V$ with $\vec{\psi}_\tau(S_1, S_2) < 1/2$, there are $S_1' \subseteq S_1$ and $S_2' \subseteq S_2$ with $\psi(S_1' \cup S_2') \leq 2\vec{\psi}_\tau(S_1, S_2)$ and $S_1' \cup S_2'$ is an induced bipartite graph in the underlying undirected graph of G.*

Proof. From Definition 4.4.9, the observation is that $\vec{\partial}_\tau(S_1, S_2)$ is a vertex cover of $E(S_1) \cup E(S_2) \cup \delta(S_1 \cup S_2)$. So, by setting $S_1' := S_1 \setminus \vec{\partial}_\tau(S_1, S_2)$ and $S_2' := S_2 \setminus \vec{\partial}_\tau(S_1, S_2)$, then (S_1', S_2') is an induced bipartite graph as there could be no edges induced in S_1' and no edges induced in S_2'. Also, $\partial(S_1' \cup S_2') \subseteq \vec{\partial}_\tau(S_1, S_2)$, as there could be no edges between $S_1' \cup S_2'$ and $V \setminus (S_1 \cup S_2 \cup \vec{\partial}_\tau(S_1, S_2))$. Therefore,

$$\pi(\partial(S_1' \cup S_2')) \leq \pi\big(\vec{\partial}_\tau(S_1, S_2)\big) = \vec{\psi}_\tau(S_1, S_2) \cdot \pi(S_1 \cup S_2) \leq 2\vec{\psi}_\tau(S_1, S_2) \cdot \pi(S_1' \cup S_2'),$$

where the last inequality uses the assumption that $\vec{\psi}_\tau(S_1, S_2) = \pi\big(\vec{\partial}_\tau(S_1, S_2)\big)/\pi(S_1 \cup S_2) < 1/2$ and so $\pi(S_1' \cup S_2') \geq \pi(S_1 \cup S_2) - \pi\big(\vec{\partial}_\tau(S_1, S_2)\big) \geq \pi(S_1 \cup S_2)/2$. We thus conclude that $\psi(S_1' \cup S_2') \leq 2\vec{\psi}_\tau(S_1, S_2)$. □

We complete the proof of Theorem 4.4.4 in Section A.2.

4.5 Higher-Order Cheeger Inequality for Vertex Expansion

The goal of this section is to prove Theorem 4.1.10. There are four main steps in the proof.

The first step is to reformulate the problem as a semidefinite program using the maximum reweighted sum of the k smallest eigenvalues $\sigma_k^*(G)$. Using von Neumann's minimax theorem, we construct the dual program of $\sigma_k^*(G)$ and see that it satisfies the so-called sub-isotropy condition. The main focus in this section will then be to relate $\sigma_k^*(G)$ and $\psi_k(G)$, rather than to relate $\lambda_k^*(G)$ and $\psi_k(G)$ directly.

The second step is to project the dual solution to $\sigma_k^*(G)$ into a low-dimensional solution. In this step, we use a similar apporach as in Section 4.3, by introducing an intermediate directed dual program and then using the Gaussian projection method. Also, we use a theorem in [LOT12] that proves that Gaussian projection approximately preserves the sub-isotropy condition.

The third step is to partition the low-dimensional solution into k disjointly supported functions each with small objective value. In this step, we closely follow the techniques in [LOT12], such as radial projection distance, smooth localization and random parititoning. A review of the techniques in [LOT12] can be found in Section 3.1.3.

The last step is to apply the Cheeger inequality for vertex expansion in Section 4.3 on these k functions to find disjoint sets with small vertex expansion. We will prove the easy direction and put together the steps to prove Theorem 4.1.10 in Section 4.5.4.

The last three steps mirror the three steps in the proof of Theorem 3.1.6, and we suggest reading Section 3.1.3 before continuing with this section.

4.5.1 Primal and Dual Programs

As mentioned in Section 4.2, the maximum reweighted k-th smallest eigenvalue $\lambda_k^*(G)$ as formulated in Definition 4.1.9 is not a convex program. Instead, we will study the following related quantity.

Definition 4.5.1 (Maximum Reweighted Sum of k Smallest Eigenvalues)**.** *Given an undirected graph $G = (V, E)$ and a probability distribution π on V, the maximum reweighted sum of k smallest eigenvalues of the normalized Laplacian matrix of G is defined as $\sigma_k^*(G) := \max_{P \geq 0} \sum_{i=1}^k \lambda_k(I - P)$, where P is subject to the same constraints stated in*

Definition 1.1.1. Note that

$$\lambda_k^*(G) \le \sigma_k^*(G) \le k \cdot \lambda_k^*(G).$$

We reformulate the primal program in Definition 4.5.1 as the semidefinite program in Proposition 4.5.2. The proof of Proposition 4.5.2 has a few small steps. First we rewrite the sum $\sum_{i=1}^k \lambda_k(I-P)$ as $\sum_{i=1}^k \lambda_k(I-\mathcal{Q})$ for a symmetric matrix $\mathcal{Q}$. Then we use Proposition 2.8.3 to write $\sum_{i=1}^k \lambda_k(I-\mathcal{Q})$ as a minimization problem using semidefinite programming. Next we apply von Neumann's minimax theorem to change the order of max-min to min-max. Then we do a change of variable and rewrite the program into a vector program form. Finally, we use linear programming duality to rewrite the inner maximization problem as a minimization problem as was done in [Roc05].

Proposition 4.5.2 (Dual Program for $\sigma_k^*(G)$)**.** *For any undirected graph $G=(V,E)$ with a self loop at each vertex and any probability distribution π on V, the following semidefinite program is dual to the primal program in Definition 4.5.1 with strong duality $\sigma_k^*(G)=\kappa(G)$ where*

$$\begin{aligned}
\kappa(G) := \min_{f:V\to\mathbb{R}^n,\ g:V\to\mathbb{R}_{\ge 0}} \quad & \sum_{v\in V} \pi(v) g(v) \\
\text{subject to} \quad & g(u)+g(v) \ge \|f(u)-f(v)\|^2 \quad \forall uv \in E \\
& \sum_{v\in V} \pi(v) f(v) f(v)^T \preceq I_n \\
& \sum_{v\in V} \pi(v) \|f(v)\|^2 = k.
\end{aligned}$$

Proof. By Definition 4.5.1, $\sigma_k^*(G) = \max_{P\ge 0} \sum_{i=1}^k \lambda_i(I-P)$, where the maximum is over all P satisfying the constraints in Definition 1.1.1. Consider the sum of eigenvalues for a fixed P that satisfies the constraints. The time reversible constraint $\pi(u)P(u,v)=\pi(v)P(v,u)$ for all $uv\in E$ is equivalent to the matrix $Q := \Pi P$ being symmetric where $\Pi := \operatorname{diag}(\pi)$. Let $\mathcal{Q} := \Pi^{-1/2} Q \Pi^{-1/2}$ be the normalized adjacency matrix of Q. Note that P and $\mathcal{Q}$ have the same spectrum, as $\mathcal{Q} = \Pi^{-1/2} Q \Pi^{-1/2} = \Pi^{1/2} P \Pi^{-1/2}$. Therefore $\sum_{i=1}^k \lambda_i(I-P) = \sum_{i=1}^k \lambda_i(I-\mathcal{Q})$ where $I-\mathcal{Q}$ is a symmetric matrix.

By Proposition 2.8.3, the sum of the k smallest eigenvalues of the symmetric matrix

$I - \mathcal{Q}$ can be written as the following semidefinite program:

$$\sum_{i=1}^{k} \lambda_i(I - \mathcal{Q}) = \min_{Y \in \mathbb{R}^{n \times n}} \quad \operatorname{tr}\big(Y \cdot (I - \mathcal{Q})\big)$$
$$\text{subject to} \quad 0 \preceq Y \preceq I$$
$$\operatorname{tr}(Y) = k.$$

Note that $I - \mathcal{Q}$ is the normalized Laplacian matrix of Q. We consider the change of variables $Y = \Pi^{1/2} Z \Pi^{1/2}$, so as to rewrite the objective function in terms of $\Pi - Q$ which is the Laplacian matrix of Q:

$$\sum_{i=1}^{k} \lambda_i(I - \mathcal{Q}) = \min_{Z \in \mathbb{R}^{n \times n}} \quad \operatorname{tr}\big(Z \cdot (\Pi - Q)\big)$$
$$\text{subject to} \quad 0 \preceq \Pi^{\frac{1}{2}} Z \Pi^{\frac{1}{2}} \preceq I$$
$$\operatorname{tr}(\Pi^{\frac{1}{2}} Z \Pi^{\frac{1}{2}}) = k.$$

Therefore, the primal program for $\sigma_k^*(G)$ can be rewritten in terms of Q as follows:

$$\begin{aligned}
\sigma_k^*(G) \;=\; \max_{Q \geq 0} \min_{Z \in \mathbb{R}^{n \times n}} \quad & \operatorname{tr}\big(Z \cdot (\Pi - Q)\big) & \\
\text{subject to} \quad & Q(u,v) = 0 & \forall uv \notin E \\
& \sum_{v \in V} Q(u,v) = \pi(u) & \forall u \in V \\
& Q(u,v) = Q(v,u) & \forall uv \in E \\
& 0 \preceq \Pi^{\frac{1}{2}} Z \Pi^{\frac{1}{2}} \preceq I & \\
& \operatorname{tr}\big(\Pi^{\frac{1}{2}} Z \Pi^{\frac{1}{2}}\big) = k. &
\end{aligned}$$

Now we write the dual program by using von Neumann's minimax theorem in Theorem 2.8.1 to switch the max-min to min-max in the objective function. Note that von Neumann's theorem can be applied because the objective function is multi-linear in Z and Q (hence concave in Q and convex in Z), the feasible region of Q is compact and convex as it is bounded and defined by linear constraints, and the feasible region of Z is compact and convex as it is bounded and defined by PSD and trace constraints. Hence, we can switch the order of $\max_Q \min_Z$ to obtain the dual program by rewriting the objective function as

$$\min_{Z \in \mathbb{R}^{n \times n}} \max_{Q \geq 0} \; \operatorname{tr}\big(Z \cdot (\Pi - Q)\big).$$

Next we rewrite this dual program into a vector program form. As $Z \succeq 0$, we can write $Z = FF^T$ where F is an $n \times n$ matrix. We denote the i-th column of F by $f_i \in \mathbb{R}^n$ for $1 \le i \le n$ and think of it as an eigenvector, denote the v-th row of F by $f(v) \in \mathbb{R}^n$ and think of it as the spectral embedding of a vertex v, and denote the (v,i)-th entry of F by $f_i(v)$ for $1 \le i \le n$ and $v \in V$. As $\Pi - Q$ is the Laplacian matrix of Q, the quadratic form for a vector $x \in \mathbb{R}^n$ is $x^T(\Pi - Q)x = \sum_{uv \in E}(x(u) - x(v))^2 \cdot Q(u,v)$, and thus the objective function can be rewritten as

$$\begin{aligned} \operatorname{tr}(F^T(\Pi - Q)F) = \sum_{i=1}^{n} f_i^T(\Pi - Q)f_i &= \sum_{i=1}^{n} \sum_{uv \in E} (f_i(u) - f_i(v))^2 \cdot Q(u,v) \\ &= \sum_{uv \in E} \|f(u) - f(v)\|^2 \cdot Q(u,v). \end{aligned}$$

Note that $\Pi^{1/2} Z \Pi^{1/2} = (\Pi^{1/2}F)(F^T\Pi^{1/2})$ and $F^T \Pi F = (F^T \Pi^{1/2})(\Pi^{1/2}F)$ have the same spectrum by Fact 2.4.2. So the first constraint can be rewritten as

$$0 \preceq F^T \Pi F = \sum_{v \in V} \pi(v) f(v) f(v)^T \preceq I,$$

and the second constraint can be rewritten as

$$\operatorname{tr}(F^T \Pi F) = \operatorname{tr}\Big(\sum_{v \in V} \pi(v) f(v) f(v)^T\Big) = \sum_{v \in V} \pi(v) \|f(v)\|^2 = k.$$

Therefore, the dual program for $\sigma_k^*(G)$ can be rewritten as follows:

$$\begin{aligned} \kappa(G) \ := \quad & \min_{f: V \to \mathbb{R}^n} \max_{Q \ge 0} \quad \sum_{uv \in E} \|f(u) - f(v)\|^2 \cdot Q(u,v) & \\ & \text{subject to} \quad Q(u,v) = 0 & \forall uv \notin E \\ & \sum_{v \in V} Q(u,v) = \pi(u) & \forall u \in V \\ & Q(u,v) = Q(v,u) & \forall uv \in E \\ & \sum_{v \in V} \pi(v) f(v) f(v)^T \preceq I_n & \\ & \sum_{v \in V} \pi(v) \|f(v)\|^2 = k. & \end{aligned}$$

Finally, as in [Roc05], note that the inner maximization problem is just a linear program in Q. For a fixed embedding $f : V \to \mathbb{R}^n$, we can use standard LP duality to rewrite the inner maximization problem into the following minimization problem:

$$\begin{aligned} \min_{g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{v\in V} \pi(v)g(v) \\ \text{subject to} \quad & g(u)+g(v) \geq \|f(u)-f(v)\|^2 \qquad\qquad \forall uv \in E, \end{aligned}$$

where $g(u)$ is a dual variable for the constraint $\sum_{v\in V} Q(u,v) = \pi(u)$. Recall that we assumed the graph has a self-loop $Q(v,v)$ at each vertex v so that the primal program is always feasible, and the primal variable $Q(v,v)$ gives the dual constraint $g(v) \geq 0$.

To summarize, we rewrite the max-min optimization problem in the primal program into a min-min optimization problem using von Neumann's minimax theorem and linear programming duality. The resulting program in the statement is a semidefinite program in the vector program form. □

4.5.2 Gaussian Projection

The second step is to project a solution to $\kappa(G)$ in Proposition 4.5.2 into a low-dimensional solution and prove that several properties are approximately preserved. The projection algorithm is a high dimensional version of the simple Guassian projection algorithm in Section 4.3.2, and is exactly the same as defined in Definition 3.1.7 and used in [LOT12]. We copy the definition here for convenience.

Definition 4.5.3 (Gaussian Projection (restatement of Definition 3.1.7))**.** *Let $f : V \to \mathbb{R}^p$ be an embedding where each vertex v is mapped to a vector $f(v) \in \mathbb{R}^p$. Given an integer $1 \leq h \leq p$, let Γ be an $h \times p$ matrix where each entry $\Gamma_{i,j}$ for $1 \leq i \leq h$ and $1 \leq j \leq p$ is an independent standard Gaussian random variable $N(0,1)$. The Gaussian projection $\bar{f} : V \to \mathbb{R}^h$ of f is an embedding of each vertex $v \in V$ to an h-dimensional vector defined as*

$$\bar{f}(v) = \frac{1}{\sqrt{h}} \cdot \Gamma\big(f(v)\big).$$

As in Section 4.3.2, we consider a related directed program $\vec{\kappa}(G)$ for the analysis of the Gaussian projection algorithm. The proof of the following lemma is the same as in Lemma 4.3.4 and Lemma 4.3.5 and is omitted.

Lemma 4.5.4 (Directed Dual Program Using Orientation for $\kappa(G)$)**.** *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of the undirected edges in E. Define*

$$\begin{aligned} \vec{\kappa}(G) := \min_{f:V\to\mathbb{R}^n} \min_{\overrightarrow{E}} \quad & \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} \|f(u) - f(v)\|^2 \\ \text{subject to} \quad & \sum_{v\in V} \pi(v) f(v) f(v)^T \preceq I_n \\ & \sum_{v\in V} \pi(v) \|f(v)\|^2 = k. \end{aligned}$$

Then $\kappa(G) \le \vec{\kappa}(G) \le 2\kappa(G)$.

As in Section 3.1.3, we shall prove that the energy, the mass, and the spreading property of f to $\vec{\kappa}(G)$ are preserved by $\bar{f}$. Here, the *energy* of the function f is defined as the objective value of $\vec{\kappa}(G)$. It is a modification of the definition in Section 3.1.3 to fit the current setting.

Definition 4.5.5 (Energy)**.** *Given a directed graph $G = (V, \overrightarrow{E})$ and a probability distribution π on V, the energy of an embedding $f : V \to \mathbb{R}^h$ is defined as*

$$\mathcal{E}(f) := \sum_{v\in V} \pi(v) \max_{u:uv\in\overrightarrow{E}} \|f(u) - f(v)\|^2 .$$

The *(π-)mass* is the LHS of the last constraint, which is denoted

$$\|f\|_\pi^2 = \sum_{v\in V} \pi(v) \|f(v)\|^2 .$$

The spreading property is related to the constraint $\sum_{v\in V} \pi(v) f(v) f(v)^T \preceq I_n$, which is called the sub-isotropy condition for the vectors $\left\{\sqrt{\pi(v)} f(v)\right\}_{v\in V}$. In [LOT12], the sub-isotropy condition is used to establish the following *spreading property*, which is used crucially in the spectral partitioning algorithm for k-way edge conductance.

Definition 4.5.6 (Spreading Property [LOT12] (restatement of Definition 3.1.8 with $w = \pi$))**.** *Let π be a probability distribution on V. For two parameters $\delta \in [0, 1]$ and $\eta \in [0, 1]$, an embedding $f : V \to \mathbb{R}^h$ is called (δ, η)-spreading if for every subset $S \subseteq V$,*

$$\operatorname{diam}_{d_f}(S) \le \delta \quad \Longrightarrow \quad \sum_{v\in S} \pi(v) \|f(v)\|^2 \le \eta \cdot \sum_{v\in V} \pi(v) \|f(v)\|^2 ,$$

where $\operatorname{diam}_{d_f}(S) := \max_{u,v\in S} d_f(u, v)$ is the diameter of the set S under the radial projection distance function d_f as defined in Definition 3.1.10.

Any feasible solution f to $\vec{\kappa}(G)$ has $\|f\|_\pi^2 = k$ and satisfies the sub-isotropy condition. Then, by Lemma 3.1.11, f is $(\delta, \frac{1}{k(1-\delta^2)})$-spreading, so we can regard $\eta = \frac{1}{k(1-\delta^2)}$ in the following. The precise parameters and also the definition of the radial projection distance are not important in this subsection.

The main result in this subsection is the following lemma which compares the energy, the mass, and the spreading property of f and of its Gaussian projection $\bar{f}$. Note that the second and the third items of the following lemma are directly from the second and third items in Lemma 3.1.9, and only the first item requires a new proof.

Lemma 4.5.7 (Dimension Reduction)**.** *Let $G = (V, \overrightarrow{E})$ be a directed graph with maximum indegree Δ_{in} and π be a probability distribution on V. Let $f : V \to \mathbb{R}^n$ be an embedding that is (δ, η)-spreading. Let $\bar{f} : V \to \mathbb{R}^h$ be a Gaussian projection of f as defined in Definition 3.1.7. By setting $h \lesssim \frac{1}{\delta^2}\log(\frac{1}{\eta\delta})$, with probability at least $1/4$, the following three properties hold simultaneously:*

$$\mathcal{E}(\bar{f}) \lesssim \left(1 + \frac{\log \Delta_{in}}{h}\right) \cdot \mathcal{E}(f) \quad \text{and} \quad \left\|\bar{f}\right\|_\pi^2 \geq \frac{1}{2}\left\|f\right\|_\pi^2 \quad \text{and} \quad \bar{f} \text{ is } \left(\frac{\delta}{4}, (1+\delta)\eta\right)\text{–spreading.}$$

Proof. The second and the third items are from Lemma 3.1.9. We will prove the first item holds with probability at least $3/4$ by using Proposition 2.10.5, and this would imply the lemma by union bound. Let Γ_i be the i-th row of Γ in Definition 3.1.7. For $uv \in \overrightarrow{E}$,

$$\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 \;=\; \sum_{i=1}^{h}\left(\frac{1}{\sqrt{h}}\Big\langle \Gamma_i, f(u) - f(v)\Big\rangle\right)^2 = \frac{1}{h}\sum_{i=1}^{h} g_i^2,$$

where $g_i = \langle \Gamma_i, f(u) - f(v)\rangle$ is an independent Gaussian random variable with mean zero and variance $\|f(u) - f(v)\|^2$. Applying Proposition 2.10.5 on each $v \in V$ with indegree at most Δ_{in},

$$\mathbb{E}\Big[\max_{u:u\to v}\left\|\bar{f}(u) - \bar{f}(v)\right\|^2\Big] \leq 4\left(1 + \frac{1 + \log \Delta_{in}}{h}\right) \cdot \max_{u:u\to v}\left\|f(u) - f(v)\right\|^2.$$

By linearity of expectation and Markov's inequality,

$$\Pr\left[\sum_{v\in V}\pi(v)\max_{u:u\to v}\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 \leq 16\Big(1 + \frac{1 + \log \Delta_{in}}{h}\Big) \cdot \sum_{v\in V}\pi(v)\max_{u:u\to v}\left\|f(u) - f(v)\right\|^2\right]$$

is at least $3/4$, implying that $\mathcal{E}(\bar{f}) \lesssim (1 + \frac{\log \Delta_{in}}{h}) \cdot \mathcal{E}(f)$ with probability at least $3/4$. □

4.5.3 Spectral Partitioning

The third step is to show that given $\bar{f} : V \to \mathbb{R}^h$ in Lemma 4.5.7, we can construct ℓ disjointly supported functions $\bar{f}_1, \ldots, \bar{f}_\ell : V \to \mathbb{R}^h$ with comparable energy and mass to that of $\bar{f}$.

Lemma 4.5.8 (Spectral Partitioning). *Let $G = (V, E)$ be an undirected graph and π be a probability distribution on V. Let $\overrightarrow{E}$ be an orientation of E and $\bar{f} : V \to \mathbb{R}^h$ be an embedding. Let ℓ be the targeted number of disjointly supported functions where $1 \leq \ell \leq k$. Suppose that there exist $\delta \in (0,1)$ and $\beta \in (0,1)$ such that*

$$\bar{f} \text{ is } \left(\frac{\delta}{4}, \frac{1}{k(1-\delta)}\right)\text{--spreading} \qquad \text{and} \qquad \delta \leq 1 - \frac{2(\ell-1)}{2(1-\beta)k-1}.$$

Then there exist embeddings $\bar{f}_1, \ldots, \bar{f}_\ell : V \to \mathbb{R}^h$ such that the supports of $\{\bar{f}_i\}_{i=1}^{\ell}$ are pairwise disjoint and

$$\mathcal{E}(\bar{f}_i) \lesssim \left(1 + \frac{h}{\delta\beta}\right)^2 \cdot \mathcal{E}(\bar{f}) \qquad \text{and} \qquad \left\|\bar{f}_i\right\|_\pi^2 \gtrsim \frac{1}{k} \left\|\bar{f}\right\|_\pi^2.$$

The proof of this step follows closely the proof in [LOT12], which we have reviewed in Section 3.1.3 and serves well as an overview of our proof.

Given an embedding $\bar{f} : V \to \mathbb{R}^h$ and a target $\ell \leq k$, we would like to find ℓ disjoint subsets $S_1, \ldots, S_\ell$ of V such that

1. for $1 \leq i \leq \ell$, the mass $\mu(S_i) := \sum_{u \in S_i} \pi(u) \left\|\bar{f}(u)\right\|^2$ of each S_i is at least $\frac{1}{2k} \cdot \mu(V)$, where $\mu(V) = \left\|\bar{f}\right\|_\pi^2$, and
2. for $1 \leq i \neq j \leq \ell$, the distance $d_{\bar{f}}(S_i, S_j) := \min_{u \in S_i, v \in S_j} d_{\bar{f}}(u, v)$ between S_i and S_j is at least 2ε for some $\varepsilon > 0$ to be determined later, where $d_{\bar{f}}$ is the radial projection distance in Definition 3.1.10.

To this end, equip V with the pseudo-metric $d_{\bar{f}}$ and consider the metric space $(V, d_{\bar{f}})$. Let $\mathcal{P} = P_1 \sqcup P_2 \sqcup \cdots \sqcup P_T$ be a $\left(\frac{\delta}{4}, \frac{ch}{\beta}, 1-\beta\right)$-padded random partitioning sampled from Theorem 3.1.13, where c is a universal constant and $\delta \in (0,1)$ and $\beta \in (0,1)$ are to be determined. By the assumption that $\bar{f}$ is $(\frac{\delta}{4}, \frac{1}{k(1-\delta)})$-spreading, the first property in Definition 3.1.12 implies that $\mu(P_i) \leq \frac{1}{k(1-\delta)} \cdot \mu(V)$ for $1 \leq i \leq T$. Denote by $B(x, r)$ the open $d_{\bar{f}}$-ball of radius r centered at x, and let $U := \{x \in V \mid B(x, \frac{\delta\beta}{4ch}) \not\subseteq P(x)\}$ be the set of points that are close to the boundaries of $\mathcal{P}$. The second property in Definition 3.1.12 implies that there exists a realization of $\mathcal{P}$ such that $\mu(V \setminus U) \geq (1-\delta)\cdot\mu(V)$. We take such a realization $\mathcal{P} = P_1 \sqcup P_2 \sqcup \cdots \sqcup P_T$ and remove all points in U to obtain $P'_i := P_i \setminus U$ for $1 \leq i \leq T$. By doing so, we end up with disjoint sets $P'_1, P'_2, \ldots, P'_T$ with the following properties:

1. $\mu(P'_i) \leq \frac{1}{k(1-\delta)} \cdot \mu(V)$ for $1 \leq i \leq T$,
2. $\sum_{i=1}^{T} \mu(P'_i) \geq (1-\beta)\cdot\mu(V)$,
3. $d_{\bar{f}}(P'_i, P'_j) \geq \frac{2\delta\beta}{ch}$ for $i \neq j \in [T]$.

Next, we will merge some of the sets $P'_1, \ldots, P'_T$ to form disjoint sets $S_1, \ldots, S_\ell$ so that $\mu(S_i) \geq \frac{1}{2k}\cdot\mu(V)$ for $1 \leq i \leq \ell$. This can be done by a simple greedy process, where we sort the P'_i by nonincreasing mass, and put consecutive sets into a group S_j until $\mu(S_j) \geq \frac{1}{2k}\cdot\mu(V)$. By the first property and the greedy process, each group S_j produced has $\mu(S_j) \leq \frac{1}{k(1-\delta)}\cdot\mu(V)$. Hence, by the second property, the greedy process will succeed to produce at least ℓ groups, each with mass at least $\frac{1}{2k}\cdot\mu(V)$, as long as

$$(1-\beta)\cdot\mu(V) - (\ell-1)\cdot\frac{\mu(V)}{k(1-\delta)} \geq \frac{\mu(V)}{2k} \quad\iff\quad \delta \leq 1 - \frac{2(\ell-1)}{2(1-\beta)k-1},$$

which is exactly the assumption we made in the statement of Lemma 4.5.8 about δ. Therefore, we can produce $S_1, \ldots, S_\ell$ satisfying the two requirements $\mu(S_i) \geq \frac{1}{2k}\cdot\mu(V)$ for $1 \leq i \leq \ell$ and $d_{\bar{f}}(S_i, S_j) \geq 2\varepsilon := \frac{2\delta\beta}{ch}$ for $i \neq j \in [\ell]$, by the third property of $P'_1, \ldots, P'_T$.

Now, we apply the smooth localization procedure in Lemma 3.1.14 on each S_i with $\varepsilon = \frac{\delta\beta}{ch}$ to obtain an embedding $\bar{f}_i : V \to \mathbb{R}^h$ for each $1 \leq i \leq \ell$. First, we check that $\bar{f}_1, \ldots, \bar{f}_\ell$ are disjointly supported. This follows from $d_{\bar{f}}(S_i, S_j) \geq 2\varepsilon$ for $i \neq j$ and the second property in Lemma 3.1.14. Second, since $\mu(S_i) \geq \frac{1}{2k}\cdot\mu(V)$ and $\bar{f}_i(v) = \bar{f}(v)$ for $v \in S_i$ by the first property in Lemma 3.1.14, it follows that $\|\bar{f}_i\|_\pi^2 = \mu(S_i) \geq \frac{1}{2k}\cdot\mu(V)$. Finally, by the third property in Lemma 3.1.14, it follows that

$$\begin{aligned}\mathcal{E}(\bar{f}_i) &= \sum_{v\in V}\pi(v)\max_{u:u\to v}\left\|\bar{f}_i(u)-\bar{f}_i(v)\right\|^2\\ &\leq \Big(1+\frac{2ch}{\delta\beta}\Big)^2\sum_{v\in V}\pi(v)\max_{u:u\to v}\left\|\bar{f}(u)-\bar{f}(v)\right\|^2 \lesssim \Big(1+\frac{h}{\delta\beta}\Big)^2\mathcal{E}(\bar{f}).\end{aligned}$$

Therefore, we conclude that $\bar{f}_1, \ldots, \bar{f}_\ell$ satisfy all the properties stated in Lemma 4.5.8.

4.5.4 Cheeger Rounding

The fourth step is to apply the results in Section 4.3 on $\bar{f}_1, \ldots, \bar{f}_l$ from Lemma 4.5.8 to obtain disjoint subsets with small vertex expansion.

Lemma 4.5.9 (Cheeger Rounding). *Let $G = (V, E)$ be an undirected graph with maximum degree Δ and π be a probability distribution π on V. Given an orientation $\overrightarrow{E}$ and an embedding $\bar{f} : V \to \mathbb{R}^h$, there exists a set $S \subseteq \operatorname{supp}(\bar{f})$ with*

$$\psi(S)^2 \lesssim \min\{h, \log \Delta\} \cdot \frac{\mathcal{E}(\bar{f})}{\mu(\bar{f})}.$$

Proof. Given $\overrightarrow{E}$ and $\bar{f}$, we apply the Gaussian projection step in Proposition 4.3.6 to obtain a 1-dimensional embedding $x : V \to \mathbb{R}$ with $\mathcal{E}(x)/\left\|x\right\|_\pi^2 \lesssim \log \Delta \cdot \mathcal{E}(\bar{f})/\left\|\bar{f}\right\|_\pi^2$. Alternatively, if $h \leq \log \Delta$, we can choose the best coordinate from $\bar{f}$ to obtain a 1-dimensional embedding $x : V \to \mathbb{R}$ with $\mathcal{E}(x)/\left\|x\right\|_\pi^2 \leq h \cdot \mathcal{E}(\bar{f})/\left\|\bar{f}\right\|_\pi^2$ as was done in [OZ22]. So we have a 1-dimensional embedding x with $\mathcal{E}(x)/\left\|x\right\|_\pi^2 \lesssim \min\{h, \log \Delta\} \cdot \mathcal{E}(\bar{f})/\left\|\bar{f}\right\|_\pi^2$.

Then, we apply the Cheeger rounding procedure in Section 4.3.3 on x, but with $c = 0$ instead of the π-weighted median of x in the construction of the ℓ_1 solution h. This ensures that the set S obtained satisfies $S \subseteq \operatorname{supp}(h) \subseteq \operatorname{supp}(x) \subseteq \operatorname{supp}(\bar{f})$ in addition to the vertex cover expansion guarantee $\vec{\psi}_\tau(S)^2 \lesssim \mathcal{E}(x)/\left\|x\right\|_\pi^2 \leq \min\{h, \log \Delta\} \cdot \mathcal{E}(\bar{f})/\left\|\bar{f}\right\|_\pi^2$.[4] Finally, we apply the postprocessing step in Lemma 4.3.11 to obtain a set $S' \subseteq S$ with $\psi(S') \leq 2\vec{\psi}_\tau(S)$ satisfying the requirements of this lemma. □

We are ready to put together the steps to prove the hard direction of the higher-order Cheeger inequality for vertex expansion.

Theorem 4.5.10 (Hard Direction for Multiway Vertex Expansion). *Let $G = (V, E)$ be an undirected graph with maximum degree Δ and π be a probability distribution on V. For any $2 \leq k \leq n$ and $0 \leq \varepsilon < 1$, letting $\xi := \max\{\varepsilon, \frac{1}{2k}\}$, it holds that*

$$\psi_{(1-\varepsilon)k}(G) \lesssim \frac{\log k}{\xi^4} \cdot \sqrt{\log \Delta \cdot \sigma_k^*(G)}.$$

[4] We give up the requirement that 0 be a π-weighted median of h, and so the produced set S may no longer satisfy $\pi(S) \leq 1/2$. The latter is however not a requirement in the current context.

Proof. The first step is to compute an optimal solution (f^*, g^*) to the dual program in Proposition 4.5.2 with objective value $\kappa(G) = \sigma_k^*(G)$. Then, we use Lemma 4.5.4 to obtain a solution $f : V \to \mathbb{R}^n$ to the directed program $\vec{\kappa}(G)$ with energy $\mathcal{E}(f) \leq 2\sigma_k^*(G)$ and $\mu(f) = k$. As f satisfies the sub-isotropy condition $\sum_u \pi(u) f(u) f(u)^T \preceq I_n$ in $\vec{\kappa}(G)$, we know from Lemma 3.1.11 that f is $\big(\delta, 1/\big(k(1-\delta^2)\big)\big)$-spreading for any $\delta \in (0,1)$ of our choice.

The second step is to apply the Gaussian projection algorithm in Lemma 4.5.7 on f with $\eta = 1/\big(k(1-\delta^2)\big)$ to obtain $\bar{f} : V \to \mathbb{R}^h$ with

$$h \asymp \frac{1}{\delta^2} \log\left(\frac{1}{\delta\eta}\right) = \frac{1}{\delta^2} \log\left(\frac{k(1-\delta^2)}{\delta}\right)$$

such that

$$\mathcal{E}(\bar{f}) \lesssim \left(1 + \frac{\log \Delta}{h}\right) \cdot \mathcal{E}(f) \quad \text{and} \quad \mu(\bar{f}) \gtrsim \mu(f) \quad \text{and} \quad \bar{f} \text{ is } \left(\frac{\delta}{4}, \frac{1}{k(1-\delta)}\right)\text{–spreading.}$$

The third step is to apply the spectral partitioning algorithm in Lemma 4.5.8 on $\bar{f}$. Let $\ell := (1-\varepsilon)k$ be the target number of output sets. By setting

$$\delta = \min\left\{\frac{1}{2}, \frac{2(k-\ell)+1}{2(k+\ell)-3}\right\} \quad \text{and} \quad \beta = \frac{2(k-\ell)+1}{4k},$$

we can check that the conditions of Lemma 4.5.8 are satisfied, and so we can construct functions $\bar{f}_1, \ldots, \bar{f}_\ell : V \to \mathbb{R}^h$ with disjoint support, such that for each $1 \leq i \leq \ell$ it holds that

$$\mathcal{E}(\bar{f}_i) \lesssim \left(1 + \frac{h}{\delta\beta}\right)^2 \cdot \mathcal{E}(\bar{f}) \quad \text{and} \quad \left\|\bar{f}_i\right\|_\pi^2 \gtrsim \frac{1}{k} \left\|\bar{f}\right\|_\pi^2.$$

The fourth step is to apply Lemma 4.5.9 to $\bar{f}_1, \ldots \bar{f}_\ell$ to obtain disjoint subsets $S_1, \ldots, S_\ell$, such that for every $1 \le i \le \ell$,

$$\begin{aligned}
\psi(S_i)^2 &\lesssim \min\{h, \log\Delta\} \cdot \frac{\mathcal{E}(\bar{f}_i)}{\|\bar{f}_i\|_\pi^2} \\
&\lesssim \min\{h, \log\Delta\} \cdot k \cdot \left(1 + \frac{h}{\delta\beta}\right)^2 \cdot \frac{\mathcal{E}(\bar{f})}{\|\bar{f}\|_\pi^2} \\
&\lesssim \min\{h, \log\Delta\} \cdot \left(1 + \frac{\log\Delta}{h}\right) \cdot \left(1 + \frac{h}{\delta\beta}\right)^2 \cdot k \cdot \frac{\mathcal{E}(f)}{\|f\|_\pi^2} \\
&\lesssim \log\Delta \cdot \left(1 + \frac{h}{\delta\beta}\right)^2 \cdot \sigma_k^*(G) \\
&\lesssim \log\Delta \cdot \frac{1}{\delta^6\beta^2} \cdot \log^2\left(\frac{k(1-\delta^2)}{\delta}\right) \cdot \sigma_k^*(G),
\end{aligned}$$

where the fourth inequality uses that $\min\{h, \log\Delta\} \cdot \left(1 + \frac{\log\Delta}{h}\right) \le 2\log\Delta$ and $\|f\|_\pi^2 = k$, and the last inequality uses the fact that $\frac{h}{\delta\beta} \ge 1$, so that $1 + h/\delta\beta \lesssim h/\delta\beta$. The above calculation implies that

$$\psi_\ell(G) \lesssim \frac{1}{\delta^3\beta} \log\left(\frac{k}{\delta}\right) \cdot \sqrt{\log\Delta \cdot \sigma_k^*(G)}.$$

Finally, we plug in $\ell = (1-\varepsilon)k$ and consider two cases. In the case when $\varepsilon \ge \frac{1}{2k}$, we see that $\delta = \Theta(\varepsilon)$ and $\beta = \Theta(\varepsilon)$, and so $\psi_\ell(G) \le \frac{1}{\varepsilon^4} \cdot \log k \cdot \sqrt{\log\Delta \cdot \sigma_k^*(G)}$. In the case when $\varepsilon < \frac{1}{2k}$, we simply set $l = k$ and see that $\delta = \Theta(1/k)$ and $\beta = \Theta(1/k)$ and so $\psi_\ell(G) \le k^4 \cdot \log k \cdot \sqrt{\log\Delta \cdot \sigma_k^*(G)}$. Combining the two cases proves the theorem. □

We prove the following easy direction in Section A.3. Note that it is about $\lambda_k^*(G)$ instead of $\sigma_k^*(G)$.

Lemma 4.5.11 (Easy Direction for Multiway Vertex Expansion)**.** *For any undirected graph $G = (V, E)$ and any probability distribution π on V, $\lambda_k^*(G) \le 2\psi_k(G)$ for any $k \ge 2$.*

Combining Lemma 4.5.11 and Theorem 4.5.10, we conclude this section with the following higher-order Cheeger inequality for vertex expansion that implies Theorem 4.1.10.

Theorem 4.5.12 (Higher-Order Cheeger Inequality for Vertex Expansion)**.** *For any undirected graph $G = (V, E)$ with maximum degree Δ and any probability distribution π on V,*

$$\frac{\sigma_k^*(G)}{k} \le \lambda_k^*(G) \lesssim \psi_k(G) \lesssim k^4 \log k \sqrt{\log\Delta \cdot \sigma_k^*(G)} \le k^{\frac{9}{2}} \log k \sqrt{\log\Delta \cdot \lambda_k^*(G)}$$

Furthermore, for any $1 > \varepsilon \geq \frac{1}{2k}$,

$$\psi_{(1-\varepsilon)k}(G) \lesssim \frac{1}{\varepsilon^4} \log k \sqrt{\log \Delta \cdot \sigma_k^*(G)} \leq \frac{1}{\varepsilon^4} \log k \sqrt{k \log \Delta \cdot \lambda_k^*(G)}.$$

4.6 Improved Cheeger Inequality for Vertex Expansion

The goal of this section is to prove Theorem 4.1.11. The proof in [KLL+13] for k-way edge conductance has two main steps. The first step is to prove that if the second eigenvector is close to a k-step function, then the approximation guarantee of threshold rounding is improved. The second step is to prove that if λ_k is large for a small k, then the second eigenfunction is close to a k-step function. Refer to Section 3.1.4 for details.

We follow the a similar plan to prove an improved version of the Cheeger inequality for $\gamma^{(1)}(G)$ in Theorem 4.3.7, where in the second step we replace $\lambda_k(G)$ by $\sigma_k^*(G)$.

The following is the precise statement of the first step for $\gamma^{(1)}(G)$, which informally says that if there is a good k-step approximation (see Definition 3.1.17) to an optimal solution to $\gamma^{(1)}(G)$, then the performance of threshold rounding is better than that in Theorem 4.3.7.

Proposition 4.6.1 (Rounding k-Step Approximation)**.** *Let* $G = (V, E)$ *be an undirected graph and* π *be a probability distribution on* V*. For any feasible solution* (f, g) *to the* $\gamma^{(1)}(G)$ *program with objective value* γ_f *and any* k*-step function* $y_f : V \to \mathbb{R}$ *approximating* f*,*

$$\psi(G) \lesssim k \cdot \gamma_f + k \left\| f - y_f \right\|_\pi \sqrt{\gamma_f}.$$

Our second step is to prove that if $\sigma_k^*(G)$ in Proposition 4.5.2 is large for a small k, then there is a good k-step approximation to a good solution to $\gamma^{(1)}(G)$.

Proposition 4.6.2 (Constructing k-Step Approximation)**.** *Let* $G = (V, E)$ *be an undirected graph and* π *be a probability distribution on* V*. For any feasible solution* (f, g) *to the* $\gamma^{(1)}(G)$ *program with objective value* γ_f*, there exists a* k*-step function* y *with*

$$\| f - y \|_\pi^2 \lesssim \frac{k \cdot \gamma_f}{\sigma_k^*(G)}.$$

Assuming Proposition 4.6.1 and Proposition 4.6.2, we prove an exact analog of the improved Cheeger's inequality in [KLL+13] for vertex expansion, with $\sigma_k^*(G)/k$ playing the role of $\lambda_k^*(G)$.

Theorem 4.6.3 (Improved Cheeger Inequality for Vertex Expansion)**.** *For any undirected graph $G=(V,E)$ and any probability distribution π on V and any $k \geq 2$,*

$$\gamma^{(1)}(G) \lesssim \psi(G) \lesssim k \cdot \gamma^{(1)}(G) \cdot \sqrt{\frac{k}{\sigma_k^*(G)}}.$$

Proof. The easy direction is proved in Lemma 4.3.8. For the hard direction, let (f^*, g^*) be an optimal solution to the $\gamma^{(1)}(G)$ program in Definition 3.2.5 with objective value γ^*, and σ^* be the optimal value of the $\sigma_k^*(G)$ program in Proposition 4.5.2. By Proposition 4.6.2, there exists a k-step function $y : V \to \mathbb{R}$ with

$$\|f^* - y\|_\pi^2 \lesssim \frac{k \cdot \gamma^*}{\sigma^*}.$$

Applying Proposition 4.6.1 with y, it follows that

$$\psi(G) \lesssim k \cdot \left(\gamma^* + \|f^* - y\|_\pi \cdot \sqrt{\gamma^*}\right) \lesssim k \cdot \gamma^* \left(1 + \sqrt{\frac{k}{\sigma^*}}\right) \lesssim k \cdot \gamma^* \cdot \sqrt{\frac{k}{\sigma^*}},$$

and the proof is complete. □

We can now prove Theorem 4.1.11.

Proof of Theorem 4.1.11. Note that $\sigma_{2k}^*(G) \geq k \cdot \lambda_k^*(G)$[5] by plugging the reweighting P that maximizes $\lambda_k(I-P)$ in Definition 4.1.9 into Definition 4.5.1, so that

$$\sigma_{2k}^*(G) \geq \sum_{j=1}^{2k} \lambda_j(I-P) \geq \sum_{j=k}^{2k} \lambda_j(I-P) \geq k\lambda_k(I-P).$$

Therefore,

$$\lambda_2^*(G) = \gamma(G) \leq \gamma^{(1)}(G) \lesssim \psi(G) \lesssim 2k \cdot \gamma^{(1)}(G) \cdot \sqrt{\frac{2k}{\sigma_{2k}^*(G)}} \lesssim \frac{k \cdot \lambda_2^*(G) \cdot \log \Delta}{\sqrt{\lambda_k^*(G)}},$$

where we use $\gamma(G) = \lambda_2^*(G)$ in Proposition 3.2.3 and $\gamma(G) \leq \gamma^{(1)}(G) \lesssim \log \Delta \cdot \gamma(G)$ in Proposition 4.3.6. □

[5]In the proof in [KLT22], we used the relation $\sigma_k^*(G) \geq \lambda_k^*(G)$, resulting in the factor $\sqrt{k}$ loss as mentioned in the footnote to Theorem 4.1.11.

Remark 4.6.4 (Tight Examples)**.** *We remark that Theorem 4.6.3 is tight. The loss in Theorem 4.1.11 is because of the factor* $\log\Delta$ *loss in the dimension reduction step.*

As an example, let G *be an* n*-cycle, where* n *is odd, and* π *the uniform distribution. Suppose* $k \ll n$*. Then* $\sigma_k^*(G) = \sigma_k(G)$*, because the only possible "reweighting" is the one with equal edge weight. Since* $\psi(G) = \Theta(1/n), \gamma^{(1)}(G) = O(1/n^2)$ *(choose* $f : V \to \mathbb{R}$ *that maps vertex* $l \in [n]$ *to point* $C(1 - \frac{4}{n} \cdot \min(l, n-l))$*, where* $C = \Theta(1)$ *is a normalizing factor, and* $g(l) \equiv \frac{8C^2}{n^2}$*), and*

$$\sigma_k(G) = \sum_{l=0}^{k-1} \left(1 - \cos\left(\frac{2\pi l}{n}\right)\right) = \Theta\left(\frac{k^3}{n^2}\right),$$

one can verify that in this case the hard direction of Theorem 4.6.3 is tight.

We prove Proposition 4.6.1 and Proposition 4.6.2 in the following two subsections.

4.6.1 Rounding k-Step Approximation

We prove Proposition 4.6.1 in this subsection. We follow the proof of the hard direction of Theorem 4.3.7, with the key modification being how we define the ℓ_1 solution $h : V \to \mathbb{R}$.

First, we do some preprocessing on the solution (f, g) as in Section 4.3. Given a feasible solution (f, g) to the $\gamma^{(1)}(G)$ program with objective value γ_f, we use Lemma 4.3.4 and Lemma 4.3.5 to obtain a solution $(f, \overrightarrow{E})$ to the directed program $\vec{\gamma}^{(1)}(G)$ in Definition 4.3.3 with objective value at most $2\gamma_f$. With Proposition 4.3.10 and Lemma 4.3.11 in mind, it suffices then to construct an $h : V \to \mathbb{R}$ such that 0 is a π-weighted median of h, and

$$\frac{\sum_{v \in V} \pi(v) \max_{u: uv \in \overrightarrow{E}} |h(u) - h(v)|}{\sum_{v \in V} \pi(v)|h(v)|} =: \eta_h \lesssim k \cdot \gamma_f + k \left\|f - y_f\right\|_\pi \sqrt{\gamma_f}.$$

We will follow closely the proof of Proposition 3.1.18. Let $y_f : V \to \mathbb{R}$ be a k-step function, taking values $t_1 < t_2 < \cdots < t_k$. Choose $c \in \mathbb{R}$ to be a π-weighted median of f, and define $h : V \to \mathbb{R}$ so that

$$h(u) := \int_c^{f(u)} \nu(t)\, dt,$$

where $\nu(t) := \min_{i \in [k]} |t - t_i|$. Note that the integral is negative when $f(u) < c$ and positive when $f(u) > c$, and so 0 is a π-weighted median of h.

Now, we lower bound the denominator and upper bound the numerator. In the denominator, the same calculations give

$$\sum_{v\in V}\pi(v)|h(v)| \gtrsim \frac{1}{k}\sum_{v\in V}\pi(v)(f(v)-c)^2 \geq \frac{1}{k}.$$

In the numerator, we have

$$\begin{aligned}
& \sum_{v\in V}\pi(v) \max_{u:uv\in\overrightarrow{E}} |h(u)-h(v)| \\
= \;& \sum_{v\in V}\pi(v) \max_{u:uv\in\overrightarrow{E}} \left|\int_{f(v)}^{f(u)} \nu(t)\,dt\right| \\
\leq \;& \sum_{v\in V}\pi(v) \max_{u:uv\in\overrightarrow{E}} \left[\frac{1}{2}|f(u)-f(v)|(2\nu(f(v))+|f(u)-f(v)|)\right] \quad (\because \nu(t) \text{ is 1-Lipschitz}) \\
\overset{(*)}{\lesssim} \;& \gamma_f + \sqrt{\gamma_f \cdot \sum_{v\in V}\pi(v)\nu(f(v))^2} \leq \gamma_f + \|f-y_f\|_\pi \sqrt{\gamma_f}.
\end{aligned}$$

Again, the step $(*)$ is where Cauchy-Schwarz inequality is applied. Combining the two bounds yields the desired result.

4.6.2 Constructing k-Step Approximation

We prove Proposition 4.6.2 in this subsection. The high level plan is similar to that in [KLL+13], and the proof follows closely that of Proposition 3.1.19. Given a feasible solution (f, g) to the $\gamma^{(1)}(G)$ program, we aim to construct a good k-step approximation y of f using a simple procedure. If we fail to do so, then we show that f can be used to construct a good k-dimensional solution $\bar{f} = (\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_k)$ to the $\sigma_k^*(G)$ program, contradicting that the value of $\sigma_k^*(G)$ is large. Therefore, the simple process must succeed to find a good k-step approximation y_f of f.

Let $M > 0$ be a parameter to be determined later. Let $t_0 = -\infty$ and successively choose $t_1, t_2, \ldots$ so that $t_i > t_{i-1}$ is the smallest real number such that the following function

$$\bar{f}_i(u) := \begin{cases} \min\left(f(u)-t_{i-1}, t_i - f(u)\right), & \text{if } t_{i-1} < f(u) \leq t_i \\ 0, & \text{otherwise} \end{cases}$$

satisfies $\left\|\bar{f}_i\right\|_\pi^2 \geq M$. The role of the functions $\bar{f}_i$ is to measure how well the two threshold values t_{i-1} and t_i approximate the values of the function at between them. If such a t_i

does not exist, we set $t_i = \infty$ and terminate the process. The process always terminates within n steps, and if it terminates with $t_{k+1} = \infty$ then the following function (which is determined once f and the t_i's are fixed)

$$y_f(u) := \arg \min_{t_i : i \in [k]} |f(u) - t_i|$$

is a k-step function. Observe also that the h_i's have disjoint support, and in fact

$$\sum_{i=1}^{k+1} \left\| \bar{f}_i \right\|_\pi^2 = \| f - y_f \|_\pi^2 .$$

Consider the scenario that the process does *not* terminate after k steps. That means $\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_k$ are all well-defined and each having mass $\left\| \bar{f}_i \right\|_\pi^2$ *exactly* M. We will construct from $\bar{f}_1, \ldots, \bar{f}_k$ a solution $(\bar{f}, \bar{g})$ to the $\sigma_k^*(G)$ program with small objective value. Define $\bar{f} : V \to \mathbb{R}^n$ and $\bar{g} : V \to \mathbb{R}$ as follows:

$$\bar{f}(v) := \Big(\frac{\bar{f}_1(v)}{\sqrt{M}}, \ldots, \frac{\bar{f}_k(v)}{\sqrt{M}}, 0, \ldots, 0 \Big)^T \quad \text{and} \quad \bar{g}(v) := \frac{1}{M} g(v).$$

We will check that $(\bar{f}, \bar{g})$ is a feasible solution to the $\sigma_k^*(G)$ program defined in Proposition 4.5.2. Define $S_i := \operatorname{supp} \bar{f}_i \subseteq V$. For the sub-isotropy condition, note that each $\bar{f}(v)$ has at most one nonzero entry, and

$$\begin{aligned}
& \sum_{v \in V} \pi(v) \bar{f}(v) \bar{f}(v)^T \\
= \ & \operatorname{diag} \Big(\frac{1}{M} \sum_{v \in S_1} \pi(u) \bar{f}_1(u)^2, \frac{1}{M} \sum_{v \in S_2} \pi(v) \bar{f}_2(u)^2, \ldots, \frac{1}{M} \sum_{v \in S_k} \pi(v) \bar{f}_k(v)^2, 0, \ldots, 0 \Big) \\
= \ & \frac{1}{M} \operatorname{diag} \Big(\left\| \bar{f}_1 \right\|_\pi^2, \left\| \bar{f}_2 \right\|_\pi^2, \ldots, \left\| \bar{f}_k \right\|_\pi^2, 0, \ldots, 0 \Big) \\
= \ & \operatorname{diag}(1, 1, \ldots, 1, 0, \ldots, 0) \ \preceq \ I_n.
\end{aligned}$$

The mass constraint is satisfied as

$$\sum_{v \in V} \pi(v) \left\| \bar{f}(v) \right\|^2 = \operatorname{tr} \left(\sum_{u \in V} \pi(u) \bar{f}(u) \bar{f}(u)^T \right) = \operatorname{tr} \Big(\operatorname{diag} \big(1, 1, \ldots, 1, 0, \ldots, 0 \big) \Big) = k.$$

For the constraint on each edge $uv \in E$,

$$\left\| \bar{f}(u) - \bar{f}(v) \right\|^2 = \frac{1}{M} \sum_{i=1}^{k} \big(\bar{f}_i(u) - \bar{f}_i(v) \big)^2 \le \frac{1}{M} \big(f(u) - f(v) \big)^2 \le \frac{1}{M} \big(g(u) + g(v) \big) = \bar{g}(u) + \bar{g}(v),$$

where for the first inequality we consider two cases: (i) suppose $u \in S_i$ and $v \in S_j$ for $i = j$, then $\sum_{l=1}^{k} \left(\bar{f}_l(u) - \bar{f}_l(v)\right)^2 = \left(\bar{f}_i(u) - \bar{f}_i(v)\right)^2 \le \left(f(u) - f(v)\right)^2$, and (ii) suppose $u \in S_i$ and $v \in S_j$ for $i \neq j$, then $\sum_{l=1}^{k} \left(\bar{f}_l(u) - \bar{f}_l(v)\right)^2 = \left(\bar{f}_i(u) - \bar{f}_i(v)\right)^2 + \left(\bar{f}_j(u) - \bar{f}_j(v)\right)^2 \le \left(f(u) - f(v)\right)^2$ since $\left|\bar{f}_i(u) - \bar{f}_i(v)\right| + \left|\bar{f}_j(u) - \bar{f}_j(v)\right| \le \left|f(u) - f(v)\right|$.

Therefore, $(\bar{f}, \bar{g})$ is a feasible solution to the $\sigma_k^*(G)$ program, and its objective value is

$$\sum_{v \in V} \pi(v)\bar{g}(v) = \frac{1}{M} \sum_{v \in V} \pi(v) g(v) = \frac{1}{M} \cdot \gamma_f \ge \sigma_k^*(G).$$

Choose $M = 2\gamma_f / \sigma_k^*(G)$ so that the above inequality *fails*. This means that the process terminates after at most k steps, and with $t_k = \infty$, which gives

$$\|f - y_f\|_\pi^2 = \sum_{i=1}^{k} \left\|\bar{f}_i\right\|_\pi^2 \le kM \lesssim \frac{k\gamma_f}{\sigma_k^*(G)}.$$

This completes the proof of Proposition 4.6.2.

4.7 Vertex Expansion of 0/1-Polytopes

(In this section, we revert to using $|V|$ to denote the number of vertices of a graph and reserve n for other uses.)

The goal of this section is to present a construction of 0/1-polytopes with poor vertex expansion as described in Theorem 4.1.12, which has implications about sampling from the uniform distribution as described in Section 4.1.5.

A 0/1-polytope is defined by a subset of vertices in the boolean hypercube $\{0,1\}^n$. Our examples are based on the following simple probabilistic construction.

Definition 4.7.1 (Probabilistic Construction)**.** *Let n be an even number and $k < n/2$. For a binary string $x \in \{0,1\}^n$, denote its 1-norm by $|x| := \sum_{i=1}^{n} |x_i|$. The set of vertices of our constructed polytope is the union of three subsets:*

1. *A left part $L := \{x \in \{0,1\}^n \mid |x| = k\}$ that consists of all binary strings with k ones.*
2. *A right part $R := \{x \in \{0,1\}^n \mid |x| = n - k\}$ that consists of all binary strings with $n - k$ ones.*

3. *A middle part* $M \subset \{x \in \{0,1\}^n \mid |x| = n/2\}$ *that consists of* $\Theta(4^k n^2)$ *uniformly random binary strings with* $n/2$ *ones.*

The graph $G_Q = (V, E)$ of a polytope Q is defined as the 1-skeleton of the polytope Q. Our plan is to prove that for a random polytope Q constructed in Definition 4.7.1, the middle part M "blocks" all the edges between L and R in G_Q with constant probability. This would imply that $\partial L \subseteq M$ and $\partial R \subseteq M$, and thus $\psi(L), \psi(R) \lesssim 4^k n^2/n^k$ are very small.

The organization of this section is as follows. First, in Section 4.7.1, we provide a sufficient condition for two binary strings x, y to have no edge in G_Q, using geometric arguments. Then, in Section 4.7.2, we outline the main probabilistic argument to prove Theorem 4.1.12, by using a union bound over the set of linear threshold functions. Finally, we show Corollary 4.1.13 in Section 4.7.3. We defer all the proofs in Section 4.7.1 and Section 4.7.2 to Section A.4.

As mentioned in Remark 4.1.14 in the introduction, Gillmann [Gil07] has constructed similar examples of 0/1-polytopes with poor vertex expansion. Following the same simple argument as in Section 4.7.3, one obtains analogous lower bounds on the fastest mixing time of these polytopes. We remark that the construction in [Gil07] is similar to our construction, but the proofs are different and so we present our proofs even though the results follow from [Gil07].

4.7.1 A Sufficient Condition for Edge Blocking

Let Q be a 0/1-polytope and $G_Q = (V, E)$ be its graph/1-skeleton. For two binary strings $x, y \in \{0,1\}^n$, if xy is an edge in G_Q, then there is a separating hyperplane l with $l(x), l(y) \geq 0$ while $l(z) < 0$ for all other binary strings z in the 0/1-polytope Q.

In the construction of Q in Definition 4.7.1, if $x \in L$ and $y \in R$ then $\frac{1}{2}(x+y)$ has 1-norm equal to $n/2$. If xy is an edge in G_Q, then there is a separating hyperplane l with $l\big(\frac{1}{2}(x+y)\big) \geq 0$ while $l(z) < 0$ for all other binary strings z in the middle part M. So, if we could establish that $\frac{1}{2}(x+y)$ is in the convex hull $\text{conv}(M)$ of M for all $x \in L$ and $y \in R$, then there are no edges between L and R in the graph G_Q. This is the sufficient condition that we will formalize.

In the analysis, we use the following definitions to group the pairs of vertices $x \in L$, $y \in R$ based on their common patterns.

Definition 4.7.2 (Patterns)**.** *For $n \in \mathbb{N}$, a pattern is an element $p \in \{0, 1, \vee\}^n$, where 0, 1, and $\vee$ are regarded as symbols.*

The support of a pattern $p \in \{0, 1, \vee\}^n$ is defined as $\text{supp}(p) := \{i \in [n] \mid p_i \neq \vee\}$*. We also define* $\text{supp}_0(p) := \{i \in [n] \mid p_i = 0\}$ *and* $\text{supp}_1(p) := \{i \in [n] \mid p_i = 1\}$*.*

Given two binary strings $x, y \in \{0, 1\}^n$, their common pattern $p^{(x,y)} \in \{0, 1, \vee\}^n$ is defined as

$$p_i^{(x,y)} = \begin{cases} 0, & \text{if } x_i = y_i = 0 \\ 1, & \text{if } x_i = y_i = 1 \\ \vee, & \text{if } x_i \neq y_i. \end{cases}$$

Given a pattern $p \in \{0, 1, \vee\}^n$ and a binary string $x \in \{0, 1\}^n$, x is said to match p if and only if $p_i \neq \vee$ implies $p_i = x_i$.

For each pattern p, we consider a potential separating hyperplane of the following specific form, which will be convenient for the probability analysis.

Definition 4.7.3 (Consistent Affine Function)**.** *Let $p \in \{0, 1, \vee\}^n$ be a pattern. An affine function $l : (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n \mapsto \beta + \sum_i \alpha_i u_i$ is called p-consistent if*

$$\alpha_i = 0 \text{ for } i \in \text{supp}(p) \qquad \text{and} \qquad \beta + \frac{1}{2} \sum_{i : i \notin \text{supp}(p)} \alpha_i = 0.$$

If p is the common pattern of x and y, then l being p-consistent implies $l\big(\frac{1}{2}(x + y)\big) = 0$.

We formulate the sufficient condition described above for the middle part M blocking the edge xy for $x \in L$ and $y \in R$ using the definitions that we have developed. The proof is deferred to Section A.4.

Lemma 4.7.4 (Blocking One Edge)**.** *Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 4.7.1. Let $x \in L$, $y \in R$ and $p = p^{(x,y)} \in \{0, 1, \vee\}^n$ be the common pattern of x and y. If for any p-consistent affine function l there exists a point $z \in M$ matching the pattern p and satisfying $l(z) \geq 0$, then there is no edge connecting x and y in the graph of Q.*

The following is a sufficient condition for the middle part M blocking all the edges between L and R, by considering all possible common patterns of an $x \in L$ and a $y \in R$.

Lemma 4.7.5 (Blocking All Edges)**.** *Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 4.7.1. Suppose for every pattern $p \in \{0, 1, \vee\}^n$ with $|\text{supp}_0(p)| = |\text{supp}_1(p)| \leq k$ and for any p-consistent affine function $l : \mathbb{R}^n \to \mathbb{R}$, there is $z \in M$ matching the pattern p with $l(z) \geq 0$. Then there are no edges between L and R in the graph of Q.*

4.7.2 Probabilistic Analysis

Our plan is to use the sufficient condition in Lemma 4.7.5 to prove that a random M with not many points can block all the edges between L and R. To this end, we prepare with two simple lemmas about the probability of $z \in M$ satisfying $l(z) \geq 0$ and matching a particular pattern p with $|\operatorname{supp}_0(p)| = |\operatorname{supp}_1(p)| \leq k$.

The geometric intuition of the first lemma is simple: when we restrict a p-consistent affine function l on the coordinates in $[n] \setminus \operatorname{supp}(p)$, then l is an "unbiased" hyperplane that goes through the point $\frac{1}{2} \cdot \vec{1}$ on $[n] \setminus \operatorname{supp}(p)$ (because of the second condition in Definition 4.7.3), and thus a random vertex in M matching the pattern p lies on the non-negative side of l with probability at least $1/2$.

Lemma 4.7.6. *Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 4.7.1. Let p be the common pattern of $x \in L$ and $y \in R$, and l be a p-consistent affine function. Let Z be the uniform distribution on $\{z \in \{0,1\}^n : |z| = \frac{n}{2}\}$. Then,*

$$\Pr_{z \sim Z}\left[\, l(z) \geq 0 \mid z \text{ matches pattern } p \,\right] \geq \frac{1}{2}.$$

The second lemma gives a lower bound on the probability that a random point $z \sim Z$ matches a pattern p with $|\operatorname{supp}_0(p)| = |\operatorname{supp}_1(p)| \leq k$.

Lemma 4.7.7. *Let $p \in \{0, 1, \vee\}^n$ be a pattern with $|supp_0(p)| = |supp_1(p)| = s \leq k$, and let Z be the uniform distribution on $\{z \in \{0,1\}^n : |z| = n/2\}$. Then*

$$\Pr_{z \sim Z}[z \text{ matches pattern } p] \gtrsim 4^{-s}.$$

With the above two lemmas, we can show that for any pattern p with $|\operatorname{supp}_0(p)| = |\operatorname{supp}_1(p)| = s \leq k$ and any p-consistent affine function l, the probability that a random point in $\{z \in \{0,1\}^n : |z| = n/2\}$ matches the pattern p and satisfies $l(z) \geq 0$ with probability not too small. So, by adding enough number of random points in the middle part M, such a point z exists in M with high probability for a fixed p and l. Then, we would like to use a union bound over p and l to prove that there will be no edges between L and R in the graph of the polytope with constant probability.

One technical issue of this approach is that there are infinitely many affine functions $l : \mathbb{R}^n \to \mathbb{R}$. Note, however, that we only care about the values of l on the hypercube vertices. This reduces the number of different functions to 2^{2^n}. Indeed, we only care about whether $l(z) \geq 0$ for $z \in \{0,1\}^n$ for Lemma 4.7.5. Therefore, we only need to apply a union bound over the set of linear threshold functions over the boolean hypercube, which further reduces the number of different such functions to 2^{n^2}.

Proposition 4.7.8 ([Cov65]). *The number of linear threshold functions on $\{0,1\}^n$ is at most 2^{n^2}. A linear threshold function on $\{0,1\}^n$ is a function of the form $\tau : \{0,1\}^n \to \{0,1\}$, where*

$$\tau(u_1,\ldots,u_n) = \begin{cases} 1, & \text{if } \beta + \sum_i \alpha_i u_i \geq 0; \\ 0, & \text{if } \beta + \sum_i \alpha_i u_i < 0, \end{cases}$$

for some $\alpha_1,\ldots,\alpha_n,\beta \in \mathbb{R}$.

The complete proof of Theorem 4.1.12 can be found in Section A.4.

4.7.3 Mixing Time

Corollary 4.1.13 follows from Theorem 4.1.12 and the easy direction of Cheeger's inequality for vertex expansion in Theorem 3.2.4.

Proof of Corollary 4.1.13. Let $Q = L \cup M \cup R$ be a 0/1-polytope from Definition 4.7.1. The number of vertices of Q is $|L| + |M| + |R| \lesssim \binom{n}{k} \leq (en/k)^k$. By Theorem 4.1.12, the vertex expansion of the graph G_Q is $\psi(G_Q) \lesssim (4k)^k/n^{k-2}$. Therefore, by the easy direction in Theorem 3.2.4, the mixing time of any reversible chain $P \in \mathbb{R}^{V\times V}$ on G_Q with stationary distribution $\pi = \frac{1}{|V|}\vec{1}$ is at least

$$\tau^*_{\text{mix}}(G) \gtrsim \frac{1}{\psi(G_Q)} \gtrsim \frac{n^{k-2}}{(4k)^k} \gtrsim \left(\frac{en}{k}\right)^{k-2} \gtrsim |V|^{1-\frac{2}{k}},$$

where the second last inequality is by the assumption that k is a constant and so only the exponent of n matters, and the last inequality is by $|V| \leq (en/k)^k$ explained above. □

The implication of Corollary 4.1.13 has been discussed in Section 4.1.5 and we won't repeat here.

4.8 Tight Example to Cheeger Inequality for Vertex Expansion

The goal of this section is to construct a family of tight examples to Theorem 4.1.3 when $\pi = \vec{1}/|V|$. The graphs constructed will have non-constant maximum degree Δ and satisfy

$$\frac{\psi(G)^2}{\log \Delta} \asymp \lambda_2^*(G).$$

These examples are suggested to us by Shayan Oveis Gharan.

The graphs are realized as proximity graphs on $\mathbb{S}^{k-1}$. We employ the following notations: μ denotes the normalized Lebesgue measure on $\mathbb{S}^{k-1}$. $d(\cdot,\cdot)$ denotes the geodesic distance on $\mathbb{S}^{k-1}$. For $\theta \in \mathbb{S}^{k-1}$ and $r > 0$, $B(\theta, r)$ denotes the set of points on $\mathbb{S}^{k-1}$ of geodesic distance less than r from θ. $\mathrm{Cap}(r)$ denotes a generic spherical cap on $\mathbb{S}^{k-1}$ of radius r. For $S \subseteq \mathbb{S}^{k-1}$, we use S_r or $S + \mathrm{Cap}(r)$ to denote the set of points of distance less than r from S.

Definition 4.8.1 (Spherical Proximity Graph)**.** *Given k, a (γ, δ)-spherical proximity graph in k dimensions is a graph $G_k = (V, E)$ that can be constructed as follows:*

1. *Partition $\mathbb{S}^{k-1}$ into n cells: $S_1, S_2, \ldots, S_n$, such that each cell has diameter at most γ and measure between $\varepsilon := \mu(\mathrm{Cap}(\gamma/4))$ and 2ε. Take a point $x_u \in S_u$ for each $u \in [n]$.*
2. *Set $V = [n]$ and $E = \{(u, v) \in V \times V : d(x_u, x_v) < \delta\}$.*

We shall show that, for suitable choices of γ and δ, the resulting graph will be a vertex expander with $\lambda_2^* = O(1/\log\Delta)$.

Theorem 4.8.2 (Tight Example of Theorem 4.1.3)**.** *There exist constants $0 < c_1 < c_2$ such that, for any $(c_1/\sqrt{k}, c_2/\sqrt{k})$-spherical proximity graph G_k, the following hold:*

- *the maximum degree Δ of G_k is $2^{O(k)}$;*
- *$\psi(G_k) = \Theta(1)$; and*
- *$\lambda_2^*(G_k) = O(1/\log\Delta)$.*

The rest of the section is organized as follows. First, we show that the construction of the spherical proximity graph is indeed possible. Then, we prove the degree and expansion bounds, by relating them to volume ratios on the sphere. Finally, we prove the bound on reweighted eigenvalue, using the spherical embedding of the graph. All the proofs in this section are deferred to Section A.5.

4.8.1 The Graph Construction

The following proposition is modified from [GM12, Lemma 8.3.22] and shows that we can always construct a spherical proximity graph.

Proposition 4.8.3 (Constructing Spherical Proximity Graph)**.** *For every k and $\gamma > 0$, there exists n such that $\mathbb{S}^{k-1}$ can be partitioned into cells $S_1, \ldots, S_n$ with $\operatorname{diam}(S_i) \leq \gamma$ and $\mu(S_i) \in [\varepsilon, 2\varepsilon]$ for all $i \in [n]$, where $\varepsilon := \mu(\operatorname{Cap}(\gamma/4))$. Therefore, $n = \Theta(\mu(\operatorname{Cap}(\gamma/4))^{-1})$.*

4.8.2 Degree and Expansion Bounds

Now, we establish bounds on the maximum degree and vertex expansion of G_k. We do so by connecting the quantities with the continuous notion of volume.

Recall our choice of parameters $\gamma = c_1/\sqrt{k}$ and $\delta = c_2/\sqrt{k}$, where $0 < c_1 < c_2$ are constants.

Lemma 4.8.4 (Degree, Expansion, and Volume)**.** *Let $G_k = (V, E)$ be the (γ, δ)-spherical proximity graph constructed above.*

- *The maximum degree of G_k is at most*
$$\frac{\mu(\operatorname{Cap}(\delta + \gamma))}{\mu(\operatorname{Cap}(\gamma/4))}.$$
- *For any $T \subseteq V$ with $|T| \leq |V|/2$, if we let $S_T := \cup_{i \in T} S_i$, then*
$$\psi(T) \geq \frac{\mu(S_T + \operatorname{Cap}(\delta - \gamma)) - \mu(S_T)}{2\mu(S_T)}.$$

After relating the graph degree and vertex expansion to volumes of portions of sphere, we shall now use results from high-dimensional geometry to obtain bounds on the maximum degree and vertex expansion.

Proposition 4.8.5 (Degree Bound)**.** *For constants $0 < c_1 < c_2$, it holds that*
$$\frac{\mu(\operatorname{Cap}(\delta + \gamma))}{\mu(\operatorname{Cap}(\gamma/4))} \leq 2^{O(k)}.$$

Proposition 4.8.6 (Small-Volume Expansion of the Sphere)**.** *For constants $0 < c_1 < c_2$ with $c_2 - c_1$ sufficiently large, it holds that for all $T \subseteq V$ with $|T| \leq |V|/2$,*
$$\frac{\mu(S_T + \operatorname{Cap}(\delta - \gamma)) - \mu(S_T)}{\mu(S_T)} \geq \Omega(1).$$

Combining Lemma 4.8.4, Proposition 4.8.5, and Proposition 4.8.6 gives the first two parts of Theorem 4.8.2.

4.8.3 Reweighted eigenvalue bound

Finally, we bound $\lambda_2^*(G)$. On a high level, if a graph can be embedded in $\mathbb{S}^2$ via $u \mapsto f(u)$, such that the centre of mass is the origin and that each edge uv satisfies $\|f(u) - f(v)\| \leq \delta$, then the embedding certifies that $\lambda_2^*(G) \leq \delta^2$. This is basically the proof. The details can be found in Section A.5.

Proposition 4.8.7 (Reweighted Eigenvalue Bound)**.** *Let G_k be the graph from Theorem 4.8.2. For constants $0 < c_1 < c_2$, it holds that*

$$\lambda_2^*(G_k) \lesssim \frac{1}{k} \lesssim \frac{1}{\log \Delta}.$$

Proposition 4.8.7 gives the last part of Theorem 4.8.2. Thus, the proof of Theorem 4.8.2 is complete.

4.9 Concluding Remarks

We present a new spectral theory which relates (i) reweighted eigenvalues, (ii) vertex expansion and (iii) fastest mixing time. This is analogous to the classical spectral theory which relates (i) eigenvalues, (ii) edge conductance and (iii) mixing time. This spectral approach for vertex expansion via reweighted eigenvalues has the advantage that most existing results and proofs for edge conductances and eigenvalues have a close analog for vertex expansion and reweighted eigenvalues with almost tight bounds. The results presented here are not exhaustive, and we fully expect that other results relating eigenvalues and edge conductances also have an analog for vertex expansion using reweighted eigenvalues.

To conclude, we believe that our work provides an interesting spectral theory for vertex expansion, as the formulations have the natural interpretation as reweighted eigenvalues and also have close connections to other important problems such as fastest mixing time and the reweighting conjectures in approximation algorithms.

Acknowledgements

We thank Shayan Oveis Gharan for suggesting the tight example for Theorem 4.1.3 in Section 4.8, Robert Wang for suggesting the connection to the 0/1-polytope expansion conjecture, and Sam Olesker-Taylor for providing insightful comments regarding the presentation.

Chapter 5

Cheeger Inequalities for Directed Graphs and Hypergraphs using Reweighted Eigenvalues

In this chapter, we derive Cheeger inequalities for directed graphs and hypergraphs using the reweighted eigenvalue approach. The goal is to develop a new spectral theory for directed graphs and an alternative spectral theory for hypergraphs.

The first main result is a Cheeger inequality relating the vertex expansion $\vec{\psi}(G)$ of a directed graph G to the vertex-capacitated maximum reweighted second eigenvalue $\vec{\lambda}_2^{v*}(G)$ that

$$\vec{\lambda}_2^{v*}(G) \lesssim \vec{\psi}(G) \lesssim \sqrt{\vec{\lambda}_2^{v*}(G) \cdot \log \frac{\Delta}{\vec{\lambda}_2^{v*}(G)}}.$$

This provides a combinatorial characterization of the fastest mixing time of a directed graph by vertex expansion, and builds a new connection between reweighted eigenvalues, vertex expansion, and fastest mixing time for directed graphs.

The second main result is a stronger Cheeger inequality relating the edge conductance $\vec{\phi}(G)$ of a directed graph G to the edge-capacitated maximum reweighted second eigenvalue $\vec{\lambda}_2^{e*}(G)$ that

$$\vec{\lambda}_2^{e*}(G) \lesssim \vec{\phi}(G) \lesssim \sqrt{\vec{\lambda}_2^{e*}(G) \cdot \log \frac{1}{\vec{\lambda}_2^{e*}(G)}}.$$

This provides a certificate for a directed graph to be an expander and a spectral algorithm to find a sparse cut in a directed graph, playing a similar role as Cheeger's inequality in certifying graph expansion and in the spectral partitioning algorithm for undirected graphs.

We also use this reweighted eigenvalue approach to derive the improved Cheeger inequality for directed graphs, and furthermore to derive several Cheeger inequalities for hypergraphs that match and improve the existing results in [Lou15, CLTZ18]. These are supporting results that this provides a unifying approach to lift the spectral theory for undirected graphs to more general settings.

5.1 Our Results

We formulate reweighted eigenvalues for directed graphs and hypergraphs. The main idea is to reduce the study of expansion properties in directed graphs and hypergraphs to the basic setting of edge conductances in undirected graphs. We show that this provides an intuitive and unifying approach to lift the spectral theory for undirected graphs to more general settings.

For convenience, we recall the definition of directed vertex expansion and directed edge conductance, which are the main combinatorial quantities of interest in this chapter.

Definition 5.1.1 (Directed Vertex Expansion (restatement of Definition 2.3.2))**.** *Let $G = (V, E, \pi)$ be a vertex-weighted directed graph. For a subset $S \subseteq V$, let $\partial^+(S) := \{v \notin S \mid \exists u \in S \text{ with } uv \in E\}$ be the set of out-neighbors of S. The directed vertex expansion of a set $S \subseteq V$ and of the graph G are defined as*[1]

$$\vec{\psi}(S) := \frac{\min\left\{\pi\big(\partial^+(S)\big), \pi\big(\partial^+(S^c)\big)\right\}}{\min\left\{\pi(S), \pi(S^c)\right\}} \quad \text{and} \quad \vec{\psi}(G) := \min_{\emptyset \neq S \subset V} \vec{\psi}(S).$$

Note that $\vec{\psi}(S) \leq 1$ for all $S \subseteq V$ as $\partial^+(S^c) \subseteq S$.

Definition 5.1.2 (Directed Edge Conductance [Yos16, Yos19] (restatement of Definition 2.3.1))**.** *Let $G = (V, E, w)$ be an edge-weighted directed graph. For a subset $S \subseteq V$, let $\delta^+(S) := \{uv \in E \mid u \in S \text{ and } v \notin S\}$ be the set of outgoing edges of S and*

[1]When specialized to undirected graphs (by considering the bidirected graph), the current definitions are slightly different from that in Section 2.3.1. We remark that the two definitions of $\psi(G)$ are within a factor of 2 of each other. The current definitions have the advantages that $\psi(S) \leq 1$ and are more convenient in the proofs.

$\mathrm{vol}_w(S) := \sum_{v\in S}\sum_{u\in V}(w(uv)+w(vu))$ *be the volume of S. The directed edge conductance of a set $S \subseteq V$ and of the graph G are defined as*

$$\vec{\phi}(S) := \frac{\min\left\{w\big(\delta^+(S)\big), w\big(\delta^+(S^c)\big)\right\}}{\min\left\{\mathrm{vol}_w(S), \mathrm{vol}_w(S^c)\right\}} \quad \textit{and} \quad \vec{\phi}(G) := \min_{\emptyset \neq S \subset V} \vec{\phi}(S).$$

5.1.1 Cheeger Inequality for Directed Vertex Expansion

Classical spectral theory connects (i) undirected edge conductance, (ii) second eigenvalue, and (iii) mixing time of random walks on undirected graphs. We present a new spectral formulation that connects (i) directed vertex expansion, (ii) reweighted second eigenvalue, and (iii) fastest mixing time of random walks on directed graphs.

To certify that a directed graph $G = (V, E)$ has large vertex expansion, our idea is to find the best reweighted *Eulerian* subgraph $G' = (V, E, w')$ of G with arc weight $w'(uv)$ for $uv \in E$ and weighted degrees $\sum_{u\in V} w'(uv) = \sum_{u\in V} w'(vu) = \pi(v)$ for all $v \in V$, and then use the edge conductance of G' as a lower bound on the vertex expansion of G. Since the weighted directed graph G' is Eulerian, the edge conductance of G' is equal to the edge conductance of the underlying undirected graph G'' with edge weight $w''(uv) = \frac{1}{2}\big(w'(uv) + w'(vu)\big)$. Now, as the graph G'' is undirected, we can use Cheeger's inequality to lower bound the edge conductance of G'' by the second smallest eigenvalue of its normalized Laplacian matrix. This leads to the following formulation of the reweighted second eigenvalue for directed vertex expansion (see Proposition 5.3.1 for more about this reduction).

Definition 5.1.3 (Maximum Reweighted Spectral Gap with Vertex Capacity Constraints)**.** *Given a directed graph $G = (V, E)$ and a weight function $\pi : V \to \mathbb{R}^+$, the maximum reweighted spectral gap with vertex capacity constraints is defined as*

$$\begin{aligned}
\vec{\lambda}_2^{v*}(G) \;:=\; \max_{A \geq 0} \quad & \lambda_2\left(I - \Pi^{-\frac{1}{2}}\left(\frac{A + A^T}{2}\right)\Pi^{-\frac{1}{2}}\right) & \\
\text{subject to} \quad & A(u,v) = 0 & \forall uv \notin E \\
& \sum_{v\in V} A(u,v) = \sum_{v\in V} A(v,u) & \forall u \in V \\
& \sum_{v\in V} A(u,v) = \pi(u) & \forall u \in V
\end{aligned}$$

where A is the adjacency matrix of the reweighted Eulerian subgraph and $\Pi := \mathrm{diag}(\pi)$ *is the diagonal degree matrix of A. Then $\frac{1}{2}(A + A^T)$ is the adjacency matrix of the underlying* *undirected graph of the reweighted Eulerian subgraph, $\mathcal{L} := I - \frac{1}{2}\Pi^{-1/2}(A + A^T)\Pi^{-1/2}$ is its normalized Laplacian matrix, and $\lambda_2(\mathcal{L})$ is the second smallest eigenvalue of $\mathcal{L}$.*

To ensure that the optimization problem for $\vec{\lambda}_2^{v*}(G)$ is always feasible, we assume that the graph has a self-loop at each vertex. In the context of Markov chains, this corresponds to allowing a non-zero holding probability on each vertex.

Our first main result is a Cheeger-type inequality that relates $\vec{\lambda}_2^{v*}(G)$ and $\vec{\psi}(G)$, proving that the directed vertex expansion is large if and only if the reweighted eigenvalue is large.

Theorem 5.1.4 (Cheeger Inequality for Directed Vertex Expansion)**.** *For any directed graph $G = (V, E)$ and any weight function $\pi : V \to \mathbb{R}_{\geq 0}$,*

$$\vec{\lambda}_2^{v*}(G) \lesssim \vec{\psi}(G) \lesssim \sqrt{\vec{\lambda}_2^{v*}(G) \cdot \log \frac{\Delta}{\vec{\psi}(G)}} \lesssim \sqrt{\vec{\lambda}_2^{v*}(G) \cdot \log \frac{\Delta}{\vec{\lambda}_2^{v*}(G)}},$$

where Δ is the maximum (unweighted) degree of a vertex of G.

Since directed vertex expansion is more general than undirected vertex expansion and Theorem 1.1.2 is tight up to a constant factor, we know that the $\log \Delta$ term in Theorem 5.1.4 is necessary. But we do not know whether the $\log(1/\vec{\psi}(G))$ term in Theorem 5.1.4 is necessary or not.

The Fastest Mixing Time Problem: The notion of reweighted eigenvalue for undirected graphs was first formulated in [BDX04] for studying the fastest mixing time problem on *reversible* Markov chains. It turns out that the reweighted eigenvalue $\vec{\lambda}_2^{v*}(G)$ in Definition 5.1.3 can be used to study the fastest mixing time problem on *general* Markov chains.

Definition 5.1.5 (Fastest Mixing Time on General Markov Chain)**.** *Given a directed graph $G = (V, E)$ and a probability distribution π on V, the fastest mixing time problem is defined as*

$$\begin{aligned}
\tau^*_{\text{mix}}(G) \;:=\; \min_{P \geq 0} \quad & \tau_{\text{mix}}(P) \\
\text{subject to} \quad & P(u,v) = 0 && \forall uv \notin E \\
& \sum_{u \in V} P(v,u) = 1 && \forall v \in V \\
& \sum_{u \in V} \pi(u) P(u,v) = \pi(v) && \forall v \in V
\end{aligned}$$

where P is the transition matrix of the Markov chain. The constraints are to ensure that P has nonzero entries only on the edges of G, that P is a row stochastic matrix, and that the stationary distribution of P is π. The objective is to minimize the mixing time $\tau_{\text{mix}}(P)$ to the stationary distribution π.

For the fastest mixing time problem on reversible Markov chains introduced in Section 3.2.2, we are given an undirected graph $G = (V, E)$ and a probability distribution π, and the last set of constraints in Definition 5.1.5 is replaced by the stronger requirement that $\pi(u)P(u, v) = \pi(v)P(v, u)$ for all $uv \in E$. With this stronger requirement, P has real eigenvalues and it is well known (c.f. Proposition 2.6.2) that $\tau_{\text{mix}}(P) \lesssim \frac{1}{1-\alpha_2(P)} \cdot \log(\frac{1}{\pi_{\min}})$, where $\alpha_2(P)$ is the second largest eigenvalue of P and $\pi_{\min} := \min_{v \in V} \pi(v)$. Thus, the reweighted eigenvalue formulation in [BDX04] is to find such a transition matrix P that maximizes the spectral gap $1 - \alpha_2(P)$, which can be solved by a semidefinite program and can be used as a proxy for upper bounding the fastest mixing time (see Section 3.2.2).

This is a well-motivated problem in the study of Markov chains and has generated considerable interest (see the references in [OZ22]), but there had been no known combinatorial characterization of the fastest mixing time for quite some time. The result by Oleskar-Taylor and Zanetti [OZ22], and our subsequent improvement in Theorem 4.1.3, give a combinatorial characterization of the fastest mixing time of reversible Markov chains by the vertex expansion of the graph.

For general Markov chains, P may have complex eigenvalues, and there was no known efficient formulation for the fastest mixing time problem. We observe that the reweighted spectral gap $\vec{\lambda}_2^{v*}(G)$ in Definition 5.1.3 provides such a formulation through the results in [Fil91, Chu05]. An interesting consequence of Theorem 5.1.4 is a combinatorial characterization of the fastest mixing time of general Markov chains, showing that small directed vertex expansion is the only obstruction of fastest mixing time.

Theorem 5.1.6 (Fastest Mixing Time and Directed Vertex Expansion)**.** *For any directed graph $G = (V, E)$ with maximum total degree Δ, and any probability distribution π on V,*

$$\frac{1}{\vec{\psi}(G)} \lesssim \tau^*(G) \lesssim \frac{1}{\vec{\psi}(G)^2} \cdot \log \frac{\Delta}{\vec{\psi}(G)} \cdot \log \frac{1}{\pi_{\min}}.$$

(Note that the lower bound is improved from that in [LTW23].)

Together, Theorem 5.1.4 and Theorem 5.1.6 connect (i) the reweighted second eigenvalue, (ii) directed vertex expansion, and (iii) fastest mixing time on directed graphs, in a similar way that classical spectral graph theory connects (i) the second eigenvalue, (ii) undirected edge conductance, and (iii) mixing time on undirected graphs. Theorem 5.1.6 is a significant generalization of the aforementioned results on reversible Markov chains to general Markov chains, and we believe it to be of independent interest.

5.1.2 Cheeger Inequality for Directed Edge Conductance

Two key applications of Cheeger's inequality are to use the second eigenvalue to certify whether an undirected graph is an expander graph, and to provide a spectral algorithm for graph partitioning that is useful in many areas. We present a new inequality for directed graphs for these purposes.

We use the same approach as described in Section 5.1.1 to prove a Cheeger-type inequality for directed edge conductance.[2] To certify that a directed graph $G = (V, E, w)$ has large edge conductance, we find the best reweighted Eulerian subgraph G' with edge weight $w'(uv) \le w(uv)$ for each $uv \in E$, and use the edge conductance of G' (with respect to the volumes using w) to provide a lower bound on the edge conductance of G. Then, the edge conductance of G' is reduced to the edge conductance of the underlying undirected graph G'' with edge weight $w''(uv) = \frac{1}{2}(w'(uv) + w'(vu))$, and the second smallest eigenvalue of the normalized Laplacian matrix of G'' is used to provide a lower bound on the edge conductance of G''. See Proposition 5.3.2 for a proof.

Definition 5.1.7 (Maximum Reweighted Spectral Gap with Edge Capacity Constraints (Restatement of Definition 1.2.1)). *Given a directed graph $G = (V, E)$ and a weight function $w : E \to \mathbb{R}_{\ge 0}$, the maximum reweighted spectral gap with edge capacity constraints is defined as*

$$\begin{aligned}
\vec{\lambda}_2^{e*}(G) \;:=\; \max_{A \ge 0} \quad & \lambda_2\bigg(D^{-\frac{1}{2}}\Big(D_A - \frac{A + A^T}{2}\Big)D^{-\frac{1}{2}}\bigg) & \\
\text{subject to} \quad & A(u,v) = 0 & \forall uv \notin E \\
& \sum_{v \in V} A(u,v) = \sum_{v \in V} A(v,u) & \forall u \in V \\
& A(u,v) \le w(uv) & \forall uv \in E
\end{aligned}$$

where A is the adjacency matrix of the reweighted Eulerian subgraph, D_A is the diagonal degree matrix of $(A + A^T)/2$ with $D_A(v,v) = \sum_{u \in V} \frac{1}{2}(A(u,v) + A(v,u))$, and D is the

[2] The reader may wonder whether it is possible to reduce directed edge conductance to directed vertex expansion, and use Theorem 5.1.4 to obtain a Cheeger-type inequality for directed edge conductance. This is indeed possible, but the result obtained in this way will have a dependency on the maximum total degree Δ as in Theorem 5.1.4, while the result that we present in Theorem 5.1.8 has no such dependency.

diagonal degree matrix of G with $D(v,v) = \sum_{u \in V}(w(uv)+w(vu))$ equal to the total weighted degree of v in G.

Our second main result is a stronger Cheeger inequality that relates $\vec{\lambda}_2^{e*}(G)$ and $\vec{\phi}(G)$.

Theorem 5.1.8 (Cheeger Inequality for Directed Edge Conductance)**.** *For any directed graph $G = (V, E)$ and any weight function $w : E \to \mathbb{R}_{\geq 0}$,*

$$\vec{\lambda}_2^{e*}(G) \lesssim \vec{\phi}(G) \lesssim \sqrt{\vec{\lambda}_2^{e*}(G) \cdot \log \frac{1}{\vec{\phi}(G)}} \lesssim \sqrt{\vec{\lambda}_2^{e*}(G) \cdot \log \frac{1}{\vec{\lambda}_2^{e*}(G)}}.$$

An important point about Theorem 5.1.8 is that there is no dependence on the maximum degree of G as in Theorem 5.1.4 or on the number of vertices of G as in [ACMM05, Yos19]. As a consequence, $\vec{\lambda}_2^{e*}(G)$ is a polynomial time-computable quantity that can be used to certify whether a directed graph has constant edge conductance. This is similar to the role of the second eigenvalue in Cheeger's inequality to certify whether an undirected graph has constant edge conductance.

Also, as in the proof of Cheeger's inequality, the proof of Theorem 5.1.8 provides a polynomial time algorithm to return a set S with $\vec{\phi}(S) \leq \sqrt{\vec{\phi}(G) \log 1/\vec{\phi}(G)}$. Since many real-world networks are directed [Yos16], we hope that this "spectral" algorithm will find applications in clustering and partitioning for directed graphs, as the classical spectral partitioning algorithm does in clustering and partitioning for undirected graphs [SM00, Lux07].

Fill [Fil91] and Chung [Chu05] defined some symmetric matrices for directed graphs, and related their eigenvalues to Cheeger constant and to mixing time (see Section 3.3.1). The main difference between our formulation and Chung's formulation is that we search for an *optimal* reweighting while Chung used a specific vertex-based reweighting by the stationary distribution. We note that the Cheeger constant in (3.9) could be very different from the directed edge conductance in Definition 5.1.2 and the directed vertex expansion in Definition 5.1.1; see Section 5.9.1 for some examples. We remark that many subsequent works used Cheeger constant as the objective for clustering and partitioning for directed graphs, and these examples illustrate their limitations in finding sets of small directed edge conductance or directed vertex expansion, which are much more suitable notions for clustering and partitioning; see [Yos16] for related discussions.

Additional Previous Work: Besides the matrices in [Fil91, Chu05], there are other Hermitian matrices associated to a directed graph studied in the literature. Guo and

Mohar [GM17] and Liu and Li [LL15] defined the Hermitian adjacency matrix H of a directed graph as $H(u,v) = 1$ if both $uv, vu \in E$, $H(u,v) = \imath$ if $uv \in E$ and $vu \notin E$ where $\imath$ is the imaginary unit, $H(u,v) = -\imath$ if $uv \notin E$ and $vu \in E$, and $H(u,v) = 0$ if both $uv, vu \notin E$. There are also other Hermitian matrices defined for clustering directed graphs [LS20, CLSZ20] and for the Max-2-Lin problem [LSZ19]. We confirm that there are no known relations between the eigenvalues of these Hermitian matrices and the expansion properties of a directed graph.

Yoshida [Yos16] introduced a nonlinear Laplacian operator for directed graphs and used it to define the second eigenvalue λ_G, and later [Yos19] supplied an SDP approximation for λ_G. This gives a polynomial time computable quantity $\tilde{\lambda}_G$ that satisfies $\tilde{\lambda}_G \lesssim \vec{\phi}(G) \lesssim \sqrt{\tilde{\lambda}_G \cdot \log n}$ (see Section 3.3.2). We note that this result is comparable to but improved by our result[3] for $\vec{\phi}(G)$ in (5.1), and cannot be used for certifying constant directed edge conductance as in Theorem 5.1.8. We also note that this result is dominated by the SDP-based $O(\sqrt{\log n})$-approximation algorithm for $\vec{\phi}(G)$ in [ACMM05] (see Section 3.6.3) that we recap below. To our knowledge, this is the only spectral formulation known in the literature that relates to directed edge conductance, and no spectral formulation was known for directed vertex expansion. We also believe that the reweighted eigenvalue approach is simpler and more intuitive than the nonlinear Laplacian operator approach.

Agarwal, Charikar, Makarychev, Makarychev [ACMM05] gave an SDP-based $O\big(\sqrt{\log n}\big)$-approximation algorithm for the directed sparsest cut problem on a directed graph $G = (V, E)$, where the objective is to find a set S that minimizes $|\delta^+(S)| / \min\{|S|, |S^c|\}$. We note that in the unweighted case, directed vertex expansion and directed edge conductance can be reduced to directed sparsest cut via standard reductions. In the weighted case, the SDP for directed sparsest cut can be slightly modified to obtain a $O\big(\sqrt{\log n}\big)$-approximation algorithm for directed edge conductance; see Section 5.9.2. To our knowledge, it was not known that the SDP in [ACMM05] can be used to certify whether a directed graph has constant edge conductance as in Theorem 5.1.8, as the analysis using triangle inequalities based on [ARV09] has a $\sqrt{\log n}$ factor loss. We show in Section 5.9.2 that the SDP in [ACMM05] is stronger than the SDP for directed edge conductance in Proposition 5.4.2. Therefore, using the new analysis in this chapter, we prove that the SDP in [ACMM05] also provides a polynomial time computable quantity to certify constant directed edge conductance as in Theorem 5.1.8.

[3] We remark that we can use the Johnson-Lindenstrauss lemma to do the dimension reduction step as in [OZ22], and this would give $\vec{\phi}(G) \lesssim \sqrt{\vec{\lambda}_2^{e*}(G) \cdot \log n}$ as well.

5.1.3 Generalizations of Cheeger's Inequality for Directed Graphs

For undirected graphs, there are several interesting generalizations of Cheeger's inequality: Trevisan's result [Tre09] that relate λ_n to bipartite edge conductance, the higher-order Cheeger's inequality [LOT12, LRTV12] that relates λ_k to k-way edge conductance, and the improved Cheeger's inequality [KLL$^+$13] that relates λ_2 and λ_k to edge conductance. See Section 3.1 for a review of these results.

As in Chapter 4, we study whether there are close analogs of these results for directed graphs, using reweighted eigenvalues for directed vertex expansion in Definition 5.1.1 and directed edge conductance in Definition 5.1.2. Perhaps surprisingly, we show that the natural analogs of Trevisan's result and higher-order Cheeger's inequality do not hold, but we obtain analogs of the improved Cheeger's inequality for directed vertex expansion and directed edge conductance. See Section 5.10 for these results.

Chan, Tang and Zhang [CTZ15] gave a higher-order Cheeger inequality for directed graphs. Roughly speaking, they showed that there are k disjoint subsets $S_1, \ldots, S_k \subseteq V$ with $\lambda_k(\tilde{\mathcal{L}}) \lesssim h(S_i) \lesssim k^2 \cdot \sqrt{\lambda_k(\tilde{\mathcal{L}})}$ for $1 \leq i \leq k$, where $h(S_i)$ is the Cheeger constant in (3.9) and $\lambda_k(\tilde{\mathcal{L}})$ is the k-th smallest eigenvalue of the Laplacian in (3.8). The proof is a direct application of the higher-order Cheeger inequality for undirected graphs on the reweighted subgraph by the stationary distribution. In Section 5.9.1, we show an example that rules out the possibility of having a higher-order Cheeger inequality for directed graphs relating $\lambda_k(\tilde{\mathcal{L}})$ to k-way directed edge conductance.

5.1.4 Cheeger Inequalities for Hypergraph Edge Conductance

We also formulate reweighted eigenvalues for hypergraphs and use them to derive Cheeger-type inequalities for hypergraphs, as supporting results that reweighted eigenvalues provide a unifying approach to study expansion properties in different settings. The idea is simply to consider the "clique-graph" of the hypergraph H, and find the best reweighted subgraph of the clique-graph to certify the edge conductance of H, subject to the constraint that the total weight of the "clique-edges" of a hyperedge e is bounded by $w(e)$.

Definition 5.1.9 (Maximum Reweighted Spectral Gap for Hypergraphs)**.** *Given an edge-weighted hypergraph $H = (V, E, w)$, the maximum reweighted spectral gap for H is defined* *as*

$$\begin{aligned} \gamma_2^*(H) \;:=\; \max_{A \geq 0} \quad & \lambda_2\Big(D^{-\frac{1}{2}}\big(D_A - A\big)D^{-\frac{1}{2}}\Big) \\ \text{subject to} \quad & \sum_{u,v \in e} c(u,v,e) \leq w(e) && \forall e \in E \\ & A(u,v) = \sum_{e \in E : u,v \in e} c(u,v,e) && \forall u,v \in V. \end{aligned}$$

In this formulation, there is a clique-edge variable $c(u,v,e)$ for each pair of vertices u,v in a hyperedge e, with the constraints that the total weight of the clique-edges in e is bounded by $w(e)$. Then, A is the adjacency matrix of the reweighted subgraph of the clique-graph with edge weight $A(u,v)$ equal to the sum of the weight of the clique-edges involving u and v, D_A is the diagonal degree matrix of A with $D_A(v,v) = \sum_{u \in V} A(u,v)$, and D is the diagonal degree matrix of H with $D(v,v) = \sum_{e \in E : v \in e} w(e)$ equal to the weighted degree of v in H.

There is a spectral theory for hypergraphs based on a continuous time diffusion process with several Cheeger-type inequalities proven in [Lou15, CLTZ18]; see Section 3.4.1. We show that the reweighted eigenvalue approach can be used to provide a simpler and more intuitive way to obtain similar results.

Theorem 5.1.10 (Cheeger Inequality for Hypergraph Edge Conductance)**.** *For any edge-weighted hypergraph $H = (V,E,w)$ of rank r,*

$$\gamma_2^*(H) \lesssim \phi(H) \lesssim \sqrt{\gamma_2^*(H) \cdot \log r}.$$

We also obtain generalizations of Cheeger's inequalities for hypergraphs using other reweighted eigenvalues such as $\gamma_k^*(H)$ and a new result about improved Cheeger inequality for hypergraphs. Using the reweighted eigenvalue approach, we can define the maximum reweighted k-th eigenvalue γ_k^* as in Definition 5.1.9 and prove the following analog of higher-order Cheeger inequality for hypergraph edge conductance in Theorem 5.11.1: for any $\varepsilon \geq 1/k$, there are disjoint subsets $S_1, \ldots, S_{(1-\varepsilon)k}$ with

$$\phi(S_i) \lesssim \sqrt{k} \cdot \varepsilon^{-4} \cdot \log k \cdot \sqrt{\log r} \cdot \sqrt{\gamma_k^*}$$

for all $i \leq (1-\varepsilon)k$. This bound is comparable to that in [CLTZ18] when $\varepsilon \approx 1/k$, and is an improvement when $\varepsilon = \Theta(1)$ by a factor of more than k^2. This also improves the approximation algorithm for the small-set hypergraph edge conductance problem in [CLTZ18] by a factor of more than k. In addition, we also prove an analog of the improved Cheeger's inequality [KLL+13] for hypergraphs. See Section 5.11 for the precise statements of all these results, and see Section 3.4.1 for a more detailed survey of the results in [CLTZ18].

Compared to the spectral theory in [Lou15, CLTZ18] for hypergraphs using the continuous time diffusion process, we believe that the reweighted eigenvalue approach is simpler and more intuitive. The definitions of the hypergraph diffusion process and its eigenvalues are quite technically involved and require considerable effort to make rigorous [CTWZ19]. The reweighted eigenvalue approach allows us to recover and improve their results on hypergraph partitioning, and also to obtain a new result. Since their spectral theory for hypergraph partitioning is gaining more attention in machine learning lately (see e.g. [LM18]), we believe that it would be beneficial to have an alternative approach that is easier to understand and to prove new results and to have efficient implementations.

5.2 Our Techniques

Conceptually, our contribution is to come up with new spectral formulations for expansion properties in directed graphs and hypergraphs, and to show that the reweighted eigenvalue approach provides a unifying method to reduce expansion problems in more general settings to the basic setting of edge conductances in undirected graphs.

Technically, the proofs are based on the framework developed in Chapter 4 in relating reweighted eigenvalues to undirected vertex expansion in Theorem 1.1.2. Recall that there are two main steps in proving Theorem 1.1.2. The first step is to construct the dual SDP for the reweighted eigenvalue, and to do random projection to obtain a 1-dimensional solution to the dual program. The second step is to analyze the threshold rounding algorithm for the 1-dimensional solution. Below we highlight some new elements in our proofs for directed graphs.

For directed graphs, we identify a key parameter for our analysis.

Definition 5.2.1 (Asymmetric Ratio of Directed Graphs)**.** *Given an edge-weighted graph* $G = (V, E, w)$*, the asymmetric ratio of a set* $S \subseteq V$ *and of the graph* G *are defined as*

$$\alpha(S) := \frac{w(\delta^+(S))}{w(\delta^+(S^c))} \quad \text{and} \quad \alpha(G) := \max_{\emptyset \neq S \subset V} \alpha(S).$$

Given a vertex-weighted graph $G = (V, E, \pi)$*, we define the* π*-induced weight of an edge* $uv \in E$ *as* $w_\pi(uv) = \min\{\pi(u), \pi(v)\}$*, and the asymmetric ratio of a set* $S \subseteq V$ *and of the graph are defined as above using the edge weight function* w_π*.*

We note that the asymmetric ratio of an edge-weighted graph was defined in [EMPS16] with the name "α-balanced" and was used in the analysis of oblivious routing in directed graphs. The asymmetric ratio is a measure of how close a directed graph is to an undirected graph for our purpose, as when $\alpha(G) = 1$ the directed graph is Eulerian and so its edge conductance is the same as the edge conductance of the underlying undirected graph.

This parameter is defined to satisfy two useful properties. The first is that it can be used to prove more refined Cheeger inequalities that

$$\vec{\phi}(G) \leq \sqrt{\vec{\lambda}_2^{e*}(G) \cdot \log \alpha(G)} \quad \text{and} \quad \vec{\psi}(G) \leq \sqrt{\vec{\lambda}_2^{v*}(G) \cdot \log \big(\Delta \cdot \alpha(G)\big)}. \tag{5.1}$$

The second is that it can be related to the directed edge conductance and directed vertex expansion such that $\alpha(G) \leq 1/\vec{\phi}(G)$ (Lemma 5.5.1) and $\alpha(G) \leq \Delta/\vec{\psi}(G)$ (Lemma 5.5.2). Combining the two properties gives Theorem 5.1.8 and Theorem 5.1.4.

We highlight two new elements in the proofs of (5.1), one in dimension reduction and one in threshold rounding. In the dimension reduction step, the Johnson-Lindenstrauss lemma can be used to project to a 1-dimensional solution with a factor of $\log n$ loss as in [OZ22]. For undirected vertex expansion, this was improved to a factor of $\log \Delta$ loss in two ways: one is the Gaussian projection method [LRV13, KLT22] in Chapter 4, while the other is a better analysis of dimension reduction for maximum matching [JPV22] in Section 3.2.3. For directed edge conductance and directed vertex expansion, the SDP is more complicated and we do not know how to extend the Gaussian projection method to improve on the $O(\log n)$ loss; see Section 5.6.1 for discussions. Instead, we extend the approach in [JPV22] to prove that random projections only lose a factor of $\log \alpha(G)$ with high probability. When the asymmetric ratio is small, we use Hoffman's result in Lemma 5.5.4 about bounded-weighted circulations to prove a "large optimal property" of the SDPs (see Lemma 5.5.5), and use it to adapt the proof in [JPV22] for maximum weighted Eulerian subgraphs; see Section 5.6 for details.

In the threshold rounding step of the 1-dimensional solution, we consider the dual SDP of $\vec{\lambda}_2^{v*}(G)$ and $\vec{\lambda}_2^{e*}(G)$ as in Proposition 4.3.1. Unlike the dual SDP for undirected vertex expansion, these dual SDPs (see Lemma 5.7.5) has some negative terms from a vertex potential function $r : V \to \mathbb{R}$. The new idea in our threshold rounding is to not just consider the ordering defined by the vertex embedding function $f : V \to \mathbb{R}$ as usual, but to consider the two orderings defined by $f \pm r$ and show that threshold rounding will work on one of these two orderings. See Section 5.7 for details.

The generalizations of Cheeger inequalities for directed graphs and all Cheeger-type inequalities for hypergraphs are based on the same proofs of the corresponding results in Chapter 4 with no new ideas involved. These results show that the reweighted eigenvalue approach provides a unifying method to lift the spectral theory for undirected edge conductance to obtain new results in more general settings in a systematic way.

Finally, we note that the maximum degree Δ for undirected vertex expansion, the asymmetric ratio $\alpha(G)$ for directed edge conductance and directed vertex expansion, and the maximum hyperedge size r for hypergraph edge conductance all play the same role as a measure of how close the respective problem is to the basic problem of undirected edge conductance. The trivial reductions to undirected edge conductance lose a factor of Δ for undirected vertex expansion, a factor of $\alpha(G)$ for directed edge-conductance (by just ignoring the directions), and a factor of r for hypergraph edge conductance (by just considering the clique graph). In comparison, the reductions through the reweighted eigenvalue approach only lose a factor of $\log \Delta$ in (4.1), a factor of $\log \alpha(G)$ in (5.1), and a factor of $\log r$ in Theorem 5.1.10 respectively.

Remark 5.2.2 (Eulerian Reweighting in Directed Laplacian Solvers)**.** *We note that the idea of reducing the problem for a directed graph to an Eulerian directed graph was also used in directed Laplacian solvers [CKP+16, CKP+17]. As in [Chu05], they also use the same reweighting by the stationary distribution to obtain an Eulerian graph from a directed graph. (Furthermore, they introduced a notion of spectral sparsification of Eulerian directed graphs.) We believe that the idea of reducing to Eulerian directed graphs and the concept of asymmetric ratio will find more applications in solving problems on directed graphs.*

Remark 5.2.3 (More Applications of Reweighting)**.** *Finally, as a technical remark, we note that some careful reweighting schemes are crucially used in the construction of the diffusion process [Lou15, CLTZ18], and also in recent exciting developments in hypergraph spectral sparsification [CKN20, KKTY22] (called balanced weight assignments). This suggests that the concept of* reweighting *is central to these recent developments, and it would be very interesting to find connections between the different reweighting methods used in this work and these previous works.*

5.2.1 Chapter Plan

We begin the rest of the chapter by proving the two main results Theorem 5.1.4 and Theorem 5.1.8. First, we prove the easy direction of the two results in Section 5.3, and write the semidefinite programs for the reweighted eigenvalues in Section 5.4. Then, we show some properties of the asymmetric ratio in Section 5.5, and use these properties and the proof in [JPV22] to analyze a random projection algorithm to construct 1-dimensional spectral solutions to the semidefinite programs in Section 5.6. Then, we analyze a new threshold rounding algorithm for the 1-dimensional solutions to the dual programs, and prove the hard direction of the two results in Section 5.7. After that, we show Theorem 5.1.6 about fastest mixing time using [Fil91, Chu05] in Section 5.8, and provide details about the relations with some previous work in Section 5.9.

Finally, we discuss generalizations of Cheeger inequalities for directed graphs in Section 5.10 and derive Cheeger-type inequalities for hypergraphs in Section 5.11.

5.3 Easy Directions by Reductions

There are two ways to prove the easy directions in Theorem 5.1.4 and Theorem 5.1.8. A standard way is to construct a solution to $\vec{\lambda}_2^{v*}(G)$ or $\vec{\lambda}_2^{e*}(G)$ with small objective value when the directed vertex expansion or the directed edge conductance is small. Here instead, we show how to use the reduction idea discussed in the introduction to this chapter to prove the easy directions, as this is how we came up with the formulations and the reduction is the main theme here.

Proposition 5.3.1 (Easy Direction for Directed Vertex Expansion)**.** *For any directed graph $G = (V, E)$ with weight function $\pi : V \to \mathbb{R}_{\geq 0}$, it holds that $\vec{\lambda}_2^{v*}(G) \leq 2\vec{\psi}(G)$.*

Proof. The idea is to reduce directed vertex expansion of G to the directed edge conductance of the reweighted Eulerian subgraph defined by A in Definition 5.1.3, and then further reduce to the underlying undirected graph defined by $\frac{1}{2}(A + A^T)$ and use classical Cheeger's inequality to lower bound its edge conductance by the second eigenvalue of its normalized Laplacian matrix.

Let $w(uv) := A(u, v)$ be the edge weight in the Eulerian reweighted subgraph for $uv \in E$. For any nonempty $S \subset V$, by Definition 5.1.1 of directed vertex expansion and Definition 5.1.2 of directed edge conductance,

$$\vec{\psi}(S) = \frac{\min\left\{\pi(\partial^+(S)), \pi(\partial^+(S^c))\right\}}{\min\{\pi(S), \pi(S^c)\}} \geq \frac{2 \cdot \min\left\{w(\delta^+(S)), w(\delta^-(S))\right\}}{\min\{\mathrm{vol}_w(S), \mathrm{vol}_w(S^c)\}} = 2\vec{\phi}(S)$$

where we use the degree constraints in Definition 5.1.3 to establish that $w(\delta^+(S)) \leq \pi(\partial^+(S))$ and $w(\delta^-(S)) \leq \pi(\partial^+(S^c))$ (note that they are not necessarily equalities because of the self-loops), and $\mathrm{vol}_w(S) = 2\pi(S)$ for every nonempty $S \subset V$.

As the edge-weighted directed graph $G' = (V, E, w)$ is Eulerian, it holds that $w(\delta^+(S)) = w(\delta^-(S))$ for every nonempty $S \subset V$, and thus the directed edge conductance of G' is equal to half the edge conductance of the underlying undirected graph G'' with edge weight $w''(uv) = \frac{1}{2}\big(w(uv) + w(vu)\big)$, because

$$2\vec{\phi}(S) = \frac{\min\left\{w(\delta^+(S)), w(\delta^-(S))\right\}}{\frac{1}{2}\cdot \min\{\mathrm{vol}_w(S), \mathrm{vol}_w(S^c)\}} = \frac{w''(\delta(S))}{\min\{\mathrm{vol}_{w''}(S), \mathrm{vol}_{w''}(S^c)\}} = \phi(S).$$

As the graph G'' is undirected, we can use Cheeger's inequality in (1.1) to lower bound the edge conductance of G'' by the second smallest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(A) := I - \frac{1}{2}\Pi^{-1/2}(A + A^T)\Pi^{-1/2}$. Therefore, for any nonempty $S \subset V$, $\vec{\psi}(S) \geq 2\vec{\phi}(S) = \phi(S) \geq \lambda_2(\mathcal{L}(A))/2$. Since this holds for any nonempty $S \subset V$ and any weighted Eulerian subgraph defined by A satisfying the constraints in Definition 5.1.3, we conclude that $2\vec{\psi}(G) \geq \max_A \lambda_2(\mathcal{L}(A)) = \vec{\lambda}_2^{v*}(G)$. □

The proof of the easy direction of Theorem 5.1.8 is similar.

Proposition 5.3.2 (Easy Direction for Directed Edge Conductance)**.** *For any directed graph $G = (V, E)$ with weight function $w : E \to \mathbb{R}_{\geq 0}$, it holds that $\vec{\lambda}_2^{e*}(G) \leq 2\vec{\phi}(G)$.*

Proof. Let $w'(uv) := A(u, v)$ be the edge weight in the Eulerian reweighted subgraph G' in Definition 5.1.7. Let G'' be the underlying undirected graph with edge weight $w''(uv) := \frac{1}{2}(w'(uv) + w'(vu))$, with an additional self-loop on each vertex so that the weighted degree $\deg_{w'}(v)$ on each vertex v is exactly equal to the total degree $\deg_w(v) = \sum_{u\in V}(w(uv) + w(vu))$ of v in G. Then, by the edge capacity constraints and the Eulerian constraints in Definition 5.1.7, for any nonempty $S \subset V$,

$$\begin{aligned}\vec{\phi}(S) &= \frac{\min\left\{w(\delta^+(S)), w(\delta^-(S))\right\}}{\min\{\mathrm{vol}_w(S), \mathrm{vol}_w(S^c)\}} \\ &\geq \frac{\min\left\{w'(\delta^+(S)), w'(\delta^-(S))\right\}}{\min\{\mathrm{vol}_w(S), \mathrm{vol}_w(S^c)\}} = \frac{w''(\delta(S))}{\min\{\mathrm{vol}_{w''}(S), \mathrm{vol}_{w''}(S^c)\}} = \phi(S).\end{aligned}$$

Let $D := \mathrm{diag}(\deg_w)$ be the diagonal degree matrix of G'', and $\mathcal{L} := D^{-1/2}(D - \frac{1}{2}(A + A^T))D^{-1/2} = I - \frac{1}{2}D^{-1/2}(A + A^T)D^{-1/2}$ be the normalized Laplacian matrix of G''. As G'' is undirected, it follows from Cheeger's inequality in (1.1) that $\phi(S) \geq \lambda_2(\mathcal{L})/2$. Since this holds for any nonempty $S \subset V$ and any weighted subgraph defined by A satisfying the constraints in Definition 5.1.7, we conclude that $\vec{\phi}(G) \geq \phi(G'') \geq \max_A \lambda_2(\mathcal{L})/2 = \vec{\lambda}_2^{e*}(G)/2$. □

5.4 Semidefinite Programs

We show that the optimization problems of reweighted eigenvalues in these directed settings can also be formulated as SDP's, and so they can be approximated arbitrarily well in polynomial time. The construction is similar to that of the semidefinite program for undirected vertex expansion in Proposition 3.2.3, but presented in a min-max form.

Proposition 5.4.1 (SDP for Reweighted Second Eigenvalue with Vertex Capacity Constraints)**.** *Given a directed graph $G = (V, E)$ and a weight function $\pi : V \to \mathbb{R}_{\geq 0}$, the optimization problem in Definition 5.1.3 can be written as*

$$\begin{aligned}
\vec{\lambda}_2^{v*}(G) = \min_{f:V\to\mathbb{R}^n} \max_{A\geq 0} \quad & \frac{1}{2}\sum_{uv\in E} A(u,v)\cdot\|f(u)-f(v)\|^2 & \\
\text{subject to} \quad & A(u,v) = 0 & \forall uv \notin E \\
& \sum_{v\in V} A(u,v) = \sum_{v\in V} A(v,u) & \forall u \in V \\
& \sum_{v\in V} A(v,u) = \pi(u) & \forall u \in V \\
& \sum_{v\in V} \pi(v)\cdot f(v) = \vec{0} & \\
& \sum_{v\in V} \pi(v)\cdot \|f(v)\|^2 = 1. &
\end{aligned}$$

Proof. Let $\mathcal{L} := I - \frac{1}{2}\Pi^{-1/2}(A + A^T)\Pi^{-1/2}$ be the normalized Laplacian matrix in the objective function $\max_A \lambda_2(\mathcal{L})$ in Definition 5.1.3. By (2.2) in the preliminary chapter,

$$\lambda_2(\mathcal{L}) = \min_{f\perp\pi} \frac{\sum_{(u,v)\in\binom{V}{2}} \frac{1}{2}\big(A(u,v)+A(v,u)\big)\cdot(f(u)-f(v))^2}{\sum_v \pi(v)f(v)^2}.$$

Then we write $f \perp \pi$ as the second last constraint and normalize the denominator to 1 as the last constraint. By (2.5), the SDP relaxation where we replace $f : V \to \mathbb{R}$ by $f : V \to \mathbb{R}^n$ is an exact relaxation. After the SDP relaxation, the feasible domain becomes convex, and so we may apply von Neumann minimax theorem in Theorem 2.8.1 to switch the order of $\max_A \min_f$ in Definition 5.1.3 to $\min_f \max_A$ as in the statement of this lemma. □

The same construction is used for $\vec{\lambda}_2^{e*}(G)$ in Definition 5.1.7 and the proof is omitted.

Proposition 5.4.2 (SDP for Reweighted Second Eigenvalue with Edge Capacity Constraints)**.** *Given a directed graph $G = (V, E, w)$, the optimization problem in Definition 5.1.7 can be written as*

$$
\begin{aligned}
\vec{\lambda}_2^{e*}(G) := \min_{f:V\to\mathbb{R}^n} \max_{A\geq 0} \quad & \frac{1}{2}\sum_{uv\in E} A(u,v)\cdot\|f(u)-f(v)\|^2 \\
\text{subject to} \quad & A(u,v) = 0 && \forall uv \notin E \\
& \sum_{v\in V} A(u,v) = \sum_{v\in V} A(v,u) && \forall u \in V \\
& A(u,u) \leq w(uv) && \forall uv \in E \\
& \sum_{v\in V} \deg_w(v)\cdot f(v) = \vec{0} \\
& \sum_{v\in V} \deg_w(v)\cdot \|f(v)\|^2 = 1.
\end{aligned}
$$

We will use these SDP's to prove the two main results.

5.5 Asymmetric Ratio

A key parameter in our proofs is the asymmetric ratio $\alpha(G)$ in Definition 5.2.1. This parameter satisfies two useful properties. One is that $\alpha(G)$ can be used to bound the directed edge conductance and directed vertex expansion. Another is that directed graphs with bounded asymmetric ratio satisfy the "large optimal property" that we will describe in Section 5.5.2, which can be used in the proof in [JPV22] to provide a better analysis of the random projection algorithm for dimension reduction of the SDP solutions.

5.5.1 Asymmetric Ratio and Expansion Properties

The relation between asymmetric ratio of an edge-weighted graph and directed edge conductance is simple.

Lemma 5.5.1 (Asymmetric Ratio and Directed Edge Conductance)**.** *For any directed graph $G = (V, E, w)$, it holds that $\alpha(G) \leq 1/\vec{\phi}(G)$.*

Proof. Let $S \subset V$ be a nonempty set. Suppose $\mathrm{vol}_w(S) \le \mathrm{vol}_w(S^c)$; the other case is similar. Then, by the definition of directed edge conductance in Definition 5.1.2, $w(\delta^+(S)) \ge \vec{\phi}(G) \cdot \mathrm{vol}_w(S)$ and $w(\delta^-(S)) \ge \vec{\phi}(G) \cdot \mathrm{vol}_w(S)$. On the other hand, $w(\delta^+(S)) \le \mathrm{vol}_w(S)$ and $w(\delta^-(S)) \le \mathrm{vol}_w(S)$. Therefore, $\alpha(S) = w(\delta^+(S))/w(\delta^-(S)) \le 1/\vec{\phi}(G)$ for any nonempty $S \subset V$, and we conclude that $\alpha(G) \le 1/\vec{\phi}(G)$. □

The relation between asymmetric ratio of vertex-weighted graph and directed vertex expansion is less trivial and has a dependency on the maximum total degree Δ.

Lemma 5.5.2 (Asymmetric Ratio and Directed Vertex Expansion)**.** *For any directed graph $G = (V, E)$ with vertex weights $\pi : V \to \mathbb{R}^+$, it holds that $\alpha(G) \lesssim \Delta/\vec{\psi}(G)$.*

Proof. To upper bound $\alpha(G)$ for a vertex-weighted graph, by Definition 5.2.1, we need to upper bound $\alpha(S) = w_\pi(\delta^+(S))/w_\pi(\delta^+(S^c))$ and $\alpha(S^c) = w_\pi(\delta^+(S^c))/w_\pi(\delta^+(S))$ for any nonempty $S \subset V$, where $w_\pi(uv) = \min\{\pi(u), \pi(v)\}$ is the π-induced edge weight for $uv \in E$. We assume without loss of generality that $\pi(S) \le \pi(V)/2$.

For the numerators, note that $w_\pi(\delta^+(S)) \le \sum_{u \in S} \sum_{v:uv \in E} w_\pi(uv) \le \sum_{u \in S} \Delta \cdot \pi(u) = \Delta \cdot \pi(S)$, and similarly $w_\pi(\delta^+(S^c)) = w_\pi(\delta^-(S)) \le \sum_{u \in S} \sum_{v:vu \in E} w_\pi(vu) \le \sum_{u \in S} \Delta \cdot \pi(u) = \Delta \cdot \pi(S)$. Therefore, the numerators for $\alpha(S)$ and $\alpha(S^c)$ are at most $\Delta \cdot \pi(S)$. For the denominators, we claim that $w_\pi(\delta^+(S)) \ge \frac{1}{3}\vec{\psi}(G) \cdot \pi(S)$ and $w_\pi(\delta^+(S^c)) \ge \frac{1}{3}\vec{\psi}(G) \cdot \pi(S)$. This claim implies that the denominators for $\alpha(S)$ and $\alpha(S^c)$ are at least $\frac{1}{3}\vec{\psi}(G) \cdot \pi(S)$, and the lemma follows immediately.

To prove the claim, we first consider the lower bound on $w_\pi(\delta^+(S))$. Let $\varepsilon := \vec{\psi}(G)/3$. Suppose by contradiction that $w_\pi(\delta^+(S)) < \varepsilon \cdot \pi(S)$. Let $C_S := \{u \in S \mid \exists v \text{ with } uv \in \delta^+(S) \text{ and } \pi(u) \le \pi(v)\}$ and $C_{S^c} := \{v \in S^c \mid \exists u \text{ with } uv \in \delta^+(S) \text{ and } \pi(v) \le \pi(u)\}$. Since each $u \in C_S$ contributes at least $\pi(u)$ weight to $w_\pi(\delta^+(S))$ and these contributions are disjoint, it follows that $\pi(C_S) \le w_\pi(\delta^+(S)) < \varepsilon \cdot \pi(S)$. By the same argument, $\pi(C_{S^c}) < \varepsilon \cdot \pi(S)$. Note that, by definition of C_S and C_{S^c}, each edge in $\delta^+(S)$ has at least one vertex in $C_S \cup C_{S^c}$. This implies that $\partial^+(S - C_S) \subseteq C_S \cup C_{S^c}$, but this leads to the contradiction that

$$\vec{\psi}(S - C_S) \le \frac{\pi(\partial^+(S - C_S))}{\pi(S - C_S)} \le \frac{\pi(C_S \cup C_{S^c})}{\pi(S) - \pi(C_S)} < \frac{2\varepsilon \cdot \pi(S)}{(1 - \varepsilon) \cdot \pi(S)} = \frac{2\vec{\psi}(G)}{3(1 - \vec{\psi}(G)/3)} \le \vec{\psi}(G).$$

The lower bound on $w_\pi(\delta^+(S^c))$ is by a similar argument. Suppose by contradiction that $w_\pi(\delta^+(S^c)) < \varepsilon \cdot \pi(S)$. Let $C_{S^c} := \{u \in S^c \mid \exists v \text{ with } uv \in \delta^+(S^c) \text{ and } \pi(u) \le \pi(v)\}$ and $C_S := \{v \in S \mid \exists u \text{ with } uv \in \delta^+(S^c) \text{ and } \pi(v) \le \pi(u)\}$. Once again, it follows that

$\pi(C_S), \pi(C_{S^c}) \le w_\pi(\delta^+(S^c)) < \varepsilon \cdot \pi(S)$, and $\partial^+(S^c - C_{S^c}) \subseteq C_S \cup C_{S^c}$. But this leads to the contradiction that

$$\begin{aligned}\vec{\psi}(S^c - C_{S^c}) = \frac{\pi(\partial^+(S^c - C_{S^c}))}{\min\{\pi(S^c - C_{S^c}), \pi(S + C_{S^c})\}} &\le \frac{\pi(C_S \cup C_{S^c})}{\min\{\pi(S^c) - \varepsilon \cdot \pi(S), \pi(S)\}} \\ &< \frac{2\varepsilon \cdot \pi(S)}{(1-\varepsilon)\cdot \pi(S)} \le \vec{\psi}(G),\end{aligned}$$

where the second inequality uses that $\pi(S^c) \ge \pi(S)$. This completes the proof of the claim. □

5.5.2 Asymmetric Ratio and Large Optimal Property

Consider the semidefinite programs for $\vec{\lambda}_2^{v*}(G)$ and $\vec{\lambda}_2^{e*}(G)$ in Proposition 5.4.1 and Proposition 5.4.2. When the geometric embedding $f : V \to \mathbb{R}^n$ in the outer minimization problem is fixed, the inner maximization problem is simply to find a maximum weighted Eulerian subgraph A with vertex capacity constraints in Proposition 5.4.1 and with edge capacity constraints in Proposition 5.4.2. The following are trivial upper bounds on the optimal values of the inner maximization problems.

Claim 5.5.3 (Maximum Weighted Eulerian Subgraph with Capacity Constraints)**.** *Given a directed graph $G = (V, E)$ and an embedding $f : V \to \mathbb{R}^n$, let $\nu_f^{v*}(G)$ and $\nu_f^{e*}(G)$ be the objective values of the inner maximization problem in Proposition 5.4.1 and Proposition 5.4.2 respectively. Then*

$$\nu_f^{v*}(G) \le \frac{1}{2}\sum_{uv\in E} w_\pi(uv)\cdot \|f(u) - f(v)\|^2 \quad \text{and} \quad \nu_f^{e*}(G) \le \frac{1}{2}\sum_{uv\in E} w(uv)\cdot \|f(u) - f(v)\|^2,$$

where $w_\pi(uv) = \min\{\pi(u), \pi(v)\}$ is the π-induced edge weight function defined in Definition 5.2.1.

In the undirected vertex expansion problem [OZ22, JPV22], when $\pi(v) = \frac{1}{n}$ for all $v \in V$, the inner maximization problem is equivalent to the maximum weighted fractional matching problem. Jain, Pham and Vuong [JPV22] used the fact that any graph with maximum degree Δ has an edge coloring with at most $\Delta + 1$ colors to show that the inner maximization problem has a solution with weight at least $1/(\Delta + 1)$ fraction of the trivial upper bound. They then used this "large optimal property" to analyze a dimension reduction algorithm for maximum weighted matching; see Section 3.2.3.

We observe that the asymmetric ratio $\alpha(G)$ in Definition 5.2.1 can play the same role as Δ to establish the large optimal property for the maximum weighted Eulerian subgraph problems in Claim 5.5.3. The proof uses the following characterization of asymmetric ratio by Hoffman (see also [EMPS16, Theorem 2.3]), restated using our terminologies.

Lemma 5.5.4 (Hoffman's Circulation Lemma)**.** *Let $G = (V, E, w)$ be a directed graph. Then G has asymmetric ratio at most α if and only if there exists an Eulerian reweighting A of G such that*

$$\sum_{v:uv\in E} A(u,v) = \sum_{v:vu\in E} A(v,u) \ \forall u \in V \quad \text{and} \quad w(uv) \leq A(u,v) \leq \alpha \cdot w(uv) \ \forall uv \in E.$$

The large optimal property in terms of asymmetric ratio is a simple consequence of Hoffman's circulation lemma.

Lemma 5.5.5 (Large Optimal Property)**.** *Given a directed graph $G = (V, E)$ and an embedding $f : V \to \mathbb{R}^n$, let $\nu_f^{v*}(G)$ and $\nu_f^{e*}(G)$ be the objective values of the inner maximization problem in Proposition 5.4.1 and Proposition 5.4.2 respectively. Then*

$$\nu_f^{v*}(G) \geq \frac{1}{2\Delta \cdot \alpha(G)} \sum_{uv\in E} w_\pi(uv) \cdot \|f(u) - f(v)\|^2$$

and

$$\nu_f^{e*}(G) \geq \frac{1}{2\alpha(G)} \sum_{uv\in E} w(uv) \cdot \|f(u) - f(v)\|^2 .$$

Proof. First, consider $\nu_f^{e*}(G)$ in Proposition 5.4.2 with weight function $w : E \to \mathbb{R}_{\geq 0}$. Let A be an Eulerian reweighting of G with weight function w given in Lemma 5.5.4. As $w(uv) \leq A(u,v) \leq \alpha(G) \cdot w(uv)$ for $uv \in E$, the scaled-down subgraph $A/\alpha(G)$ satisfies the edge capacity constraints and is a feasible solution to the inner maximization problem in Proposition 5.4.2, with objective value $\frac{1}{2}\sum_{uv\in E} \frac{A(u,v)}{\alpha(G)} \cdot \|f(u) - f(v)\|^2 \geq \frac{1}{2\alpha(G)} \sum_{uv\in E} w(uv) \cdot \|f(u) - f(v)\|^2$.

Similarly, consider $\nu_f^{v*}(G)$ in Proposition 5.4.1 with weight function $\pi : V \to \mathbb{R}_{\geq 0}$ and induced function $w_\pi : E \to \mathbb{R}_{\geq 0}$. Let A be an Eulerian reweighting of G with weight function w_π given in Lemma 5.5.4. For each vertex u, the weighted degree is $\sum_{v:uv\in E} A(u,v) \leq \alpha(G) \cdot \sum_{v:uv\in E} w_\pi(uv) \leq \alpha(G) \cdot \Delta \cdot \pi(u)$. Therefore, scaling down A by a factor of $\Delta \cdot \alpha(G)$ satisfies the vertex capacity constraints and is a feasible solution to the inner maximization problem of Proposition 5.4.1, with objective value $\frac{1}{2\Delta\cdot\alpha(G)} \sum_{uv\in E} w_\pi(uv) \cdot \|f(u) - f(v)\|^2$. □

We will use Lemma 5.5.5 in the analysis of the dimension reduction step in the next subsection.

5.6 Dimension Reduction

The goal of this section is to obtain a good low-dimensional solution to the semidefinite programs in Proposition 5.4.1 and Proposition 5.4.2.

Definition 5.6.1 (Low-Dimensional Solutions to Semidefinite Programs)**.** *Define*

$$\vec{\lambda}_v^{(k)}(G) := \min_{f:V\to\mathbb{R}^k} \max_{A\geq 0} \frac{1}{2} \sum_{uv\in E} A(u,v)\cdot \|f(u)-f(v)\|^2$$

to be the objective value of the SDP in Proposition 5.4.1 when restricting f to be a k-dimensional embedding and subjecting to the same constraints.

Define $\vec{\lambda}_e^{(k)}(G)$ similarly as the objective value of the SDP in Proposition 5.4.2 when restricting f to be a k-dimensional embedding subjecting to the same constraints.

The main result that we will prove in this subsection is that there is a good 1-dimensional solution when the asymmetric ratio of the graph is small.

Theorem 5.6.2 (One Dimensional Solutions to Semidefinite Programs)**.** *Let $\vec{\lambda}_v^{(k)}(G)$ and $\vec{\lambda}_e^{(k)}(G)$ be as defined in Definition 5.6.1. Then*

$$\vec{\lambda}_v^{(1)}(G) \lesssim \log\big(\Delta\cdot\alpha(G)\big)\cdot\vec{\lambda}_2^{v*}(G) \quad\text{and}\quad \vec{\lambda}_e^{(1)}(G) \lesssim \log\alpha(G)\cdot\vec{\lambda}_2^{e*}(G).$$

Remark 5.6.3. *Using the tight example in Section 4.8 for undirected vertex expansion and a standard reduction from undirected vertex expansion to directed edge conductance, we can show that the second inequality in Theorem 5.6.2 is tight up to a constant factor. However, as this example has large maximum degree, we cannot conclude that the first inequality in Theorem 5.6.2 is also tight.*

5.6.1 Previous Work

For undirected vertex expansion, there are two different proofs of an analogous dimension reduction result with a factor of $\log\Delta$ loss. One is the proof presented in Section 4.3.2, and the other is by Jain, Pham, and Vuong [JPV22], reviewed in Section 3.2.3.

In Section 4.3.2, the approach was to first construct the dual SDP of $\lambda_2^*(G)$, where the objective function is of the form $\min_{f:V\to\mathbb{R}^n}\sum_{v\in V}\pi(v)\cdot\max_{u:uv\in E}\|f(u)-f(v)\|^2$. Since each maximum is over at most Δ terms, one can use the analysis of the Gaussian projection method in [LRV13] to directly project f to a 1-dimensional solution, and prove that the expected maximum is at most a factor of $O(\log \Delta)$ larger using properties of Gaussian random variables. For the semidefinite programs for $\vec{\lambda}_2^{v*}(G)$ and $\vec{\lambda}_2^{e*}(G)$, however, the objective function of the dual SDP is of the form $\min_{f:V\to\mathbb{R}^n} \sum_{v\in V} \pi(v)\cdot \max_{u:uv\in E} \big(\|f(u)-f(v)\|^2 - r(u)+r(v)\big)$ where $r(u)$ is a real number (see Lemma 5.7.5 for the 1-dimensional version). Since the contribution of $-r(u)+r(v)$ could be negative, the same approach of projecting f does not work anymore.[4]

Instead, we will follow the two-step approach of projecting the (primal) SDP solution in [JPV22]. In the first step, the n-dimensional solution to $\lambda_2^*(G)$ is projected to a $O(\log \Delta)$-dimensional solution, while the objective value only increases by a constant factor. Then the $O(\log \Delta)$-dimension solution is reduced to a 1-dimension solution, by choosing the best coordinate and losing a factor of $O(\log \Delta)$ as in [OZ22]. See Section 3.2.3 for details.

To analyze the first step, they proved Theorem 3.2.14 for maximum weighted matchings. We observe that their proof only needs the large optimal property of maximum matching as discussed in Section 5.5.2, but not any other property specific to matchings, and so it also works for maximum weighted Eulerian subgraphs in our problems with Lemma 5.5.5 about their large optimal property in place.

5.6.2 Random Projection

The arguments in this subsection are essentially the same as in [JPV22]. We cannot directly use their theorem as a black box and so we reproduce their arguments here. We will use the standard Gaussian projection algorithm as defined in Definition 3.1.7 when deriving the higher-order Cheeger inequalities. The main technical result is the following adaptation of Theorem 3.2.14 concerning dimension reduction for maximum matchings in [JPV22].

Theorem 5.6.4 (Dimension Reduction for Maximum Weighted Eulerian Subgraphs). *Let $\vec{\lambda}_v^{(k)}(G)$ and $\vec{\lambda}_e^{(k)}(G)$ be as defined in Definition 5.6.1. There exists a constant C such that*

$$\vec{\lambda}_v^{\big(C\cdot\log(\Delta\cdot\alpha(G))\big)}(G) \lesssim \vec{\lambda}_2^{v*}(G) \quad \text{and} \quad \vec{\lambda}_e^{\big(C\cdot\log\alpha(G)\big)}(G) \lesssim \vec{\lambda}_2^{e*}(G).$$

Proof. The proofs of the two inequalities are essentially the same, and we explain the proof of the second inequality here.

[4] We have an example showing that the random projection algorithm in Section 4.3.2 will lose a factor of $\log\alpha(G)$, even when the maximum degree is constant.

Let $G = (V, E)$ be a directed graph and $f : V \to \mathbb{R}^n$ be an optimal embedding of the vertices in G such that $\nu_f^{e*}(G) = \vec{\lambda}_2^{e*}(G)$. Let $\bar{f} : V \to \mathbb{R}^k$ be obtained from f via Gaussian projection. We would like to use $\bar{f}$ as a solution to $\vec{\lambda}_e^{(k)}(G)$. First, note that $\bar{f}$ is obtained by applying a (random) linear operator to f, and so the $\sum_{v \in V} \deg_w(v) \cdot f(v) = 0$ constraint in the SDP in Proposition 5.4.2 is also satisfied by $\bar{f}$. But the normalization constraint $\sum_{v \in V} \deg_w(v) \cdot \|\bar{f}(v)\|^2 = 1$ may not be satisfied, and the objective value $\nu_{\bar{f}}^{e*}(G) = \max_{A \geq 0} \frac{1}{2} \sum_{uv \in E} A(u,v) \cdot \|\bar{f}(u) - \bar{f}(v)\|^2$ may be bigger than $\nu_f^{e*}(G)$. Our plan is to prove that

$$\vec{\lambda}_e^{(k)}(G) \leq \frac{\nu_{\bar{f}}^{e*}(G)}{\sum_{v \in V} \deg_w(v) \|\bar{f}(v)\|^2} \lesssim \frac{\nu_f^{e*}(G)}{\sum_{v \in V} \deg_w(v) \|f(v)\|^2} = \vec{\lambda}_2^{e*}(G), \tag{5.2}$$

when the dimension $k \geq C \cdot \log \alpha(G)$ for some large enough constant C, and this would imply that a scaled version of $\bar{f}$ will satisfy the constraint with objective value at most $O\big(\vec{\lambda}_2^{e*}(G)\big)$.

The main job is to bound $\nu_{\bar{f}}^{e*}(G)$, for which we use the arguments in Theorem 3.2.14. Given $\bar{f} : V \to \mathbb{R}^k$, let $\mathcal{B} = \{uv \in E \mid \|\bar{f}(u) - \bar{f}(v)\|^2 \geq e^{2\varepsilon} \cdot \|f(u) - f(v)\|^2\}$ be the set of "bad edges" where the projected length is considerably longer than the original length. We can bound $\nu_{\bar{f}}^{e*}(G)$ in terms of the edges in $\mathcal{B}$ as follows. For any Eulerian subgraph A that satisfies the constraints in Proposition 5.4.2, twice its objective value is

$$\begin{aligned} & \sum_{uv \notin \mathcal{B}} A(u,v) \|\bar{f}(u) - \bar{f}(v)\|^2 + \sum_{uv \in \mathcal{B}} A(u,v) \|\bar{f}(u) - \bar{f}(v)\|^2 \\ \leq \quad & 4e^{2\varepsilon} \nu_f^{e*}(G) + \sum_{uv \in \mathcal{B}} w(uv) \big(\|\bar{f}(u) - \bar{f}(v)\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 \big) \end{aligned}$$

by essentially the same arguments as in Theorem 3.2.14. Since the upper bound on the last line no longer depends on A, it follows that

$$\begin{aligned} \mathbb{E}_{\bar{f}}[2\nu_{\bar{f}}^{e*}(G)] \quad & \leq \quad 4e^{2\varepsilon} \nu_f^{e*}(G) + \mathbb{E}_{\bar{f}} \Big[\sum_{uv \in \mathcal{B}} w(uv) \big(\|\bar{f}(u) - \bar{f}(v)\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 \big) \Big] \\ & = \quad 4e^{2\varepsilon} \nu_f^{e*}(G) + \sum_{uv \in E} w(uv) \cdot \mathbb{E}_{\bar{f}} \big[\mathbb{1}_{\mathcal{E}_{u,v}} \big(\|\bar{f}(u) - \bar{f}(v)\|^2 - e^{2\varepsilon} \|f(u) - f(v)\|^2 \big) \big] \\ & \leq \quad 4e^{2\varepsilon} \nu_f^{e*}(G) + e^{-c\varepsilon^2 k} \sum_{uv \in E} w(uv) \|f(u) - f(v)\|^2 \\ & \leq \quad 4e^{2\varepsilon} \nu_f^{e*}(G) + 2e^{-c\varepsilon^2 k} \cdot \alpha(G) \cdot \nu_f^{e*}(G), \end{aligned}$$

where the second last inequality is by the second property in Lemma 3.2.13, and the last inequality is by the large optimal property in Lemma 5.5.5. By choosing some constant $\varepsilon \leq 1/4$ and $k \gtrsim \frac{1}{c\varepsilon^2} \log \alpha(G)$, it follows that

$$\mathbb{E}_{\bar{f}}[\nu_{\bar{f}}^{e*}(G)] \leq 4\big(e^{2\varepsilon} + e^{-c\varepsilon^2 k}\alpha(G)\big) \cdot \nu_f^{e*}(G) \lesssim \nu_f^{e*}(G).$$

Finally, the same proof as in Theorem 3.2.14 yields the denominator lower bound that, using the same choice of ε and k, with probability at least 9/10 we have

$$\sum_{v \in V} \deg_w(v) \cdot \left\|\bar{f}(v)\right\|^2 \geq e^{-2\varepsilon}\big(1 - 10e^{-c\varepsilon^2 k}\big) \sum_{v \in V} \deg_w(v) \left\|f(v)\right\|^2 \gtrsim \sum_{v \in V} \deg_w(v) \left\|f(v)\right\|^2.$$

Therefore, (5.2) follows by combining the upper bound on the numerator and this lower bound on the denominator.

The proof of the first inequality is the same, with $\deg_w(v)$ replaced by $\pi(v)$, $w(uv)$ replaced by $w_\pi(uv)$, and with $\alpha(G)$ in the large optimal property replaced by $\Delta \cdot \alpha(G)$ as stated in Lemma 5.5.5. □

By choosing the best coordinate from a k-dimensional embedding, one can achieve the following bound. The proof is standard and is omitted; see [OZ22, Proposition 2.9].

Lemma 5.6.5 (One Dimensional Solution from k-Dimensional Solution)**.** *Let $\vec{\lambda}_v^{(k)}(G)$ and $\vec{\lambda}_e^{(k)}(G)$ be as defined in Definition 5.6.1. Then*

$$\vec{\lambda}_v^{(1)}(G) \leq k \cdot \vec{\lambda}_v^{(k)}(G) \quad \text{and} \quad \vec{\lambda}_e^{(1)}(G) \leq k \cdot \vec{\lambda}_e^{(k)}(G)$$

Theorem 5.6.2 follows immediately from Theorem 5.6.4 and Lemma 5.6.5.

5.7 Rounding Algorithms

The main goal in this section is to show how to find a set of small directed vertex expansion (respectively directed edge conductance) from a solution to $\vec{\lambda}_v^{(1)}(G)$ (respectively $\vec{\lambda}_e^{(1)}(G)$).

Theorem 5.7.1 (Rounding One Dimensional Solution)**.** *For any vertex-weighted directed graph $G = (V, E, \pi)$,*

$$\vec{\psi}(G) \lesssim \sqrt{\vec{\lambda}_v^{(1)}(G)}.$$

For any edge-weighted directed graph $G = (V, E, w)$,

$$\vec{\phi}(S) \lesssim \sqrt{\vec{\lambda}_e^{(1)}(G)}.$$

Assuming Theorem 5.7.1, we can complete the proofs of the two main results.

Proof of Theorem 5.1.4 and Theorem 5.1.8. The easy directions are proved in Proposition 5.3.1 and Proposition 5.3.2. For the hard directions, first we solve the semidefinite programs for $\vec{\lambda}_2^{v*}(G)$ in Proposition 5.4.1 and $\vec{\lambda}_2^{e*}(G)$ in Proposition 5.4.2. Then, we use the dimension reduction result in Theorem 5.6.2 to obtain 1-dimensional solutions to the semidefinite programs with $\vec{\lambda}_v^{(1)}(G) \lesssim \log(\Delta \cdot \alpha(G)) \cdot \vec{\lambda}_2^{v*}(G)$ and $\vec{\lambda}_e^{(1)}(G) \lesssim \log \alpha(G) \cdot \vec{\lambda}_2^{e*}(G)$. Then, we apply the rounding result in Theorem 5.7.1 to establish that

$$\vec{\psi}(G) \lesssim \sqrt{\log(\Delta \cdot \alpha(G)) \cdot \vec{\lambda}_2^{v*}(G)} \quad \text{and} \quad \vec{\phi}(G) \lesssim \sqrt{\log \alpha(G) \cdot \vec{\lambda}_2^{e*}(G)}. \tag{5.3}$$

Finally, we use the inequality $\alpha(G) \lesssim \Delta / \vec{\psi}(G)$ in Lemma 5.5.2 and $\alpha(G) \leq 1/\vec{\phi}(G)$ in Lemma 5.5.1 to obtain the final forms in Theorem 5.1.4 and Theorem 5.1.8. □

We remark that all the steps in the proofs of the two main results can be implemented in polynomial time, and so these give efficient "spectral" algorithms to find a set of small directed vertex expansion or small directed edge conductance.

5.7.1 Proof Structure and Auxiliary Programs

Again, we follow the two-step proof structure in Theorem 3.1.1. We first obtain a solution to the following ℓ_1 versions of $\vec{\lambda}_v^{(1)}$ and $\vec{\lambda}_e^{(1)}$.

Definition 5.7.2 (ℓ_1 Version of $\vec{\lambda}_v^{(1)}$)**.** *Given a directed graph $G = (V, E)$ with vertex weights $\pi : V \to \mathbb{R}^+$, let*

$$\begin{aligned}
\eta_v(G) := \min_{f:V\to\mathbb{R}} \max_{A \geq 0} \quad & \frac{1}{2}\sum_{uv\in E} A(u,v) \cdot |f(u) - f(v)| \\
\text{subject to} \quad & A(u,v) = 0 && \forall uv \notin E \\
& \sum_{u\in V} A(u,v) = \sum_{u\in V} A(v,u) && \forall v \in V \\
& \sum_{u\in V} A(u,v) = \pi(v) && \forall v \in V \\
& \sum_{v\in V} \pi(v) f(v) = 0 \\
& \sum_{v\in V} \pi(v) |f(v)| = 1.
\end{aligned}$$

Definition 5.7.3 (ℓ_1 Version of $\vec{\lambda}_e^{(1)}$). *Given a weighted directed graph $G = (V, E, w)$, let*

$$\begin{aligned}
\eta_e(G) := \min_{f:V\to\mathbb{R}} \max_{A\geq 0} \quad & \frac{1}{2}\sum_{uv\in E} A(u,v)\cdot |f(u)-f(v)| & \\
\text{subject to} \quad & A(u,v) = 0 & \forall uv \notin E \\
& \sum_{v\in V} A(u,v) = \sum_{v\in V} A(v,u) & \forall u \in V \\
& A(u,u) \leq w(uv) & \forall uv \in E \\
& \sum_{v\in V} \deg_w(v) f(v) = 0 & \\
& \sum_{v\in V} \deg_w(v)|f(v)| = 1. &
\end{aligned}$$

We will prove in Section 5.7.2 that there is a square root loss by going from ℓ_2^2 to ℓ_1.

Proposition 5.7.4 (Reductions from ℓ_2^2 to ℓ_1). *For any vertex-weighted directed graph $G = (V, E, \pi)$,*

$$\eta_v(G) \lesssim \sqrt{\vec{\lambda}_v^{(1)}(G)}.$$

For any edge-weighted directed graph $G = (V, E, w)$,

$$\eta_e(G) \lesssim \sqrt{\vec{\lambda}_e^{(1)}(G)}.$$

For threshold rounding, we construct the duals of $\eta_v(G)$ and $\eta_e(G)$ using linear programming duality in the inner maximization problems.

Lemma 5.7.5 (Dual Program of $\eta_v(G)$). *Given a vertex-weighted directed graph $G = (V, E, \pi)$, let*

$$\begin{aligned}
\xi_v(G) := \min_{f:V\to\mathbb{R}} \min_{\substack{q:V\to\mathbb{R}_{\geq 0}\\ r:V\to\mathbb{R}}} \quad & \sum_{v\in V} \pi(v) q(v) & \\
\text{subject to} \quad & q(v) \geq |f(u)-f(v)| - r(u) + r(v) & \forall uv \in E \\
& \sum_{v\in V} \pi(v) f(v) = 0 & \\
& \sum_{v\in V} \pi(v)|f(v)| = 1. &
\end{aligned}$$

Then $\xi_v(G) = 2\eta_v(G)$.

Proof. To write the dual program, we consider the equivalent program of $\vec{\lambda}_v^{(1)}(G)$, where we remove the self-loops and replace the constraint $\sum_{v \in V} A(v,u) = \pi(u)$ by $\sum_{v \in V} A(v,u) \leq \pi(u)$. Then we multiply the objective of $\eta_v(G)$ by a factor of 2 (to avoid the factor 1/2 carrying around). Then we associate a dual variable $q(u) \geq 0$ to each constraint $\sum_{v \in V} A(v,u) \leq \pi(u)$, and a dual variable $r(u)$ to each constraint $\sum_{v \in V} A(u,v) = \sum_{v \in V} A(v,u)$. The result follows from standard linear programming duality. □

The dual program of $\eta_e(G)$ is constructed in the same way and the proof is omitted.

Lemma 5.7.6 (Dual Program of $\eta_e(G)$)**.** *Given an edge-weighted directed graph $G = (V, E, w)$, let*

$$
\begin{aligned}
\xi_e(G) := \min_{f:V \to \mathbb{R}} \min_{\substack{q:E \to \mathbb{R}_{\geq 0} \\ r:V \to \mathbb{R}}} \quad & \sum_{uv \in E} w(uv) q(uv) \\
\text{subject to} \quad & q(uv) \geq |f(u) - f(v)| - r(u) + r(v) \qquad \forall uv \in E \\
& \sum_{v \in V} \deg_w(v) f(v) = 0 \\
& \sum_{v \in V} \deg_w(v) |f(v)| = 1.
\end{aligned}
$$

Then $\xi_e(G) = 2\eta_e(G)$.

In Section 5.7.3, we will present a threshold rounding algorithm to return a set of small directed vertex expansion (respectively directed edge conductance) from a solution to $\xi_v(G)$ (respectively $\xi_e(G)$), with only a constant factor loss.

Proposition 5.7.7 (Threshold Rounding)**.** *For any vertex-weighted directed graph $G = (V, E, \pi)$,*

$$\vec{\psi}(G) \lesssim \xi_v(G).$$

For any edge-weighted directed graph $G = (V, E, w)$,

$$\vec{\phi}(G) \lesssim \xi_e(G).$$

Note that Theorem 5.7.1 follows immediately from Proposition 5.7.4 and Proposition 5.7.7, so it remains to prove the two propositions in Section 5.7.2 and Section 5.7.3.

5.7.2 Step 1 (ℓ_2^2 to ℓ_1)

We first prove the first inequality in Proposition 5.7.4 about directed vertex expansion. Let $G = (V, E)$ be a directed graph with vertex measure $\pi : V \to \mathbb{R}^+$. Let $f : V \to \mathbb{R}$ be a solution to $\vec{\lambda}_v^{(1)}(G)$ with objective value λ_f, with A being an optimal solution to the inner maximization problem (which can be computed by linear programming). Our goal is to construct a solution to $\eta_v(G)$ in Definition 5.7.2 with objective value $O\big(\sqrt{\lambda_f}\big)$.

To this end, define $h : V \to \mathbb{R}$ by

$$h(u) := \begin{cases} (f(u) - c)^2 & \text{if } f(u) > c \\ -(f(u) - c)^2 & \text{otherwise ,} \end{cases}$$

where $c \in \mathbb{R}$ is chosen so as to satisfy the constraint $\sum_{v \in V} \pi(v) h(v) = 0$ in Definition 5.7.2. Note that such c exists and is unique.

Using $\sum_{v \in V} \pi(u) f(u) = 0$, it follows that

$$\sum_{v \in V} \pi(v)|h(v)| = \sum_{v \in V} \pi(v)(f(v) - c)^2 \geq \sum_{v \in V} \pi(v) f(v)^2 = 1, \tag{5.4}$$

and later we shall scale down h to satisfy the constraint $\sum_{v \in V} \pi(v)|h(v)| = 1$.[5] Now we bound the objective value of the ℓ_1 program in Definition 5.7.2 using g as a solution. Let B be an optimal solution to the inner maximization problem in Definition 5.7.2 after fixing g. Note that (3.2) in Theorem 3.1.1 holds for any $c \in \mathbb{R}$, and so in our case it gives the inequality $|h(u) - h(v)| \leq |f(u) - f(v)|(|f(u) - c| + |f(v) - c|)$ for all $u, v \in V$. The

[5] Compared to the original proof in [LTW23], here we do not need to show that $\sum_{v \in V} \pi(v)|h(v)| \leq 2$.

objective value to the ℓ_1 program is then

$$\begin{aligned}
& \frac{1}{2}\sum_{uv\in E} B(u,v)\cdot |h(u)-h(v)| \\
\leq \;& \frac{1}{2}\sum_{uv\in E} B(u,v)|f(u)-f(v)|(|f(u)-c|+|f(v)-c|) \\
\overset{(*)}{\leq} \;& \frac{1}{2}\sqrt{\sum_{uv\in E} B(u,v)\cdot (f(u)-f(v))^2 \cdot 2\sum_{uv\in E} B(u,v)((f(u)-c)^2+(f(v)-c)^2)} \\
= \;& \sqrt{\sum_{uv\in E} B(u,v)(f(u)-f(v))^2}\cdot \sqrt{\frac{1}{2}\sum_{v\in V}(f(v)-c)^2\cdot \Big(\sum_{u:uv\in E} B(u,v)+\sum_{u:vu\in E} B(v,u)\Big)} \\
\leq \;& \sqrt{\sum_{uv\in E} A(u,v)(f(u)-f(v))^2}\cdot \sqrt{\sum_{u\in V}\pi(v)(f(v)-c)^2},
\end{aligned}$$

where the step $(*)$ uses Cauchy-Schwarz inequality and $(a+b)^2 \leq 2(a^2+b^2)$, and the last inequality uses the optimality of A to the inner maximization problem for f. Dividing both sides by $\sum_{v\in V}\pi(v)|h(v)|$ and using (5.4) from above, we obtain

$$\frac{\sum_{uv\in E} B(u,v)|h(u)-h(v)|}{\sum_{v\in V}\pi(v)|h(v)|} \lesssim \sqrt{\lambda_f},$$

and so a scaled-down version of h would be a feasible solution to $\eta_v(G)$ with objective value $O(\sqrt{\lambda_f})$.

The proof of the second inequality about directed edge conductance is essentially the same (with $\pi(v)$ replaced by $\deg_w(v)$) and is omitted.

5.7.3 Step 2 (Threshold Rounding)

Finally, we prove Proposition 5.7.7. Again, we first prove the first inequality in Proposition 5.7.7 about directed vertex expansion. Let $G=(V,E,\pi)$ be a vertex-weighted directed graph. Let (f,q,r) be a feasible solution to $\xi_v(G)$ in Lemma 5.7.5 with objective value ξ_f. Our goal is to construct a nonempty set $S\subset V$ with $\psi(S)\lesssim \xi_f$ using threshold rounding.

In previous threshold rounding algorithms for Cheeger-type inequalities, only the embedding function $f: V\to\mathbb{R}$ is used to produce the output set, so in particular only one ordering of the vertices is considered. The new twist in our algorithm is that we would consider a few candidate orderings of the vertices. These orderings will all ensure that the threshold rounding would produce a set with small expected directed vertex boundary, and we will choose the one that gives large expected set size. To this end, define the following four functions:

- $g_1(u) := \max\{0, f(u) + r(u) - c_1\}$
- $g_2(u) := \max\{0, f(u) - r(u) - c_2\}$
- $g_3(u) := \max\{0, -f(u) + r(u) + c_2\}$
- $g_4(u) := \max\{0, -f(u) - r(u) + c_1\}$,

where c_1 is a π-weighted median of $f(u) + r(u)$, so that

$$\max(\pi(\text{supp}(g_1)), \pi(\text{supp}(g_4))) \leq \pi(V)/2.$$

Similarly, c_2 is a π-weighted median of $f(u) - r(u)$, so that

$$\max(\pi(\text{supp}(g_2)), \pi(\text{supp}(g_3))) \leq \pi(V)/2.$$

Numerator: We bound the size of the outer boundary of either S_t or S_t^c for uniformly random t, depending on whether the coefficient of $r(u)$ is -1 or $+1$ in the function g_i.

On the one hand, if we consider g_1 (similar for g_3), then we would bound the expected outer boundary size of S_t^c as:

$$\begin{aligned}
\int_0^\infty \pi(\partial^+(S_t^c))\, dt &= \sum_{v \in V} \pi(v) \int_0^\infty \mathbb{1}[v \in \partial^+(S_t^c)]\, dt \\
&= \sum_{v \in V} \pi(v) \int_0^\infty \mathbb{1}[\exists\, u \text{ with } uv \in E \text{ and } g_1(u) \leq t < g_1(v)]\, dt \\
&= \sum_{v \in V} \pi(v) \max_{u: uv \in E} \{g_1(v) - g_1(u)\} \\
&\leq \sum_{v \in V} \pi(v) \max_{u: uv \in E} \{(f(v) + r(v)) - (f(u) + r(u))\} \\
&\leq \sum_{v \in V} \pi(v) \max_{u: uv \in E} \{|f(u) - f(v)| + r(v) - r(u)\} \\
&\leq \sum_{v \in V} \pi(v) q(v).
\end{aligned}$$

On the other hand, if we consider the function g_2 (similar for g_4), then we bound the expected outer boundary size of S_t as

$$\begin{aligned}
\int_0^\infty \pi(\partial^+(S_t))\,dt &= \sum_v \pi(v) \int_0^\infty \mathbb{1}[v \in \partial^+(S_t)]\,dt \\
&= \sum_{v\in V} \pi(v) \int_0^\infty \mathbb{1}[\exists\, u \text{ with } uv \in E \text{ and } g_2(v) \le t < g_2(u)]\,dt \\
&= \sum_{v\in V} \pi(v) \max_{u:uv\in E}\{g_2(u) - g_2(v)\} \\
&\le \sum_{v\in V} \pi(v) \max_{u:uv\in E}\{(f(u) - r(u)) - (f(v) - r(v))\} \\
&\le \sum_{v\in V} \pi(v) \max_{u:uv\in E}\{|f(v) - f(u)| + r(v) - r(u)\} \\
&\le \sum_{v\in V} \pi(v) \cdot q(v).
\end{aligned}$$

To summarize, when we do threshold rounding with respect to any of g_1, g_2, g_3, g_4, it holds that

$$\int_0^\infty \min\left\{\pi(\partial^+(S_t)), \pi(\partial^+(S_t^c))\right\} dt \le \sum_{v\in V} \pi(v) q(v).$$

Denominator: For the function g_i, the expected size of S_t is given by

$$\int_0^\infty \pi(S_t)\,dt = \sum_{u\in V} \pi(u) \int_0^\infty \mathbb{1}[g_i(u) > t]\,dt = \sum_{u\in V} \pi(u) \cdot g_i(u).$$

Therefore, our goal is to show that there exists $1 \le i \le 4$ with $\sum_{u\in V} \pi(u) g_i(u) \ge \Omega(1)$. To do so, we will show that

$$\sum_{i=1}^4 \sum_{u\in V} \pi(u) \cdot g_i(u) \ge \Omega(1).$$

Note that, for any $u \in V$,

$$g_1(u) + g_4(u) = \max\{0, f(u) + r(u) - c_1\} + \max\{0, -f(u) - r(u) + c_1\} = |(f(u) + r(u)) - c_1|,$$

and

$$g_2(u) + g_3(u) = \max\{0, f(u) - r(u) - c_2\} + \max\{0, -f(u) + r(u) + c_2\} = |(f(u) - r(u)) - c_2|.$$

Thus it suffices to show that

$$\sum_{u \in V} \pi(u)\Big(\big|(f(u)+r(u))-c_1\big|+\big|(f(u)-r(u))-c_2\big|\Big) \geq \frac{1}{2}.$$

To this end, we note that either $\sum_{u \in V} \pi(u)|f(u)+r(u)| \geq 1$ or $\sum_{u \in V} \pi(u)|f(u)-r(u)| \geq 1$, because

$$\begin{aligned}
\sum_{u \in V} \pi(u)\left(|f(u)+r(u)|+|f(u)-r(u)|\right) &= \sum_{u \in V} \pi(u) \cdot 2\max(|f(u)|,|r(u)|) \\
&\geq 2\sum_{u \in V} \pi(u)|f(u)| = 2.
\end{aligned}$$

Assume without loss that $\sum_{u \in V} \pi(u)r(u) = 0$ (as we can shift every $r(u)$ by the same amount without changing anything). Then both

$$\sum_{u \in V} \pi(u)(f(u)+r(u)) = 0 \quad \text{and} \quad \sum_{u \in V} \pi(u)(f(u)-r(u)) = 0.$$

Consider first the case where $\sum_{u \in V} \pi(u)|f(u)+r(u)| \geq 1$; the other case is treated similarly. Then, since $\sum_{u \in V} \pi(u)(f(u)+r(u)) = 0$ and $\sum_{u \in V} \pi(u)|f(u)+r(u)| \geq 1$, it follows that

$$\sum_{u: f(u)+r(u) \leq 0} \pi(u)|f(u)+r(u)| = \sum_{u: f(u)+r(u) \geq 0} \pi(u)|f(u)+r(u)| = \frac{1}{2}\sum_{u} \pi(u)|f(u)+r(u)| \geq \frac{1}{2}.$$

If $c_1 \geq 0$, then

$$\begin{aligned}
\sum_{u \in V} \pi(u)|(f(u)+r(u))-c_1| &\geq \sum_{u: f(u)+r(u) \leq 0} \pi(u)|(f(u)+r(u))-c_1| \\
&\geq \sum_{u: f(u)+r(u) \leq 0} \pi(u)|f(u)+r(u)| \geq \frac{1}{2},
\end{aligned}$$

and similarly if $c_1 < 0$, then

$$\begin{aligned}
\sum_{u \in V} \pi(u)|(f(u)+r(u))-c_1| &\geq \sum_{u: f(u)+r(u) \geq 0} \pi(u)|(f(u)+r(u))-c_1| \\
&\geq \sum_{u: f(u)+r(u) \geq 0} \pi(u)|f(u)+r(u)| \geq \frac{1}{2}.
\end{aligned}$$

To summarize,

$$\sum_{i=1}^{4}\sum_{u\in V}\pi(u)g_i(u) = \sum_{u\in V}\pi(u)\Big(\big|(f(u)+r(u))-c_1\big| + \big|(f(u)-r(u))-c_2\big|\Big) \geq \frac{1}{2}.$$

Conclusion: There exists $g = g_i$ for some $1 \leq i \leq 4$, such that if we use this function for threshold rounding,

- $\int_0^\infty \min\left\{\pi(\partial^+(S_t)), \pi(\partial^+(S_t^c))\right\} dt \leq \sum_{v\in V}\pi(v)q(v) = \xi_f$;
- $\int_0^\infty \pi(S_t)\, dt \geq 1/8$;
- $\pi(S_t) \leq \pi(V)/2$ always.

Hence, we can return some $S = S_t$, whence $0 < \pi(S) \leq \pi(V)/2$ and

$$\vec{\psi}(S) = \frac{\min\left\{\pi(\partial^+(S)), \pi(\partial^+(S^c))\right\}}{\min\left\{\pi(S), \pi(S^c)\right\}} = \frac{\min(\pi(\partial^+(S)), \pi(\partial^+(S^c)))}{\pi(S)} \leq 8\xi_f.$$

The proof of the second inequality about directed edge conductance is the same (after replacing the numerator $\sum_{v\in V}\pi(v)q(v)$ by $\sum_{uv\in E} w(uv)q(uv)$ and the denominator $\sum_{v\in V}\pi(v)|f(v)|$ by $\sum_{v\in V}\deg_w(v)|f(v)|$) and is omitted.

5.8 Fastest Mixing Time

The goal of this section is to prove Theorem 5.1.6 that

$$\frac{1}{\vec{\psi}(G)} \lesssim \tau^*_{\text{mix}}(G) \lesssim \frac{1}{\vec{\psi}(G)^2} \cdot \log\frac{\Delta}{\vec{\psi}(G)} \cdot \log\frac{1}{\pi_{\min}}.$$

There are two parts to the proof. In the first part, we upper bound the fastest mixing time using Theorem 3.3.2 by Fill [Fil91] and Chung [Chu05]. In the second part, we lower bound the fastest mixing time using a combinatorial argument.

Proof of Theorem 5.1.6. Recall that in the setting of the theorem, π is not only a weight function, but a probability distribution. We assume the graph is strongly connected and so $\vec{\lambda}_2^{v*}(G) > 0$.

To prove the upper bound, we apply Theorem 3.3.2 to prove that

$$\tau^*_{\text{mix}}(G) \lesssim \left(\vec{\lambda}_2^{v*}(G)\right)^{-1} \cdot \log(\pi_{\min}^{-1}),$$

and then the result will follow from Theorem 5.1.4. Let A be an optimal reweighted Eulerian subgraph in Definition 5.1.3. Let $P := \Pi^{-1}A$ be the transition matrix of the ordinary random walk corresponding to the reweighted subgraph A. Observe that $P := \Pi^{-1}A$ is a feasible solution to Definition 5.1.5, and so is $(I+P)/2$. Therefore, by Theorem 3.3.2,

$$\tau^*_{\text{mix}}(G) \leq \tau_{\text{mix}}\left(\frac{I+P}{2}\right) \lesssim \frac{1}{\lambda_2(\tilde{\mathcal{L}}(G))} \cdot \log\left(\frac{1}{\pi_{\min}}\right) = \frac{1}{\vec{\lambda}_2^{v*}(G)} \cdot \log\left(\frac{1}{\pi_{\min}}\right),$$

where the last inequality is because $\tilde{\mathcal{L}}(G) = I - \Pi^{-\frac{1}{2}}(A + A^T)\Pi^{-\frac{1}{2}}/2$ as defined in (3.8) and $\lambda_2(\tilde{\mathcal{L}}(G)) = \vec{\lambda}_2^{v*}(G)$ by Definition 5.1.1.

To prove the lower bound, we will prove that for any feasible solution P to Definition 5.1.5,

$$\frac{1}{\vec{\psi}(G)} \lesssim \tau_{\text{mix}}(P),$$

which immediately implies the result. This argument is similar to that in [LP17, Theorem 7.4] which lower bounds mixing time using graph conductance. For the argument to work, it is essential that the definition of $\vec{\psi}(G)$ accounts for the minimum of outgoing vertex boundary and incoming vertex boundary. For any such P, consider the graph with arc weights $A(u,v) = \pi(u)P(u,v)$. Then,

$$\sum_{u \in V} \pi(u)P(u,v) = \pi(v) = \sum_{u \in V} \pi(v)P(v,u) \quad\implies\quad \sum_{u \in V} A(u,v) = \sum_{u \in V} A(v,u),$$

so that A is Eulerian. Consider a nonempty subset $S \subset V$ such that $\vec{\psi}(S) = \vec{\psi}(G)$. We will use S to define an initial distribution $p_0 : V \to \mathbb{R}_{\geq 0}$ such that

$$d_{TV}(p_t, \pi) = \frac{1}{2}\sum_{v \in V} |p_t(v) - \pi(v)| > \frac{1}{e}$$

for any $t \leq 1/(8\vec{\psi}(S))$, and it will follow that $\tau_{\text{mix}}(P) > 1/(8\vec{\psi}(S))$. Here, $p_t^T = p_0^T P^t$ is the distribution after t steps. Without loss of generality assume that $\pi(S) \leq 1/2$. We define

$$p_0(u) = \begin{cases} \pi(u)/\pi(S), \text{ if } u \in S; \\ 0, \text{ otherwise.} \end{cases}$$

By induction, we can show that $p_t(v) \leq \pi(v)/\pi(S)$ for all $t \geq 0$; indeed, for any $v \in V$,

$$p_{t+1}(v) = \sum_{u \in V} p_t(u) \cdot P(u,v) \leq \sum_{u \in V} \frac{\pi(u)}{\pi(S)} \cdot P(u,v) = \frac{\pi(v)}{\pi(S)}. \tag{5.5}$$

There are two cases to consider.

1. $\pi(\partial^+(S)) \leq \pi(\partial^+(S^c))$. In this case, we will show that $p_t(S) \geq 1 - t \cdot \vec{\psi}(S)$ for all $t \geq 0$. Indeed by (5.5), at step $t+1$, the total amount of probability mass escaping from S is at most

$$\sum_{v \in \partial^+(S)} p_t(v) \leq \frac{\pi(\partial^+(S))}{\pi(S)} = \vec{\psi}(S).$$

   Hence, for any $t \leq 1/(8\vec{\psi}(S))$, we have $p_t(S) \geq \frac{7}{8}$ and $p_t(S^c) \leq \frac{1}{8}$, so

$$d_{TV}(p_t, \pi) \geq \frac{1}{2}\big(|p_t(S^c) - \pi(S^c)| + |p_t(S) - \pi(S)|\big) \geq \frac{3}{8} > \frac{1}{e}.$$

2. $\pi(\partial^+(S)) > \pi(\partial^+(S^c))$. We will again show that $p_t(S) \geq 1 - t \cdot \vec{\psi}(S)$ for all $t \geq 0$. By (5.5) and using the Eulerian property of $A = \Pi P$, the total amount of probability mass escaping from S is at most

$$\begin{aligned} \sum_{u \in S} \sum_{v \in S^c} p_t(u) P(u,v) &\leq \frac{1}{\pi(S)} \sum_{u \in S} \sum_{v \in S^c} \pi(u) P(u,v) = \frac{1}{\pi(S)} \sum_{u \in S} \sum_{v \in S^c} \pi(v) P(v,u) \\ &= \frac{1}{\pi(S)} \sum_{u \in \partial^+(S^c)} \sum_{v \in S^c} \pi(v) P(v,u) \leq \frac{\pi(\partial^+(S^c))}{\pi(S)} = \vec{\psi}(S). \end{aligned}$$

   By the same reasoning as in the first case, for $t \leq 1/(8\vec{\psi}(S))$ we have

$$d_{TV}(p_t, \pi) \geq \frac{3}{8} > \frac{1}{e}.$$

This completes the proof of the lower bound and hence Theorem 5.1.6. □

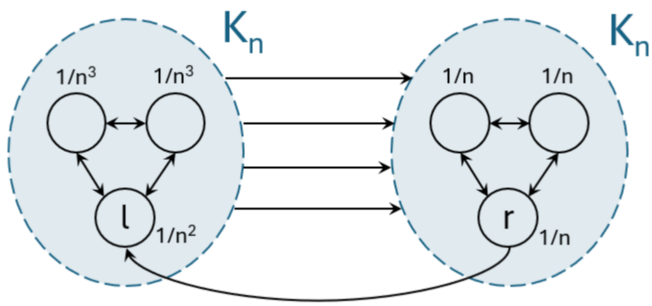
5.9 Relations with Previous Work

In this section, we show some examples where the Cheeger constant [Fil91, Chu05] is very different from directed edge conductance and directed vertex expansion, and relate the semidefinite program in [ACMM05] to the one for reweighted eigenvalue in Definition 5.1.7.

5.9.1 Cheeger Constant, Edge Conductance, and Vertex Expansion

We show two examples. In the first example, the Cheeger constant in (3.9) in Section 3.3.1 is large while the directed edge conductance and directed vertex expansion is small.

Example 5.9.1 (Large Cheeger Constant but Small Edge Conductance and Vertex Expansion)**.** *Consider the directed graph shown in the figure. Both L and R are cliques of size n. There is an arc from every vertex in L to every vertex in R. There is an arc (r, l) from a special vertex $r \in R$ to a special vertex $l \in L$.*



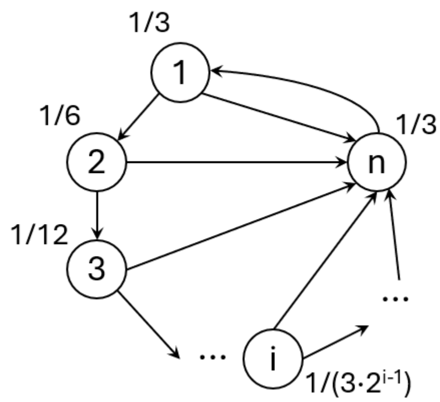
When the graph G has the same weight on each arc and the same weight on each vertex, it is clear that $\vec{\phi}(G) \leq 1/n^2$ and $\vec{\psi}(G) \leq 1/n$ as there is only one arc from R to L.

We claim that the Cheeger constant $h(G)$ in (3.9) is $\Omega(1)$. The reason is that the Cheeger constant is normalized by the probabilities in the stationary distribution π, and this will make L to have small π-weight and so both $h(L)$ and $h(R)$ become big after the normalization. More precisely, after some calculations that are omitted, we have $\pi(v) \approx 1/n$ for every vertex $v \in R$, $\pi(u) \approx 1/n^3$ for every vertex $u \in L - \{l\}$, and $\pi(l) \approx 1/n^2$. This implies that $h(L) = h(R) = \Omega(1)$, and indeed $h(G) = \Omega(1)$ after a case analysis which we omit.

This example shows that the edge conductance of the reweighted subgraph with respect to the stationary distribution does not provide a good approximation to directed edge conductance and directed vertex expansion, while an optimal reweighted subgraph does identify the bottlenecks in the directed graph.

In the second example, the k-way Cheeger constant is large but the k-way directed edge conductance is small.

Example 5.9.2 (Large k-Way Cheeger Constant but Small k-Way Edge Conductance)**.** *Let G be a directed cycle over the vertex set $[n]$. For each $i \in \{2, 3, ...n-2\}$ we add an extra edge (i, n). The figure of the graph is shown with the stationary distribution of the ordinary*



random walk on the graph where every edge has the same weight. In this example, the k-way directed edge conductance is k/n, but the graph has Cheeger constant $\Omega(1)$ because the vertices $\{2, 3, ...n-1\}$ have exponentially decreasing stationary weight.

Since large Cheeger constant implies large $\lambda_2(\tilde{L})$ implies large $\lambda_k(\tilde{L})$ ($\tilde{L}$ is the Laplacian defined by Chung in (3.8)), this example shows that $\lambda_k(\tilde{L})$ is large but the k-way directed edge conductance is small. This rules out the possibility of having a higher-order Cheeger inequality for directed graphs relating $\lambda_k(\tilde{\mathcal{L}})$ to k-way directed edge conductance.

5.9.2 Semidefinite Program for Directed Sparsest Cut

We compare the semidefinite program for $\vec{\lambda}_2^{e*}(G)$ in Proposition 5.4.2 with the semidefinite program for the directed sparsest cut probelm in [ACMM05]. Given an unweighted directed graph $G = (V, E)$, the directed sparsest cut problem is defined as

$$\vec{\varphi}(G) := \min_{S \subseteq V} \frac{\min\left\{|\delta^+(S)|, |\delta^+(S^c)|\right\}}{\min\{|S|, |S^c|\}}.$$

Agarwal, Charikar, Makarychev, and Makarychev [ACMM05] gave a semidefinite program relaxation $\mathsf{sdp}^{\Delta}(G)$ for $\vec{\varphi}(G)$ and proved that $\mathsf{sdp}^{\Delta}(G) \lesssim \vec{\varphi}(G) \lesssim \sqrt{\log n} \cdot \mathsf{sdp}^{\Delta}(G)$; see Section 3.6.3.

We note that $\mathsf{sdp}^{\Delta}(G)$ in (3.20) can be modified slightly to give a similar approximation to the directed edge conductance $\vec{\phi}(G)$ in Definition 5.1.2. Consider the semidefinite program $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$ defined as

$$\begin{aligned} \min_{f:V\cup\{0\}\to\mathbb{R}^n} \quad & \sum_{uv\in E} \left(\|f(u)-f(v)\|^2 - \|f(u)-f(0)\|^2 + \|f(v)-f(0)\|^2 \right) \\ \text{subject to} \quad & \|f(u)-f(v)\|^2 + \|f(v)-f(u')\|^2 \geq \|f(u)-f(u')\|^2 \quad \forall u,v,u' \in V\cup\{0\} \\ & \sum_{v\in V} \deg(v) f(v) = \vec{0} \\ & \sum_{v\in V} \deg(v) \|f(v)\|^2 = 1, \end{aligned}$$

Note that by Fact 2.10.4, the normalization constraint in (3.20) that

$$\sum_{u,v\in V} \|f(u)-f(v)\|^2 = n^2$$

is equivalent to the constraints

$$\sum_{u\in V} f(u) = \vec{0}, \quad \sum_{u\in V} \|f(u)\|^2 = \frac{n}{2}$$

since the remainder of the program is invariant under the translation $f(u) \mapsto f(u) + \zeta$ for any vector $\zeta \in \mathbb{R}^n$. After that, the only difference between $\mathsf{sdp}^{\Delta}(G)$ and $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$ is that the scaling of $f(v)$ by $\deg(v)$ (instead of 1 for sdp_{φ}), which corresponds to the degree weights in the denominator of the directed edge conductance in Definition 5.1.2. We note that a simple modification of the proof in [ACMM05] shows that $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G) \lesssim \vec{\phi}(G) \lesssim \sqrt{\log n} \cdot \mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$. [6]

To our knowledge, it was not known that $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$ can be used to certify whether a directed graph has constant edge conductance as in Theorem 5.1.8, as the analysis using ℓ_2^2 triangle inequalities based on [ARV09] has a $\sqrt{\log n}$ factor loss. However, we observe that the semidefinite program in Proposition 5.4.2 for $\vec{\lambda}_2^{e*}(G)$ is a weaker program than $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$.

Claim 5.9.3 ($\vec{\lambda}_2^{e*}(G)$ and $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$)**.** *For any directed graph $G = (V, E)$, it holds that $\vec{\lambda}_2^{e*}(G) \leq \mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$.*

[6] In fact, using a weighted version of the ARV structure theorem of Theorem 3.6.2 which we shall formulate and prove in Lemma 8.3.2, the results of [ACMM05] extend readily to π-weighted edge expansion on weighted directed graphs $G = (V, E, w)$ for arbitrary vertex measure π.

Proof. Consider the following equivalent characterization of $\vec{\lambda}_2^{e*}(G)$ by using LP duality in the inner maximization problem as in Lemma 5.7.6:

$$\begin{aligned}
\vec{\lambda}_2^{e*}(G) = \min_{f:V\to\mathbb{R}^n} \min_{\substack{q:E\to\mathbb{R}_{\geq 0}\\ r:V\to\mathbb{R}}} \quad & \sum_{uv\in E} q(uv) \\
\text{subject to} \quad & q(uv) \geq \|f(u)-f(v)\|^2 - r(u) + r(v) \qquad \forall uv \in E \\
& \sum_{v\in V} \deg(v) f(v) = \vec{0} \\
& \sum_{v\in V} \deg(v) \|f(v)\|^2 = 1.
\end{aligned}$$

We will show that for every feasible solution $f : V \cup \{0\} \to \mathbb{R}^n$ to $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$, there is a feasible solution $f' : V \to \mathbb{R}^n, q : E \to \mathbb{R}_{\geq 0}, r : V \to \mathbb{R}$ to $\vec{\lambda}_2^{e*}(G)$ with the same objective value. Then the lemma would follow immediately. To this end, define $f'(v) = f(v)$ for $v \in V$, $r(v) = \|f(v) - f(0)\|^2$ for $v \in V$ and $q(uv) = \|f(u) - f(v)\|^2 - r(u) + r(v)$ for $uv \in E$. Clearly, the objective values are equal. Also, we see that the constraints $q(uv) \geq 0$ are satisfied because of the ℓ_2^2 triangle inequalities. □

Therefore, Theorem 5.1.8 and Claim 5.9.3 imply that

$$\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G) \lesssim \vec{\phi}(G) \lesssim \sqrt{\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G) \cdot \log \frac{1}{\vec{\phi}(G)}},$$

where the "easy direction" $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G) \lesssim \vec{\phi}(G)$ follows because $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$ is a relaxation of directed edge conductance $\vec{\phi}(G)$. This provides a new analysis that $\mathsf{sdp}^{\Delta}_{\vec{\phi}}(G)$ can also be used to certify constant edge conductance in directed graphs.

5.10 Generalizations of Cheeger Inequalities for Directed Graphs

For undirected graphs, there are several interesting generalizations of Cheeger's inequality: Trevisan's result that relates λ_n to bipartite edge conductance [Tre09], the higher-order Cheeger's inequality that relates λ_k to k-way edge conductance [LOT12, LRTV12], and the improved Cheeger's inequality that relates λ_2 and λ_k to edge conductance [KLL+13] (see Section 3.1). Using reweighted eigenvalues for vertex expansion, close analogs of these results were obtained in Chapter 4, relating λ_n^* to bipartite vertex expansion, λ_k^* to k-way vertex expansion, and λ_2^* and λ_k^* to vertex expansion.

In this section, we study whether there are close analogs of these results for directed graphs, using reweighted eigenvalues for directed vertex expansion in Definition 5.1.1 and directed edge conductance in Definition 5.1.2. Perhaps surprisingly, we show that the natural analogs of Trevisan's result and higher-order Cheeger's inequality do not hold, but we obtain analogs of the improved Cheeger's inequality for directed vertex expansion and directed edge conductance.

5.10.1 Higher-Order Cheeger's Inequality for Directed Graphs

We attempt to develop a theory of higher reweighted eigenvalues for directed graphs and obtain directed analogues of the higher-order Cheeger inequalities in Section 3.1.3 and Section 4.5. For a directed graph $G = (V, E)$, we can define $\vec{\lambda}_k^{v*}(G)$ and $\vec{\lambda}_k^{e*}(G)$ as in Definition 5.1.1 and Definition 5.1.2, but with the objective function replaced by maximizing the k-th smallest eigenvalue. It is a basic fact (c.f. Proposition 2.5.6) that, given an undirected graph G, $\lambda_k(G) = 0$ if and only if G has at least k connected components. The following is a directed analogue of this basic fact.

Proposition 5.10.1 (Reweighted Eigenvalues and Strongly Connected Components)**.**

- *For any directed graph $G = (V, E)$ with weight function $w : E \to \mathbb{R}_{\geq 0}$, then $\vec{\lambda}_k^{e*}(G) = 0$ if and only if G has at least k strongly connected components.*
- *For any directed graph $G = (V, E)$ with weight function $\pi : V \to \mathbb{R}_{\geq 0}$, then $\vec{\lambda}_k^{v*}(G) = 0$ if and only if G has at least k strongly connected components.*

Proof. In one direction, assume G has at least k strongly connected components $S_1, \ldots, S_k$. Then, in any Eulerian reweighted subgraph A, we claim that

$$\sum_{uv \in \delta^+(S_i)} A(u, v) = \sum_{uv \in \delta^-(S_i)} A(u, v) = 0$$

for $1 \leq i \leq k$. To see this, suppose to the contrary that $uv \in \delta^+(S_i)$ and $A(u, v) > 0$, then as the edge set of any Eulerian graph can be decomposed into edge disjoint cycles, there must be a directed cycle C with $uv \in C$, but then $S_i \cup C \supseteq S_i \cup \{v\}$ is also strongly connected, contradicting that S_i is a maximally strongly connected subset. Therefore, in the underlying undirected graph defined by $\frac{1}{2}(A + A^T)$, each S_i is a set of conductance zero, and thus $\lambda_k = 0$ by the basic fact. Since this holds for any Eulerian reweighted subgraph A, it follows that $\vec{\lambda}_k^{v*}(G) = \vec{\lambda}_k^{e*}(G) = 0$.

In the other direction, assume G has less than k strongly connected components $S_1, \ldots, S_\ell$ for $\ell < k$. Then, in each strongly connected component S_i, there is an Eulerian reweighting A_i in the induced subgraph of S_i such that S_i is strongly connected. (It is not difficult to see this directly, or one can use Hoffman's result in Lemma 5.5.4.) So, there is an Eulerian reweighting such that the underlying undirected graph G' has at most $\ell < k$ connected components, and thus $\vec{\lambda}_k^{v*}(G), \vec{\lambda}_k^{e*}(G) \geq \lambda_k(G') > 0$ by the basic result. □

One might expect that there is a robust generalization of Proposition 5.10.1 relating $\vec{\lambda}_k^{v*}(G)$ and $\vec{\lambda}_k^{e*}(G)$ to k-way directed vertex expansion and k-way directed edge conductance, just as in the case $k = 2$ in Theorem 5.1.4 and Theorem 5.1.8. But in general, unlike undirected graphs, it is not true that G has at least k strongly connected components if and only if G has at least k disjoint subsets $S_1, \ldots, S_k$ each with directed edge conductance zero or directed vertex expansion zero. Note the subtlety that this is true for $k = 2$, as there is a source component and a sink component with directed edge conductance and directed vertex expansion zero.

Example 5.10.2 (Counterexample to Higher-Order Cheeger's Inequality for Directed Graphs)**.** *Consider the complete directed acyclic graph G where the vertex set is $[n]$ and there is an arc ij for every $i < j$. On the one hand, $\vec{\lambda}_k^{v*}(G) = \vec{\lambda}_k^{e*}(G) = 0$ for every $k \leq n$, as any Eulerian reweighting must have n isolated vertices (with self-loops). On the other hand, for any $k \geq 3$, at least one set has non-zero directed edge conductance. Furthermore, it can be shown that for $k \geq 2\log_2 n$, any k disjoint subsets must contain at least one subset of directed edge conductance at least $1/4$. This provides a strong counterexample that $\vec{\lambda}_k^{e*}(G)$ is small but the k-way directed edge conductance $\vec{\phi}_k(G)$ is large. A similar argument can be made for the case of directed vertex expansion.*

We believe that there is still a robust generalization of Proposition 5.10.1, such that $\vec{\lambda}_k^{v*}(G), \vec{\lambda}_k^{e*}(G)$ is small if and only if there are k disjoint subsets where each is "close" to a strongly connected component. But it is not clear how to formulate closeness to a strongly connected component, as it is a "global" property that cannot be determined by only looking at the edges incident to a subset $S \subseteq V$. On a technical level, we remark that the proofs in Section 3.1.3 and Section 4.5 can be followed to construct k disjointly-supported functions $f_1, \ldots, f_k$ from a solution to $\vec{\lambda}_k^{v*}(G)$ and $\vec{\lambda}_k^{e*}(G)$, such that each f_i has small objective value to $\vec{\lambda}_2^{v*}(G)$ and $\vec{\lambda}_2^{e*}(G)$. However, using the new threshold rounding algorithm for $\vec{\lambda}_2^{v*}(G)$ and $\vec{\lambda}_2^{e*}(G)$ based on $f_i \pm r_i$ in Section 5.7.3, we can no longer conclude that there is a subset S_i of small directed edge conductance or directed vertex expansion in the support of f_i, as the support of $f_i \pm r_i$ could be very different from that of f_i. This also indirectly shows that the new idea of doing threshold rounding on $f \pm r$ is a necessary modification.

Open Problem 5.10.3. *Prove a robust generalization of Proposition 5.10.1.*

5.10.2 Bipartite Cheeger Inequality for Directed Graphs

Another basic result in spectral graph theory is Proposition 2.5.7 that $\lambda_n(G) = 2$ if and only if G has a bipartite component S, or equivalently G has a set S of conductance zero with the induced subgraph $G[S]$ being bipartite. Trevisan [Tre09] proved a robust generalization of this basic result, by proving a Cheeger-type inequality that $\lambda_n(G)$ is close to 2 if and only if G has a set S of small conductance with the induced subgraph $G[S]$ being close to bipartite; see Section 3.1.2.

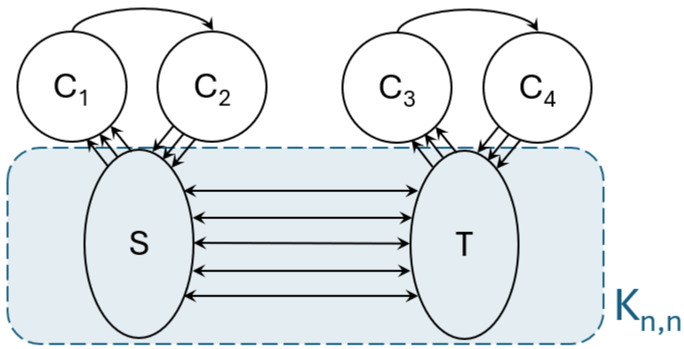
As in Section 5.10.1, we can use the n-th reweighted eigenvalue to prove an analog of the basic result for directed graphs. We omit the proof as it is similar to that in Proposition 5.10.1 and also because it is not used in other results.

Proposition 5.10.4 (Reweighted Eigenvalues and Bipartite Strongly Connected Components)**.**

- *For any directed graph $G = (V, E)$ with weight function $w : E \to \mathbb{R}_{\geq 0}$, $\vec{\lambda}_n^{e*}(G) = 2$ if and only if G has a strongly connected component S such that the induced subgraph $G[S]$ is bipartite.*
- *For any directed graph $G = (V, E)$ with weight function $\pi : V \to \mathbb{R}_{\geq 0}$, $\vec{\lambda}_n^{v*}(G) = 2$ if and only if G has a strongly connected component S such that the induced subgraph $G[S]$ is bipartite.*

As in Section 5.10.1, the natural analogs of Trevisan's result for directed graphs are not true, because the existence of a nearly strongly connected bipartite component does not imply the existence of a set S of small directed edge conductance or directed vertex expansion, with the induced subgraph $G[S]$ being close to bipartite.

Example 5.10.5 (Counterexample to Bipartite Cheeger Inequality for Directed Graphs)**.** *Consider the example shown in the figure below. In this directed graph G, $|S| = |T| = n$*



and $|C_1| = |C_2| = |C_3| = |C_4| = n/2$. Each C_i is a clique, and there is only one edge from C_1 to C_2 and only one edge from C_3 to C_4. The induced subgraph on $S \cup T$ is a complete bipartite graph $K_{n,n}$. Every vertex in S has an edge to every vertex in C_1, and every vertex in C_2 has an edge to every vertex in S. Similarly, every vertex in T has an edge to every vertex in C_3, and every vertex in C_4 has an edge to every vertex in T. Every edge in G has weight one.

On the one hand, because of the bottlenecks from C_1 to C_2 and from C_3 to C_4, any Eulerian reweighing A will have $\sum_{uv:u\in S, v\in C_1} A(u,v) = \sum_{uv:u\in C_2, v\in S} A(u,v) \leq 1$ and $\sum_{uv:u\in T, v\in C_3} A(u,v) = \sum_{uv:u\in C_4, v\in T} A(u,v) \leq 1$. Therefore, $S \cup T$ is an induced bipartite graph with small edge conductance in the underlying undirected graph $\frac{1}{2}(A + A^T)$, and one can use the easy direction of Trevisan's result to show that $\vec{\lambda}_n^{e*}(G) \geq 2 - O(1/n^2)$. On the other hand, $S \cup T$ has large directed edge conductance, and any subset with small directed edge conductance must be far from bipartite because of the edges induced in C_i. We could formally define directed bipartite edge conductance $\vec{\phi}_B(G)$ and show that $\vec{\phi}_B(G) = \Omega(1)$ is large, but we decide to omit these details. To summarize, this gives a strong counterexample where $\vec{\lambda}_n^{e*}(G)$ is very close to 2 but there does not exist any subset S with small directed edge conductance and the induced subgraph $G[S]$ being close to bipartite. A similar argument can be made for the case of directed vertex expansion.

As in Section 5.10.1, we believe that there is a robust generalization of Proposition 5.10.4 that $\vec{\lambda}_n^{e*}(G)$ and $\vec{\lambda}_n^{e*}(G)$ are close to 2 if and only if G has a nearly bipartite strongly connected component. We leave it as an open problem to formulate the combinatorial condition and to prove such a Cheeger-type inequality.

Open Problem 5.10.6. *Formulate and prove a Cheeger-type inequality relating $\vec{\lambda}_n^{e*}(G)$ and $\vec{\lambda}_n^{e*}(G)$ to a combinatorial quantity of a directed graph G.*

5.10.3 Improved Cheeger Inequality for Directed Graphs

Unlike the higher-order and bipartite Cheeger inequalities, we can extend the improved Cheeger inequality [KLL+13] to directed graphs, as this is only about 2-way partitioning (recall the discussion above Example 5.10.2). This potentially can also be used to explain the good empirical performance of the spectral algorithm in Theorem 5.1.8.

Theorem 5.10.7 (Improved Cheeger's inequality for Directed Vertex Expansion)**.** *Let $G = (V, E, \pi)$ be a vertex-weighted directed graph. For any $2 \leq k \leq n/2$,*

$$\vec{\lambda}_2^{v*}(G) \lesssim \vec{\psi}(G) \lesssim \frac{k \cdot \log(\Delta \cdot \alpha(G)) \cdot \vec{\lambda}_2^{v*}(G)}{\sqrt{\vec{\lambda}_k^{v*}(G)}} \lesssim \frac{k \cdot \log(\Delta/\vec{\psi}(G)) \cdot \vec{\lambda}_2^{v*}(G)}{\sqrt{\vec{\lambda}_k^{v*}(G)}}.$$

Theorem 5.10.8 (Improved Cheeger's inequality for Directed Edge Conductance)**.** *Let $G = (V, E, w)$ be an edge-weighted directed graph. For any $2 \leq k \leq n/2$,*

$$\vec{\lambda}_2^{e*}(G) \lesssim \vec{\phi}(G) \lesssim \frac{k \cdot \log \alpha(G) \cdot \vec{\lambda}_2^{e*}(G)}{\sqrt{\vec{\lambda}_k^{e*}(G)}} \lesssim \frac{k \cdot \log(1/\vec{\phi}(G)) \cdot \vec{\lambda}_2^{e*}(G)}{\sqrt{\vec{\lambda}_k^{e*}(G)}}.$$

The proofs of the two results are similar to that the proofs of Theorem 3.1.16 and Theorem 4.1.11 and also similar to each other, so we just provide a sketch of the proof of Theorem 5.10.7 in the following.

Note that $\vec{\lambda}_k^{v*}(G)$ not a convex optimization problem. As in Section 4.6 in the previous chapter, we change the objective in Definition 5.1.3 to maximize the sum of the k smallest eigenvalues $\sum_{i=1}^{k} \lambda_i(\mathcal{L})$, so that we can use Proposition 2.8.3 to write this as a semidefinite program, which we call $\vec{\sigma}_k^{v*}(G)$. Using the same manipulations as in Proposition 4.5.2 and

Proposition 5.4.1, we can write

$$\begin{aligned}
\vec{\sigma}_k^{v*}(G) := \min_{f:V\to\mathbb{R}^n} \max_{A\geq 0} \quad & \frac{1}{2}\sum_{uv\in E} A(u,v)\cdot \|f(u)-f(v)\|^2 \\
\text{subject to} \quad & A(u,v) = 0 && \forall uv \notin E \\
& \sum_{u\in V} A(u,v) = \sum_{u\in V} A(v,u) && \forall v\in V \\
& \sum_{u\in V} A(u,v) = \pi(v) && \forall v \in V \\
& \sum_{v\in V} \pi(v) f(v) f(v)^T \preceq I_n \\
& \sum_{v\in V} \pi(v)\, \|f(v)\|^2 = k.
\end{aligned}$$

The proof will relate $\vec{\lambda}_2^{v*}(G)$ and $\vec{\sigma}_k^{v*}(G)$ to $\vec{\psi}(G)$. We follow the same two-step approach in Section 3.1.4. The first step is to prove that if there is a 1-dimensional solution to $\vec{\lambda}_v^{(1)}(G)$ that is close to a k-step function (i.e. a function with at most k distinct values), then the approximation guarantee of threshold rounding in Section 5.7.3 is improved.

Proposition 5.10.9 (Improved Threshold Rounding)**.** *Let $G = (V, E, \pi)$ be a vertex-weighted directed graph. Given a solution $f : V \to \mathbb{R}$ to $\vec{\lambda}_v^{(1)}(G)$ with objective value λ_f and a k-step function $y_f : V \to \mathbb{R}$ approximating f, it holds that*

$$\vec{\psi}(G) \lesssim \eta_v(G) \lesssim k\cdot\lambda_f + k\,\|f - y_f\|_\pi \sqrt{\lambda_f},$$

where $\eta_v(G)$ is the ℓ_1-version of $\vec{\lambda}_v^{(1)}(G)$ in Definition 5.7.2, and $\|z\|_\pi^2 := \sum_v \pi(v)\cdot z(v)^2$ for any $z : V \to \mathbb{R}$. Note the first inequality is by Proposition 5.7.7.

The second step is to prove that if $\vec{\sigma}_k^{v*}(G)$ is large for a small k, then there is a good k-step approximation to a good solution to $\vec{\lambda}_v^{(1)}(G)$. As in Section 5.7.3, we consider the ℓ_1 dual program $\xi_v(G)$ in Lemma 5.7.5 of $\eta_v(G)$.

Proposition 5.10.10 (Constructing k-Step Approximation)**.** *Let $G = (V, E, \pi)$ be a vertex-weighted directed graph. Given a solution $f : V \to \mathbb{R}$ to $\xi_v(G)$ with objective value ξ_f, there exists a k-step function $y : V \to \mathbb{R}$ with*

$$\|f - y\|_\pi^2 \lesssim \frac{k\cdot\xi_f}{\vec{\sigma}_k^{v*}(G)}.$$

Combining the two propositions with $2k$ in place of k, using $\lambda_f = \xi_f$, applying the dimension reduction result in Theorem 5.6.2, and using the relation $\vec{\sigma}_{2k}^{v*}(G) \geq k \cdot \vec{\lambda}_k^{v*}(G)$[7], we get

$$\vec{\psi}(G) \lesssim 2k \cdot \lambda_f + 2k \left\| f - y_f \right\|_\pi \sqrt{\lambda_f} \lesssim \frac{(2k)^{1.5} \cdot \lambda_f}{\sqrt{\vec{\sigma}_{2k}^{v*}(G)}} \lesssim \frac{k \cdot \log(\Delta \cdot \alpha(G)) \cdot \vec{\lambda}_2^{v*}(G)}{\sqrt{\vec{\lambda}_k^{v*}(G)}}.$$

The proof of Proposition 5.10.9 is by combining the arguments in Proposition 4.6.1 for undirected vertex expansion and Proposition 5.7.4 for the rounding analysis. The proof of Proposition 5.10.10 is essentially the same as in Proposition 4.6.2. There are no new steps in these proofs, and so we omit them here.

5.11 Cheeger-Type Inequalities for Hypergraphs

Louis [Lou15] and Chan, Louis, Tang, Zhang [CLTZ18] developed a spectral theory for hypergraphs based on a continuous time diffusion process. They used it to derive a Cheeger inequality for hypergraph edge conductance, a higher-order Cheeger inequality for hypergraph k-way edge conductance, and a Cheeger inequality for hypergraph small-set conductance. Refer to Section 3.4 for details.

In this section, we will use the reweighted eigenvalue approach to derive similar results and compare with the results in [CLTZ18]. In addition, we will prove an improved Cheeger inequality for hypergraph edge conductance, which was not known before. Since the proofs of these results are all essentially the same as the corresponding proofs in Chapter 4, we only provide a proof for the improved Cheeger inequality for hypergraphs and just provide quick sketches for the other results.

We note that vertex expansion in a hypergraph H can simply be reduced to vertex expansion in its clique-graph G (see Remark 2.3.4), and so the results in Chapter 4 can be directly applied with $\Delta(G) \leq \Delta(H) \cdot r$, where r is the maximum size of a hyperedge in H. Therefore, we will only focus on hypergraph edge conductance in this section.

[7] Similar to Theorem 4.1.11 in the previous chapter, this result is an improvement of the original result in [LTW23], and this is thanks to using the relation $\vec{\sigma}_{2k}^{v*}(G) \geq k \cdot \vec{\lambda}_k^{v*}(G)$ instead of $\vec{\sigma}_k^{v*}(G) \geq \vec{\lambda}_k^{v*}(G)$.

5.11.1 Cheeger Inequality for Hypergraphs

In the reweighted eigenvalue approach, we use $\gamma_2^*(H)$ in Definition 5.1.9 as a relaxation to $\phi(H)$. We can prove the easy direction as in Proposition 5.3.2 by a reduction, but we actually do not need to prove it as we will see soon. As in Proposition 5.4.2, we can write $\gamma_2^*(H)$ as the following semidefinite program:

$$\begin{aligned}\gamma_2^*(H) := \min_{f:V\to\mathbb{R}^n} \max_{A\geq 0} \quad & \sum_{e\in E}\sum_{\{u,v\}\subseteq e} c(u,v,e)\left\|f(u)-f(v)\right\|^2 \\ \text{subject to} \quad & \sum_{\{u,v\}\subseteq e} c(u,v,e) \leq w(e) \qquad \forall u,v\in V \\ & \sum_{v\in V}\deg_w(v)\cdot f(v) = \vec{0} \\ & \sum_{v\in V}\deg_w(v)\cdot \left\|f(v)\right\|^2 = 1.\end{aligned}$$

By using LP duality in the inner maximization problem as in Lemma 5.7.5, it follows that

$$\begin{aligned}\gamma_2^*(H) = \min_{\substack{f:V\to\mathbb{R}^n\\ g:V\to\mathbb{R}}} \quad & \sum_{e\in E} w(e)g(e) \\ \text{subject to} \quad & g(e) \geq \left\|f(u)-f(v)\right\|^2 \quad \forall u,v\in e,\ \forall e\in E \\ & \sum_{v\in V}\deg_w(v)\cdot f(v) = \vec{0} \\ & \sum_{v\in V}\deg_w(v)\cdot \left\|f(v)\right\|^2 = 1.\end{aligned}$$

It turns out that $\gamma_2^*(H)$ in this form is exactly the same as $\tilde{\gamma}_2$ in [CLTZ18, SDP 8.3], and γ_2 in [CLTZ18] is simply this dual program restricted to one dimensional embeddings $f : V \to \mathbb{R}$ as stated in Lemma 5.7.6 (c.f. Definition 3.4.3). Therefore, Theorem 5.1.10 follows from their result in Theorem 3.4.2. We remark that it is also possible to derive Theorem 5.1.10 using the same proof as for Theorem 1.1.2.

5.11.2 Higher-Order Cheeger Inequality and Small-Set Expansion for Hypergraphs

We define $\gamma_k^*(H)$ as in Definition 5.1.9 but the objective is to maximize the k-th smallest eigenvalue of the normalized Laplacian matrix $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$. This is, however, not a convex optimization problem. Following Section 4.5, we change the objective to maximize the sum of the k smallest eigenvalues $\sum_{i=1}^{k} \lambda_i(\mathcal{L})$, so that we can use Proposition 2.8.3 to write this as a semidefinite program that we call $\sigma_k^*(H)$. Using the same manipulations as in Proposition 4.5.2, we can write

$$\begin{aligned}
\sigma_k^*(H) := \min_{f:V\to\mathbb{R}^n} \min_{g:V\to\mathbb{R}_{\geq 0}} \quad & \sum_{e\in E} g(e)\cdot w(e) \\
\text{subject to} \quad & g(e) \geq \|f(u)-f(v)\|^2 \qquad \forall\{u,v\}\subseteq e, \forall e\in E \\
& \sum_{v\in V} \deg_w(v)\cdot f(v)f(v)^T \preceq I_n \\
& \sum_{v\in V} \deg_w(v)\cdot \|f(v)\|^2 = k.
\end{aligned}$$

Retracing the same but rather long proof in Section 4.5, we can construct functions $f_1,\ldots,f_\ell$ with disjoint supports such that each is a good solution to Definition 5.1.9, and prove the exact same statement as Theorem 4.5.12.

Theorem 5.11.1 (Higher-Order Cheeger Inequality for Hypergraphs)**.** *For any undirected hypergraph $H=(V,E,w)$,*

$$\frac{1}{k}\sigma_k^*(H) \lesssim \phi_k(H) \lesssim k^4 \log k\sqrt{\log r\cdot\sigma_k^*(H)} \quad \text{and} \quad \phi_{(1-\varepsilon)k}(H) \lesssim \frac{1}{\varepsilon^4}\log k\sqrt{\log r\cdot\sigma_k^*(H)}.$$

Compared to (3.11), the result for k-way expansion is comparable with an extra factor of $\sqrt{k}$ but a factor of $\sqrt{\log r}$ less, while the result for $[(1-\varepsilon)k]$-way expansion for constant ε is an improvement by a factor of more than k^2. As a consequence, this also implies an improvement of (3.12) for small-set conductance by a factor of more than k.

5.11.3 Improved Cheeger Inequality for Hypergraphs

Using the reweighted eigenvalue approach, we can also prove an analog of the improved Cheeger's inequality as described in Section 5.10.3. This is a new result that was not obtained in [Lou15, CLTZ18]. Combining with the higher-order Cheeger inequality for hypergraphs in Theorem 5.11.1, this implies the following corollary that only depends on the combinatorial structure of H: If the k-way edge conductance $\phi_k(H)$ is large for a small k, then $\gamma_2^*(H)$ is a tighter approximation to $\phi(H)$.

Theorem 5.11.2 (Improved Cheeger's Inequality for Hypergraphs)**.** *Let $H = (V, E)$ be a hypergraph with weight function $w : E \to \mathbb{R}_{\geq 0}$. For any $2 \leq k \leq n$,*

$$\phi(H) \lesssim \frac{k \cdot \log r \cdot \gamma_2^*(H)}{\sqrt{\gamma_k^*(H)}}.$$ [8]

The proof of Theorem 5.11.2 follows the same two-step approach as in Section 3.1.4 and also in Section 5.10.3. Actually, the proof for hypergraphs is very similar to that for undirected vertex expansion in Section 4.6, and is easier than that for directed graphs in Section 5.10.3. We omit the proof for brevity.

5.12 Concluding Remarks

In this chapter, we show that the reweighted eigenvalue approach can be extended substantially to derive Cheeger inequalities for directed graphs and hypergraphs. Most notably, this develops into an interesting new spectral theory for directed graphs, which is much closer to the spectral theory for undirected graphs than what are previously known. We hope that this spectral theory will find more applications in practice, in clustering and partitioning of directed graphs and hypergraphs.

Technically, the reweighted eigenvalue approach provides an intuitive and unifying method to reduce the study of expansion properties in more general settings to the basic setting of edge conductance in undirected graphs. We believe that this approach can be used to lift more results in spectral graph theory for undirected graphs to more general settings, as the ideas are consistent with recent works on directed Laplacian solvers and hypergraph spectral sparsification that we mentioned in the end of Section 5.1.

There are some concrete open problems. The most obvious one is to prove tight bounds for the two main results Theorem 5.1.4 and Theorem 5.1.8, to settle whether the dependency on the asymmetric ratio can be completely removed or not[9]. Another one is to formulate and prove higher-order Cheeger inequality and bipartite Cheeger inequality for directed graphs as discussed in Section 5.10.

[8] Again, this is an improvement from $k^{3/2}$ to k over the original statement in [LTW23].

[9] See Remark 5.6.3 that the dimension reduction result for directed edge conductance is tight, and so a positive result removing the $\log \alpha(G)$ factor in Theorem 5.1.8 would probably need substantial new ideas. We incline to believe that the $\log \alpha(G)$ factor in Theorem 5.1.8 cannot be completely removed, but we do not have an example supporting this belief. We are less sure about what the right bound should be for Theorem 5.1.4.

Chapter 6

Reweighted Eigenvalues for Submodular Transformations

In the two previous chapters, we have used reweighted eigenvalues to derive Cheeger inequalities for different expansion quantities in generalized graphs. Encouraged by the success, the next questions are: how much further can reweighted eigenvalues be generalized? What are the limitations of the theory?

Motivated by Yoshida [Yos19] and Li and Milenkovic [LM18], who studied a very general class of cut functions called submodular transformations and proved Cheeger inequalities for them, we define reweighted eigenvalues for submodular transformations and attempt to build a spectral theory out of it. We obtain a Cheeger inequality and an improved Cheeger inequality for directed hypergraph expansion, which generalizes the corresponding results in Chapter 4 and Chapter 5. We then identify bottlenecks in our rounding algorithm for general submodular transformations, which provide evidence that reweighted eigenvalues may have limitations beyond the generalized graph models.

The plan for this chapter is as follows. We begin the chapter with Section 6.1, where we define submodular transformations and their expansion, Laplacian, and eigenvalues [LM18, Yos19]. We then review the past work by Yoshida [Yos19] and Li and Milenkovic [LM18] in Section 6.2. Following this, we present our first reweighting approach in Section 6.3. We first introduce the ℓ_1 reweighted program in Section 6.3.1 and prove a threshold rounding result relating it to the expansion. Then, in Section 6.3.2 we define reweighted eigenvalues for submodular transformations, which we show to be an SDP. We derive a Cheeger inequality for directed hypergraphs in Section 6.3.7. We also explain why difficulties arise when attempting to emulate the proof for general submodular transformations.

After that, in Section 6.3.8 we derive an improved Cheeger inequality for directed hypergraphs. We briefly discuss a second approach in Section 6.4 and conclude the chapter in Section 6.5.

6.1 Submodular Transformations

In this section, we introduce submodular transformation and its expansion, Laplacian, eigenvalue, and related concepts as defined in [LM18, Yos19]. Another excellent general reference is the book [Fuj05] by Fujishige.

6.1.1 Submodular Transformations and Expansion

Let V be a finite set. A function $f : 2^V \to \mathbb{R}$ is said to be submodular if

$$f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$$

for all $A, B \subseteq V$. We say that a submodular function f is normalized if $f(\emptyset) = f(V) = 0$ and symmetric if $f(S) = f(S^c)$ for all $S \subseteq V$.

There are two main reasons for studying algorithms for finding cuts of submodular functions. First, as proven in Proposition 6.1.1, all graph cut functions are submodular, so the submodularity property is inherent to the problems we study.

Proposition 6.1.1 (Graph Cut Functions are Submodular)**.** *Let V be a vertex set. Then, the following functions $f : 2^V \to \mathbb{R}$ are all submodular:*

(i) Let $G = (V, E)$ be a graph. f equals the edge cut function for an edge $uv \in E$, so that $f(S) = \mathbb{1}[uv \in \delta(S)]$ for $S \subseteq V$;

(ii) Let $G = (V, E)$ be a graph. f equals the vertex cut function for a vertex $v \in V$, so that $f(S) = \mathbb{1}[v \in S \wedge \partial(v) \not\subseteq S]$ for $S \subseteq V$;

(iii) Let $G = (V, E)$ be a directed graph. f equals the directed edge cut function for an arc $uv \in E$, so that $f(S) = \mathbb{1}[u \in S \wedge v \notin S]$ for $S \subseteq V$;

(iv) Let $G = (V, E)$ be a directed graph. f equals the directed vertex cut function for a vertex $v \in V$, so that $f(S) = \mathbb{1}[v \in S \wedge \partial^+(v) \not\subseteq S]$ for $S \subseteq V$;

(v) Let $H = (V, E)$ be an undirected hypergraph. f equals the hyperedge cut function for a hyperedge $e \in E$, so that $f(S) = \mathbb{1}[e \in \delta(S)]$ for $S \subseteq V$;

(vi) Let $H = (V, E)$ be a directed hypergraph. f equals the directed hyperedge cut function for a hyperedge $e = (e^-, e^+)$, so that $f(S) = \mathbb{1}[e^- \cap S \neq \emptyset \wedge e^+ \cap S^c \neq \emptyset]$ for $S \subseteq V$.

Furthermore, the functions (i) and (v) and symmetric, and all of them are normalized.

Proof. The submodularity of some of these functions are proved in [Yos19]. We will prove here that the function (vi) is submodular and leave it as an exercise to check that the functions (i) - (v) are special cases of (vi). We leave it to the reader to check that the functions (i) and (v) are symmetric and that all of them are normalized.

Let $e^-, e^+ \subseteq V$ be given and let $A, B \subseteq V$. We check the submodularity condition that

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B).$$

Observe that f can only take values 0 and 1, so there are only three cases to consider:

- If $f(A) = f(B) = 1$, there is nothing to prove.
- If $f(A) = 1$ and $f(B) = 0$ (the case where $f(A) = 0$ and $f(B) = 1$ is analogous), that means $A \cap e^- \neq \emptyset$ and $A^c \cap e^+ \neq \emptyset$, but either $e^- \subseteq B^c$ or $e^+ \subseteq B$. Suppose further that $f(A \cup B) = 1$; if not then we are done. That means that $(A \cup B)^c \cap e^+ \neq \emptyset$, or equivalently $A^c \cap B^c \cap e^+ \neq \emptyset$. In particular, $e^+ \not\subseteq B$.

  Let's show that $f(A \cap B) = 0$. From the preceding summary and the fact that $f(B) = 0$, it must be the case that $e^- \subseteq B^c$, and so $e^- \subseteq B^c \subseteq (A \cap B)^c$.
- If $f(A) = f(B) = 0$, one can show that (a) either $e^- \subseteq (A \cup B)^c = A^c \cap B^c$ or $e^+ \subseteq A \cup B$, and (b) either $e^- \subseteq (A \cap B)^c = A^c \cup B^c$ or $e^+ \subseteq A \cap B$. The proof is similar to that in the previous case. (a) implies that $f(A \cup B) = 0$ and (b) implies that $f(A \cap B) = 0$.

We conclude that f is indeed a submodular function. □

A second main reason for studying submodular functions is that submodularity is a fundamental property that frequently arises in combinatorial optimization problems in domains such as graph theory, machine learning, economics, and game theory. Submodularity can be understood as capturing a *diminishing return* phenomenon, which is present a lot of these applications. The property is formalized as Proposition 6.1.2 below. A spectral theory for submodular functions can potentially have far-reaching applications to problems in these domains. See e.g. [Fuj05] for general references to submodular functions and also [RY22, Yos19] for a more detailed account of the significance of submodularity in applications.

Proposition 6.1.2 (Diminishing Return [Fuj05])**.** *Let $f : 2^V \to \mathbb{R}$ be a submodular function, and let $S \subseteq T \subseteq V$ and $u \in V \setminus T$. Then,*

$$f(T \cup \{u\}) - f(T) \leq f(S \cup \{u\}) - f(S).$$

Proof. Take $A = T$ and $B = S \cup \{u\}$ in the definition of submodularity and rearrange. □

A submodular transformation is a collection $F = \{F_e\}_{e \in E}$ of functions $F_e : 2^V \to \mathbb{R}$ each of which is submodular [LM18, Yos19]. We say that F is a submodular transformation on (V, E) if its constituent functions are indexed by E and have common domain 2^V. We say that a submodular transformation F is symmetric (respectively, normalized) if all its constituent functions F_e are symmetric (respectively, normalized). Write $F(S) := \sum_{e \in E} F_e(S)$.

As the sum of two submodular functions is again submodular, submodular transformation can be thought of as decomposing more complex submodular functions into the sum of "simpler" submodular functions. For example, the cut size function $F(S) := |\delta(S)|$ in a graph $G = (V, E)$ can be decomposed into the sum of single-edge cut functions in Proposition 6.1.1(i).

The expansion of a submodular transformation relative to a measure $\mu : V \to \mathbb{R}^+$ is defined naturally as follows.

Definition 6.1.3 (Expansion of Submodular Transformations [LM18, Yos19])**.** *Given a submodular transformation $F = \{F_e\}_{e \in E}$ on (V, E). Suppose that a measure $\mu : V \to \mathbb{R}^+$ is given. The (μ-)expansion of a subset $S \subseteq V$ and of F are defined as*

$$\phi_\mu(S) := \frac{F(S)}{\min(\mu(S), \mu(S^c))} \qquad \text{and} \qquad \phi_\mu(F) := \min_{\emptyset \neq S \subset V} \phi_\mu(S).$$

With Proposition 6.1.1 in mind, one can verify that this definition encompasses generalized graph expansions defined in Section 2.3, by setting each F_e appropriately.

6.1.2 Lovász Extension

Cheeger inequalities relate expansion quantities with the smallest nontrivial eigenvalue of an appropriately defined Laplacian. It seems necessary to define the Laplacian of a submodular transformation in order to be able to properly develop a spectral theory.

The Lovász extension [Lov83] is a *convex* extension of submodular functions from the discrete domain 2^V to the continuous domain $\mathbb{R}^V$. It crucially makes submodular optimization amenable to continuous methods in convex optimization. See also [Fuj05, Section 6.3] for reference.

We first need the notion of the base polytope of a submodular function.

Definition 6.1.4 (Base Polytope of Submodular Functions [Lov83, Fuj05])**.** *Given a submodular function $f : 2^V \to \mathbb{R}$, its base polytope $\mathcal{B}(f)$ is defined as*

$$\mathcal{B}(f) := \{x \in \mathbb{R}^V : x(S) \leq f(S)\ \forall S \subseteq V \ \wedge\ x(V) = f(V)\}.$$

Proposition 6.1.5 (Some Properties of Base Polytopes [Fuj05])**.** *The base polytope $\mathcal{B}(f)$ of a submodular function $f : 2^V \to \mathbb{R}$ satisfies the following properties:*

1. *For $x \in \mathbb{R}^V$, the set of extreme points of the set $\arg\max_{y \in \mathcal{B}(f)} \langle y, x \rangle$ of points in $\mathcal{B}(f)$ are precisely the set of points $w \in \mathbb{R}^V$ generated by the following procedure: order the elements of V as $u_1, \ldots, u_{|V|}$ such that $x(u_1) \geq x(u_2) \geq \cdots \geq x(u_{|V|})$. For $i = 1, 2, \ldots, |V|$ set*

$$y(u_i) := f(\{u_1, \ldots, u_i\}) - f(\{u_1, \ldots, u_{i-1}\}).$$

2. *The set of all extreme points of $\mathcal{B}(f)$ are precisely the set of points $w \in \mathbb{R}^V$ generated by the following procedure: order the elements of V as $u_1, \ldots, u_{|V|}$. For $i = 1, 2, \ldots, |V|$ set*

$$y(u_i) := f(\{u_1, \ldots, u_i\}) - f(\{u_1, \ldots, u_{i-1}\}).$$

   In particular, $\mathcal{B}(f)$ has at most $|V|!$ extreme points.

3. *If f is normalized with values in $[0, M]$, and let $y \in \mathcal{B}(f)$ be a point in the base polytope of f. Then, $\|y\|_1 \leq 2M$.*

4. *If f is symmetric, then $\mathcal{B}(f)$ is symmetric, so that $y \in \mathcal{B}(f) \Leftrightarrow -y \in \mathcal{B}(f)$.*

We identify below an important subclass of submodular functions. Given a submodular function $f : 2^V \to \mathbb{R}$, we say that f is simple if f only takes values 0 and 1. We say that a submodular transformation $F = \{F_e\}_{e\in E}$ is simple if each F_e is simple. This definition is useful because when f is simple, the extreme points of $\mathcal{B}(f)$ take a particularly simple form. As we shall prove below, this is also equivalent to f being a directed hyperedge cut function.

Proposition 6.1.6. *Let $f : 2^V \to \mathbb{R}$ be a normalized submodular function with $f(S) \in [0, 1]$ for all $S \subseteq V$. The following are equivalent:*

(i) f is simple;

(ii) The nonzero extreme points of $\mathcal{B}(f)$ are all of the form $\mathbb{1}_u - \mathbb{1}_v$ for some $u, v \in V$;

(iii) f is the directed hyperedge cut function in Proposition 6.1.1(vi).

Proof. We will prove that $(i) \implies (ii) \implies (iii) \implies (i)$.

$(i) \implies (ii)$: Suppose f is simple. By the second property in Proposition 6.1.5, the extreme points of $\mathcal{B}(f)$ are of the form $y(v_i) := f(S_i) - f(S_{i-1})$ where $v_1, v_2, \ldots, v_n$ is an ordering of the elements in V and $S_i := \{v_1, \ldots, v_i\}$. Since f is simple, y can take only values $0, 1, -1$. We need to show that if $y(v_i) = -1$ for some i, then $y(v_j) = 0$ for all $j > i$. Suppose otherwise that $j > i$ is the smallest index such that $y(v_j) \neq 0$. If $y(v_j) = -1$ then $f(S_j) = f(S_{j-1}) + 1 = f(S_i) + 1 = f(S_{i-1}) + 2$ which is impossible. If $y(v_j) = 1$, let $T := S_{i-1} \cup \{v_j\}$. Note that the condition $y(v_i) = -1$ implies $f(S_{i-1}) = 1$ and $f(S_i) = 0$, and the condition $y(v_j) = 1$ implies $f(S_{j-1}) = 0$ and $f(S_j) = 1$. Then, by submodularity,

$$f(T) = f(T) + f(S_{j-1}) \geq f(T \cup S_{j-1}) + f(T \cap S_{j-1}) = f(S_j) + f(S_{i-1}) = 2,$$

which contradicts the assumption that f is simple. This proves the assertion $(i) \implies (ii)$.

$(ii) \implies (iii)$: given the condition on f, we construct the following directed hyperedge $e = (e^-, e^+)$. Let $e^- := \{u \in V : \mathbb{1}_u - \mathbb{1}_v \in \mathcal{B}(f)$ for some $v \neq u\}$, and let $e^+ := \{u \in V : \mathbb{1}_v - \mathbb{1}_u \in \mathcal{B}(f)$ for some $v \neq u\}$. We show that f is precisely the hyperedge cut function for e. Let $S \subseteq V$. There are three cases to consider:

- Case 1: $e^- \cap S = \emptyset$. In this case, consider an ordering of the vertices $v_1, v_2, \ldots, v_n$ such that $S = \{v_1, \ldots, v_k\}$. Then, the extreme point y corresponding to the vertex ordering must have $y(v_1) = y(v_2) = \cdots = y(v_k) = 0$, and so $f(S) = 0$.

- Case 2: $e^- \cap S \neq \emptyset$ but $e^+ \subseteq S$. In this case, consider an ordering of the vertices $v_1, \ldots, v_n$ such that $S = \{v_1, \ldots, v_k\}$ and $v_1 \in e^- \cap S$. By the diminishing return property, we have $f(\{v_1\}) \geq 1$, so equality must hold and $y(v_1) = 1$. Since $e^+ \subseteq S$, we must have $y(v_j) = -1$ for some $j \leq k$ and $y(v_\ell) = 0$ for all $\ell \neq 1, j$, and so $f(S) = y(v_1) + \cdots + y(v_k) = 0$.

- Case 3: $e^- \cap S \neq \emptyset$ and $e^+ \not\subseteq S$. In this case, consider an ordering of the vertices $v_1, \ldots, v_n$ such that $S = \{v_1, \ldots, v_k\}$, $v_1 \in e^- \cap S$, and $v_n \in e^+ \setminus S$. Again, we have $f(\{v_1\}) = 1$, and by diminishing return and the fact that $\mathbb{1}_u - \mathbb{1}_{v_n}$ is an extreme point of $\mathcal{B}(f)$ for some $u \in V$, we have $f(V) - f(V \setminus \{v_n\}) \leq -1$. So, the extreme point corresponding to this vertex ordering must be $\mathbb{1}_{v_1} - \mathbb{1}_{v_n}$, and $f(\{v_1, \ldots, v_\ell\}) = 1$ for $1 \leq \ell \leq n-1$. In particular, we have $f(S) = 1$.

This completes the proof of $(ii) \implies (iii)$.

Finally, $(iii) \implies (i)$ is trivial, and so the proof is complete. □

Example 6.1.7 (Base Polytope of Directed Hyperedge Cut Functions). *Let $e = (e^-, e^+)$ be a directed hyperedge and*

$$f(S) = \mathbb{1}[e^- \cap S \neq \emptyset \wedge e^+ \cap S^c \neq \emptyset]$$

be the corresponding cut function as defined in Proposition 6.1.1. As we have shown in Proposition 6.1.6, the set of extreme points of $\mathcal{B}(f)$ is precisely $\{\mathbb{1}_u - \mathbb{1}_v \mid u \in e^-, v \in e^+\}$.

We now define Lovász extension and collect some useful and essential properties.

Definition 6.1.8 (Lovász Extension [Lov83, Fuj05]). *Given a submodular function $f : 2^V \to \mathbb{R}$, the Lovász extension $\tilde{f} : \mathbb{R}^V \to \mathbb{R}$ of f is defined as*

$$\tilde{f}(x) := \max_{y \in \mathcal{B}(f)} \langle y, x \rangle.$$

Proposition 6.1.9 (Basic Properties of Lovász Extension [Lov83, Fuj05]). *Given a submodular function $f : 2^V \to \mathbb{R}$, its Lovász extension $\tilde{f} : \mathbb{R}^V \to \mathbb{R}$ satisfies the following properties:*

1. *$\tilde{f}$ is convex.*

2. *If f is normalized, then $f(S) = \tilde{f}(\mathbb{1}_S)$ for $S \subseteq V$.*

3. *If f is normalized, then $\tilde{f}(x) = \tilde{f}(x + c \cdot \mathbb{1}_V)$ for any $c \in \mathbb{R}$.*

4. *For any* $c \geq 0$, $\tilde{f}(cx) = c \cdot \tilde{f}(x)$.

5. *If* f *is symmetric, then* $f(x) = f(-x)$.

6. *For* $x \in \mathbb{R}^V$, *the value of* $\tilde{f}(x)$ *may be determined as follows: order the elements of* V *as* $u_1, \ldots, u_{|V|}$ *such that* $x(u_1) \geq x(u_2) \geq \cdots \geq x(u_{|V|})$. *Then,*

$$\tilde{f}(x) = \sum_{i=1}^{n} x(u_i) \cdot (f(\{u_1, \ldots, u_i\}) - f(\{u_1, \ldots, u_{i-1}\})).$$

The last property will be referred to as the "level set property" of Lovász extension, as it expresses $\tilde{f}$ in terms of the value of f on the level sets

$$S_i := \{u_1, \ldots, u_i\} = \{u \in V : x(u) \geq x(u_i)\}$$

of the input x. It is the essential property that enables threshold rounding.

6.1.3 Laplacian Eigenvalues

As observed in [LM18, Yos19], the Laplacian of a submodular transformation can be defined using Lovász extension.

Definition 6.1.10 (Laplacians of Submodular Transformations [LM18, Yos19])**.** *Given a normalized submodular transformation* $F = \{F_e\}_{e \in E}$ *on* (V, E), *its Laplacian* L_F *is defined as the following multi-valued function from* $\mathbb{R}^V$ *to* $\mathbb{R}^V$:

$$L_F(x) := \left\{ \sum_{e \in E} y_e \langle y_e, x \rangle \mid y_e \in \arg \max_{y \in \mathcal{B}(F_e)} \langle y, x \rangle \right\} \subseteq \mathbb{R}^V.$$

While $L_F(x)$ is multi-valued, the "quadratic form" $\langle x, L_F(x) \rangle$ is single-valued, since for any $y_e \in \arg\max_{y \in \mathcal{B}(F_e)} \langle y, x \rangle$, one has

$$\left\langle x, \sum_{e \in E} y_e \langle y_e, x \rangle \right\rangle = \sum_{e \in E} \langle y_e, x \rangle^2 = \sum_{e \in E} \left(\max_{y \in \mathcal{B}(F_e)} \langle y, x \rangle \right)^2 = \sum_{e \in E} \tilde{F}_e(x)^2,$$

which is independent of the choice of y_e. The "sum of energies" form

$$\langle x, L_F(x) \rangle = \sum_{e \in E} \tilde{F}_e(x)^2 \tag{6.1}$$

also provides an intuitive explanation for this definition of the Laplacian [Yos19].

To further motivate the definition, we shall see in Example 6.1.11 that Definition 6.1.10 generalizes the ordinary graph Laplacians in Section 2.5, as well as the nonlinear Laplacians by Yoshida [Yos16] (see Section 3.3.2) and Louis [Lou15] and Chan, Louis, Teng, and Zhang [CLTZ18] (see Section 3.4.1).

Example 6.1.11 (Comparison with Graph Laplacians)**.** *We demonstrate that Definition 6.1.10 captures several definitions of graph Laplacians that we have seen.*

- *Let $G = (V, E)$ be a graph. The unnormalized graph Laplacian has the quadratic form*

$$\sum_{uv \in E} (x(u) - x(v))^2.$$

Let F_e be the edge cut function in Proposition 6.1.1(i). By specializing the base polytope characterization in Example 6.1.7 to ordinary graphs, we know that the extreme points of $\mathcal{B}(F_e)$ are $\mathbb{1}_u - \mathbb{1}_v$ and $\mathbb{1}_v - \mathbb{1}_u$ where $e = uv \in E$. Therefore,

$$\tilde{F}_e(x) = \max_{y \in \mathcal{B}(F_e)\rangle y,x} = \max(x(u) - x(v), x(v) - x(u)) = |x(u) - x(v)|,$$

and the sum of energies form in (6.1) yields

$$\langle x, L_F(x)\rangle = \sum_{uv \in E} (x(u) - x(v))^2.$$

- *Let $G = (V, E)$ be a directed graph. Yoshida [Yos16] defined a nonlinear Laplacian to study directed graph conductance; see Section 3.3.2 for review. The unnormalized Laplacian has the quadratic form*

$$\langle x, L'x\rangle = \sum_{uv \in E} \left[(x(u) - x(v))^+\right]^2.$$

For each hyperedge $e \in E$, the base polytope of the directed edge cut function F_e in Proposition 6.1.1(iii) has extreme points $\mathbb{1}_u - \mathbb{1}_v$ and $\vec{0}$ where $e = uv \in E$, and again the Laplacian defined in Definition 6.1.10 coincides with the one above.

- *Let $H = (V, E)$ be a hypergraph. Louis [Lou15] and Chan, Louis, Tang, and Zhang [CLTZ18] defined a nonlinear Laplacian to build a spectral theory for hypergraphs; see Section 3.4.1 for review. The unnormalized Laplacian has the quadratic form*

$$\sum_{e \in E} \max_{u,v \in e} (x(u) - x(v))^2.$$

For each hyperedge $e \in E$, the base polytope of the hyperedge cut function F_e in Proposition 6.1.1(v) has extreme points $\{\mathbb{1}_u - \mathbb{1}_v\}_{u,v\in e}$, and again the Laplacian defined in Definition 6.1.10 coincides with the one above.

We can then define eigenvalues of the Laplacian L_F using the Courant-Fischer formulation in Proposition 2.4.4. Since F is assumed to be normalized, $F_e(V) = 0$ for any $e \in E$. Hence, for any $c \in \mathbb{R}$, using

$$\max_{y\in\mathcal{B}(F_e)} \langle y, c \cdot \mathbb{1}_V\rangle = \max_{y\in\mathcal{B}(F_e)} c \cdot y(V) = c \cdot F_e(V) = 0.$$

This implies that $L_F(c\cdot \mathbb{1}_V) = \{\vec{0}\}$. Moreover, $\langle x, L_F(x)\rangle \geq 0$ for all $x \in \mathbb{R}^V$, due to the sum of energies form (6.1). Therefore, it makes sense to consider 0 as the trivial eigenvalue of L_F, with corresponding eigenvector $\mathbb{1}_V$. The hope is that the smallest nontrivial eigenvalue carries more information about the expansion of a submodular transformation.

Definition 6.1.12 (Nontrivial Eigenvalue [LM18, Yos19])**.** *Given a normalized submodular transformation $F = F_e$ on (V, E) and a vertex measure $\mu : V \to \mathbb{R}^+$. The nontrivial (μ-normalized) eigenvalue of F is defined as*

$$\lambda_\mu(F) := \min_{\substack{x\in\mathbb{R}^V\\ x\perp\mu}} \frac{\langle x, L_F(x)\rangle}{\sum_{u\in V} \mu(u)x(u)^2}.$$

Continuing Example 6.1.11, we leave it to the reader to verify that Definition 6.1.12 captures the definitions of the second smallest eigenvalue of (i) the unnormalized and normalized graph Laplacians in (by taking $\mu = \mathbb{1}$ and $\mu = \deg$ respectively), (ii) the nonlinear Laplacian defined by [Yos16] (by taking $\mu = \deg$), and (iii) the hypergraph Laplacian defined by Louis [Lou15] and Chan, Louis, Tang, and Zhang [CLTZ18] (by taking $\mu = \deg$). Note that here we deliberately leave the choice of vertex measure μ open.

We review past work on spectral theory for submodular transformations before describing our reweighted eigenvalue approach. As we shall see, the past work may be described as proving a Cheeger inequality relating the nontrivial eigenvalue (which is an ℓ_2^2 quantity) to the expansion (which is an ℓ_1 quantity), then design a relaxation (that is also ℓ_2^2) that is efficiently computable and whose solution can be rounded to a solution to the Laplacian eigenvalue problem. On the contrary, the reweighted eigenvalue approach represents a different philosophy of designing the relaxation at the ℓ_1 level, then lifting the ℓ_1 program to the ℓ_2^2 program.

6.2 Review of Past Work

In this section, we review the Cheeger inequality relating the nontrivial eigenvalue and the expansion of a submodular transformation, as proven by Li and Milenkovic [LM18] in the symmetric case and Yoshida [Yos19] in the general case. Then, we review convex relaxations of the nontrivial eigenvalue $\lambda_\mu(F)$, by Yoshida [Yos19].

6.2.1 Cheeger Inequality for Submodular Transformations

Yoshida [Yos19] proved a Cheeger inequality for general submodular transformations. He considered the setting where the vertex measure $\mu(u)$ is be the number of functions F_e influenced by u, i.e.

$$\mu(u) := |\{e \in E : F_e(S \cup \{u\}) \neq F_e(S) \text{ for some } S \subseteq V\}|\,. \tag{6.2}$$

This vertex measure is equal to the degree measure in unweighted generalized graphs.

Theorem 6.2.1 (Cheeger Inequality for Submodular Transformations [Yos19, Theorem 1.3])**.** *Let $F = \{F_e\}$ be a normalized submodular transformation on (V, E), such that each constituent function F_e has range $\subseteq [0,1]$. Define the measure $\mu : V \to \mathbb{R}^+$ as in (6.2). Then,*

$$\frac{\lambda_\mu(F)}{2} \leq \phi_\mu(F) \leq \sqrt{2\lambda_\mu(F)}.$$

Proof. First we prove the easy direction. Let $S \subseteq V$ be such that

$$\phi_\mu(S) = \frac{F(S)}{\min(\mu(S), \mu(S^c))}$$

is minimized. Define the following solution $x : V \to \mathbb{R}$ to $\lambda_\mu(F)$ as

$$x(u) = \begin{cases} \frac{1}{\mu(S)}, & \text{if } u \in S, \\ \frac{-1}{\mu(S^c)}, & \text{otherwise.} \end{cases}$$

Then, $\sum_{u \in V} \mu(u)x(u) = 0$, and by the first property of base polytopes in Proposition 6.1.5

and telescoping sum we have

$$\begin{aligned}
\lambda_\mu(F) &\leq \frac{\langle x, L_F(x)\rangle}{\sum_{u\in V}\mu(u)x(u)^2} \\
&= \left(\frac{1}{\mu(S)}+\frac{1}{\mu(S^c)}\right)^{-1}\cdot\sum_{e\in E}\left[\frac{F_e(S)-F_e(\emptyset)}{\mu(S)}-\frac{F_e(V)-F_e(S)}{\mu(S^c)}\right]^2 \\
&= \left(\frac{1}{\mu(S)}+\frac{1}{\mu(S^c)}\right)^{-1}\cdot\sum_{e\in E}F_e(S)^2\cdot\left(\frac{1}{\mu(S)}+\frac{1}{\mu(S^c)}\right)^2 \\
&\leq \left(\frac{1}{\mu(S)}+\frac{1}{\mu(S^c)}\right)\cdot\sum_{e\in E}F_e(S) \qquad (\because F_e(S)\in[0,1]) \\
&\leq 2\phi_\mu(S).
\end{aligned}$$

For the hard direction we break the proof into two steps per Theorem 3.1.1.

Step 1 (ℓ_2^2 to ℓ_1). Given $x : V \to \mathbb{R}$, our goal is to construct $h : V \to \mathbb{R}$ such that

$$\frac{\sum_{e\in E}\tilde{F}_e(h)}{\sum_{v\in V}\mu(v)|h(v)|} \leq \sqrt{2\cdot\frac{\sum_{uv\in E}\tilde{F}_e(x)^2}{\sum_{v\in V}\mu(v)x(v)^2}}.$$

Let $c \in \mathbb{R}$ be a μ-weighted median of x; that is, $\mu(\{v \in V : x(v) > c\}) \leq \mu(V)/2$ and $\mu(\{v \in V : x(v) < c\}) \leq \mu(V)/2$. Define $h : V \to \mathbb{R}$ so that

$$h(u) := \begin{cases}(x(u)-c)^2, & \text{if } f(u) \geq c; \\ -(x(u)-c)^2, & \text{if } f(u) < c.\end{cases}$$

Check that 0 is a μ-weighted median of h. In the denominator,

$$\sum_{v\in V}\mu(v)|h(v)| = \sum_{v\in V}\mu(v)(x(v)-c)^2 \geq \sum_{v\in V}\mu(v)x(v)^2.$$

where the inequality is due to $\sum_{v\in V}\mu(v)x(v) = 0$ and using Fact 2.10.3. In the numerator, we use crucially the fact that h preserves the order in x: if $u_1, \ldots, u_n$ is an ordering of the elements in V such that $x(u_1) \geq x(u_2) \geq \cdots \geq x(u_n)$, then $h(u_1) \geq h(u_2) \geq \cdots \geq h(u_n)$ as well. For each $e \in E$, let $i_1^e, \ldots, i_{n_e}^e$ be the subsequence of indices such that $u_{i_j^e}$ influences

F_e. Using Proposition 6.1.5,

$$\begin{aligned}
\sum_{e\in E}\tilde{F}_e(h) &= \sum_{e\in E}\sum_{i=1}^{n} h(u_i)\cdot(F_e(S_i)-F_e(S_{i-1})) \\
&= \sum_{e\in E}\sum_{j=1}^{n_e-1} F_e(S_{i_j^e})(h(u_{i_j^e})-h(u_{i_{j+1}^e})) \\
&\le \sum_{e\in E}\sum_{j=1}^{n_e-1} F_e(S_{i_j^e})(x(u_{i_j^e})-x(u_{i_{j+1}^e}))\cdot(|x(u_{i_j^e})-c|+|x(u_{i_{j+1}^e})-c|) \quad \text{(by (3.2))} \\
&\le \sqrt{\sum_{e\in E}\sum_{j=1}^{n_e-1} F_e(S_{i_j^e})(x(u_{i_j^e})-x(u_{i_{j+1}^e}))^2\cdot 2\sum_{e\in E}\|F_e\|_\infty\cdot\sum_{j=1}^{n_e}|x(u_{i_j^e})-c|^2} \\
&\le \sqrt{\sum_{e\in E}\sum_{j=1}^{n_e-1}\left[F_e(S_{i_j^e})(x(u_{i_j^e})-x(u_{i_{j+1}^e}))\right]^2\cdot 2\,\|x-c\|_\mu^2} \\
&= \sqrt{\sum_{e\in E}\tilde{F}_e(x)^2\cdot 2\sum_{v\in V}\mu(v)|h(v)|},
\end{aligned}$$

where the last inequality uses the definition of μ in (6.2). In conclusion,

$$\frac{\sum_{uv\in E}\tilde{F}_e(h)}{\sum_{v\in V}\mu(v)|h(v)|}\le\sqrt{2\cdot\frac{\sum_{uv\in E}\tilde{F}_e(x)^2}{\sum_{v\in V}\mu(v)|h(v)|}}\le\sqrt{2\cdot\frac{\sum_{uv\in E}\tilde{F}_e(x)^2}{\sum_{v\in V}\mu(v)x(v)^2}},$$

as claimed.

Step 2 (threshold rounding). We will exhibit a subset $S\subseteq V$ such that $0<\mu(S)\le\mu(V)/2$ and

$$\min(\phi_\mu(S),\phi_\mu(S^c))=\frac{\min(F(S),F(S^c))}{\mu(S)}\le\frac{\sum_{e\in E}\tilde{F}_e(h)}{\sum_{v\in V}\mu(v)|h(v)|}.$$

Let $t\in\mathbb{R}$ be a parameter, and define $S_t\subseteq V$ as follows:

$$S_t:=\begin{cases}\{v\in V: h(v)>t\} & \text{if } t\ge 0\\ \{v\in V: h(v)<t\} & \text{if } t<0.\end{cases}$$

Note that, since 0 is a μ-weighted median of h, $\mathrm{vol}(S_t)$ is at most $\mathrm{vol}(V)/2$ for any $t \in \mathbb{R}$. The "average" denominator is

$$\int_{-\infty}^{\infty} \mu(S_t)\, dt = \sum_{v \in V} \mu(v) \int_{-\infty}^{\infty} \mathbb{1}[v \in S_t]\, dt = \sum_{v \in V} \mu(v)|h(v)|,$$

and letting $h(u_1) \geq h(u_2) \geq \cdots \geq h(u_n)$ the "average" numerator is

$$\begin{aligned}
\int_{-\infty}^{\infty} \min(F(S_t), F(S_t^c))\, dt &\leq \sum_{e \in E} \int_{-\infty}^{0} F_e(S_t^c)\, dt + \sum_{e \in E} \int_{0}^{\infty} F_e(S_t)\, dt \\
&= \int_{-\infty}^{\infty} F_e\left(\{u \in V : h(u) > t\}\right)\, dt \\
&= \sum_{e \in E} \sum_{i=1}^{n-1} (h(u_i) - h(u_{i+1})) F_e(\{u_1, \ldots, u_i\}) \\
&= \sum_{e \in E} \tilde{F}_e(h),
\end{aligned}$$

where the last equality is again by the level set property of $\tilde{F}_e$ in Proposition 6.1.9.

By the averaging argument, the proof is complete. □

Note that the nonnegativity of the submodular functions F_e is essential, but it is possible to derive a weighted version of Theorem 6.2.1 where the functions F_e are assumed to be in range $[0, W_e]$ instead of $[0, 1]$ uniformly.

Formally, Theorem 6.2.1 matches the classical Cheeger inequality (Theorem 3.1.1), even up to constant factors. One crucial difference between them, however, is that while there is a fast algorithm to compute a cut with conductance at most $\sqrt{2\phi(G)}$ Theorem 6.2.1 does not come with a polynomial-time algorithm that finds low-expansion subsets. Indeed, $\lambda_\mu(F)$ is NP-hard to compute under the small-set expansion hypothesis [Yos19].

6.2.2 SDP Relaxations for Submodular Transformations

To deal with the computational hardness issue, Yoshida [Yos19] designed an SDP relaxation of the $\lambda_\mu(F)$ program in Definition 6.1.12. Guarantees on both the running time and the approximation factor, depending on the "complexity" of the submodular functions F_e, are proven. Below we describe the SDP relaxation and state the main results from [Yos19]. For proofs and further details, we refer the reader to their paper.

Suppose first that the submodular transformation F is symmetric, i.e. $F_e(S) = F_e(S^c)$ for all $S \subseteq V$. In this case, the base polytope $\mathcal{B}(F_e)$ is symmetric by Proposition 6.1.5, and the energy of a submodular function F_e has a particularly simple form:

$$\tilde{F}_e(x)^2 = \left[\max_{y\in\mathcal{B}(F_e)} \langle y, x\rangle\right]^2 = \max_{y\in\mathcal{B}(F_e)} \left[\langle y, x\rangle^2\right]. \tag{6.3}$$

Using this, the $\lambda_\mu(F)$ program in Definition 6.1.12 can be rewritten as the following:

$$\begin{aligned}
\min_{\substack{x:V\to\mathbb{R}\\ \eta_e\in\mathbb{R}}} \quad & \sum_{e\in E} \eta_e \\
\text{subject to} \quad & \langle y, x\rangle^2 \leq \eta_e \qquad \forall e \in E,\ \forall y \in \mathcal{B}(F_e) \\
& \sum_{u\in V} \mu(u)x(u)^2 = 1 \\
& \sum_{u\in V} \mu(u)x(u) = 0.
\end{aligned}$$

This admits a natural SDP relaxation as below.

Definition 6.2.2 (SDP Relaxation, Symmetric Case [Yos19, LM18])**.** *Let $F = \{F_e\}_{e\in E}$ be a normalized submodular transformation on (V, E). Suppose further that F is symmetric. Define the $\mathsf{sdp}_\mu(F)$ program as follows:*

$$\begin{aligned}
\min_{\substack{X:V\to\mathbb{R}^N\\ \eta_e\in\mathbb{R}^N}} \quad & \sum_{e\in E} \|\eta_e\|^2 \\
\text{subject to} \quad & \left\|y^T X\right\|^2 \leq \|\eta_e\|^2 \qquad \forall e \in E,\ \forall y \in \mathcal{B}(F_e) \\
& \sum_{u\in V} \mu(u) \|X(u)\|^2 = 1 \\
& \sum_{u\in V} \mu(u)X(u) = \vec{0}.
\end{aligned}$$

It is easy to see that $\mathsf{sdp}_\mu(F)$ is a relaxation of $\lambda_\mu(F)$, and Yoshida upper bounds the objective value of the SDP relaxation in terms of $\lambda_\mu(F)$ via random projection. Note that as presented, the SDP has uncountably many constraints, one for each point in one of the base polytopes. Yoshida proposed two ways to remedy this. One is to consider y to be an extreme point of $\mathcal{B}(F_e)$ only. This decreases the number of constraints to at most $|V|! \cdot |E|$ (by Proposition 6.1.5) and even down to $\text{poly}(|V|, |E|)$ in many cases. Another is to find an ε-cover for the base polytopes, decreasing the number of constraints to the number of ε-balls in the cover, but incurring an additive error. See [Yos19] for more details. We state the main approximation results in Theorem 6.2.5 after introducing the SDP relaxation in the general case.

If the additional assumption that F is symmetric is dropped, then the energy of its constituent function may no longer enjoy the simple form (6.3). Yoshida wrote the energy in the form

$$\tilde{F}_e(x)^2 = \left[\max_{y \in \mathcal{B}(F_e)} \langle y, x \rangle\right]^2 = \frac{1}{2} \cdot \max_{y \in \mathcal{B}(F_e)} \left[\langle y, x \rangle^2 + \langle y, x \rangle \cdot |\langle y, x \rangle|\right], \tag{6.4}$$

and his idea for designing an SDP relaxation in the general case was to introduce additional vector variables to represent the signed quantity $\langle y, x \rangle$.

Definition 6.2.3 (SDP Relaxation, General Case [Yos19])**.** *Let $F = \{F_e\}_{e \in E}$ be a normalized submodular transformation on (V, E). Define the $\mathsf{sdp}_\mu(F)$ program as follows:*

$$\begin{aligned}
\min_{\substack{X: V \to \mathbb{R}^N \\ \eta_e, \xi_y \in \mathbb{R}^N}} \quad & \frac{1}{2} \sum_{e \in E} \|\eta_e\|^2 \\
\text{subject to} \quad & \left\|y^T X\right\|^2 + \langle y^T X, \xi_y \rangle \le \|\eta_e\|^2 \qquad \forall e \in E, \forall y \in \mathcal{B}(F_e) \\
& \langle \xi_y, \xi_1 \rangle \ge \|\xi_y\|^2 \qquad \forall e \in E, \forall y \in \mathcal{B}(F_e) \\
& \|\xi_y\|^2 = \left\|y^T X\right\|^2 \qquad \forall e \in E, \forall y \in \mathcal{B}(F_e) \\
& \sum_{u \in V} \mu(u) \|X(u)\|^2 = 1 \\
& \sum_{u \in V} \mu(u) X(u) = \vec{0}.
\end{aligned}$$

Compared to the SDP relaxation in the symmetric case, additional vector variables ξ_y for $y \in \cup_{e \in E} \mathcal{B}(F_e)$ and ξ_1 indicating the "positive" direction are introduced. The term $\langle y^T X, \xi_y \rangle$ in the first constraint is supposed to represent the term $\langle y, x \rangle \cdot |\langle y, x \rangle|$ in (6.4), the second set of constraints asserts that ξ_y aligns with the "positive" direction indicated by ξ_1, and the third set of constraints ensures that the vector energy $\left\|y^T X\right\|^2 + \langle y^T X, \xi_y \rangle$ is between 0 and $2\left\|y^T X\right\|^2$. Similarly to the symmetric case, some manipulations are needed to decrease the number of constraints. In particular, here there is one variable ξ_y for each constraint, and so the ambient dimension N for the vector variables, and hence the efficiency of solving the SDP, depends heavily on the number of constraints.

Theorem 6.2.4 (SDP Relaxation Guarantees, Symmetric Case [Yos19]). *Let $F = \{F_e\}_{e \in E}$ be a nonnegative, normalized submodular transformation on (V, E), with $n := |V|$ and $m := |E|$. Let $\mu : V \to \mathbb{R}^+$ be the measure given in (6.2). Suppose that F is symmetric.*

- *For any $\varepsilon > 0$,*

$$\mathsf{sdp}_\mu(F) \leq \lambda_\mu(F) \leq O\left(\frac{\log n}{\varepsilon^2}\mathsf{sdp}_\mu(F) + \varepsilon B^2\right),$$

  where B is the maximum ℓ_2 norm of a point in any of the $\mathcal{B}(F_e)$. Furthermore, there is a randomized algorithm that computes a solution x to $\lambda_\mu(F)$ in Definition 6.1.12, so that $x \perp \mu$ and

$$\frac{\langle x, L_F(x)\rangle}{\sum_{u\in V}\mu(u)x(u)^2} \leq O\left(\frac{\log n}{\varepsilon^2}\mathsf{sdp}_\mu(F) + \varepsilon B^2\right),$$

  and the runtime of the algorithm is $O(\mathrm{poly}(nm)^{\mathrm{poly}(1/\varepsilon)})$.

- *Suppose that the number of extreme points of each $\mathcal{B}(F_e)$ is upper bounded by N. Then, the above upper bound on $\lambda_\mu(F)$ can be replaced by $O(\log N \cdot \mathsf{sdp}_\mu(F))$ and there is a randomized algorithm computing a solution x to $\lambda_\mu(F)$ with runtime $O(\mathrm{poly}(nmN))$.*

Theorem 6.2.5 (SDP Relaxation Guarantees, General Case [Yos19]). *Let $F = \{F_e\}_{e \in E}$ be a nonnegative, normalized submodular transformation on (V, E), with $n := |V|$ and $m := |E|$. Let $\mu : V \to \mathbb{R}^+$ be the measure given in (6.2).*

- *For any $\varepsilon > 0$,*

$$\mathsf{sdp}_\mu(F) \leq \lambda_\mu(F) \leq O\left(\left(\frac{\log n}{\varepsilon^4} + \frac{\log n \log m}{\varepsilon^2}\right)\mathsf{sdp}_\mu(F) + \varepsilon B^2\right),$$

  where B is the maximum ℓ_2 norm of a point in any of the $\mathcal{B}(F_e)$. Furthermore, there is a randomized algorithm that computes a solution x to $\lambda_\mu(F)$ in Definition 6.1.12, so that $x \perp \mu$ and

$$\frac{\langle x, L_F(x)\rangle}{\sum_{u\in V}\mu(u)x(u)^2} \leq O\left(\left(\frac{\log n}{\varepsilon^4} + \frac{\log n \log m}{\varepsilon^2}\right)\mathsf{sdp}_\mu(F) + \varepsilon B^2\right),$$

  and the runtime of the algorithm is $O(\mathrm{poly}(nm)^{\mathrm{poly}(1/\varepsilon)})$.

- *Suppose that the number of extreme points of each $\mathcal{B}(F_e)$ is upper bounded by N. Then, the above upper bound on $\lambda_\mu(F)$ can be replaced by $O((\log^2 N + \log m \log N) \cdot \mathsf{sdp}_\mu(F))$ and there is a randomized algorithm computing a solution x to $\lambda_\mu(F)$ with runtime $O(\mathrm{poly}(nmN))$.*

6.3 First Approach: Distribution on Extreme Points

In this section, we describe our first approach for defining reweighted eigenvalues on submodular transformations. It is based on the idea that every point $y \in \mathcal{B}(f)$ can be written as a convex combination of the extreme points of $\mathcal{B}(f)$.

6.3.1 ℓ_1 Program and Reweighting

Let $F = \{F_e\}_{e \in E}$ be a nonnegative, normalized submodular transformation on (V, E), so that $F_e(S) \geq 0$ for all $S \subseteq V$ and $e \in E$. Given $\mu : V \to \mathbb{R}^+$, the μ-expansion of F in Definition 6.1.3 can be written as the following integer program:

$$\phi_\mu(F) = \min_{x \in \{0,1\}^V} \frac{\sum_{e \in E} \tilde{F}_e(x)}{\min(\sum_{u \in V} \mu(u)x(u), \sum_{u \in V} \mu(u)(1 - x(u)))}.$$

Replacing the denominator with the normalization constraints

$$\sum_{u \in V} \mu(u)x(u) = 0 \qquad \text{and} \qquad \sum_{u \in V} \mu(u)|x(u)| = 1,$$

and relaxing the codomain of x from $\{0, 1\}$ to $\mathbb{R}$, we obtaining the following fractional program:

Definition 6.3.1 (ℓ_1 Program)**.** *Let $F = \{F_e\}_{e \in E}$ be a nonnegative, normalized submodular transformation on (V, E) and μ be a measure on V. Define the $\eta_\mu(F)$ program as*

$$\begin{aligned} \eta_\mu(F) := \quad & \min_{x \in \mathbb{R}^V} \sum_{e \in E} \tilde{F}_e(x) \\ \text{subject to} \quad & \sum_{u \in V} \mu(u)x(u) = 0 \\ & \sum_{u \in V} \mu(u)|x(u)| = 1. \end{aligned}$$

The following lemma asserts that $\eta_\mu(F)$ are $\phi_\mu(F)$ are up to constants the same.

Lemma 6.3.2. *Let $F = \{F_e\}_{e \in E}$ be a nonnegative, normalized submodular transformation on (V, E) and let μ be a measure on V. Then,*

$$\frac{\phi_\mu(F)}{2} \leq \eta_\mu(F) \leq \phi_\mu(F).$$

Proof. We first prove the "easy direction" that $\eta_\mu(F) \le \phi_\mu(F)$. Let $S \subseteq V$ such that $\phi_\mu(S) = \phi_\mu(F)$; that is,

$$\frac{\sum_{e\in E} F_e(S)}{\min(\mu(S), \mu(S^c))} = \phi_\mu(F).$$

Implicit is the assumption that $\mu(S) > 0$ and $\mu(S^c) > 0$. Define $x \in \mathbb{R}^V$ such that

$$x(u) := \begin{cases} \frac{1}{2\mu(S)}, & \text{if } u \in S; \\ \frac{-1}{2\mu(S^c)}, & \text{otherwise.} \end{cases}$$

Check that x is feasible for the $\eta_\mu(F)$ program, and that

$$\tilde{F}_e(x) = \tilde{F}_e\left(x + \frac{1}{2\mu(S^c)}\mathbb{1}\right) = \left[\frac{1}{2\mu(S)} + \frac{1}{2\mu(S^c)}\right] F_e(S) \le \frac{F_e(S)}{\min(\mu(S), \mu(S^c))}.$$

Here, we used the third and fourth properties in Proposition 6.1.9. So, $\eta_\mu(F) \le \phi_\mu(F)$.

Next, let's prove the "hard" direction that $\phi_\mu(F)/2 \le \eta_\mu(F)$. A direct proof is not difficult, but we simply do a reduction to the threshold rounding step in the proof of Theorem 6.2.1. To this end, given a feasible $x : V \to \mathbb{R}$ to $\eta_\mu(F)$, we just need to show that $h(u) := x(u) - c$, where c is a μ-weighted median of x, satisfies

$$\sum_{u\in V} \mu(u)|h(u)| \ge \frac{1}{2}.$$

(Note that $\tilde{F}_e(h) = \tilde{F}_e(x)$ by Proposition 6.1.9.) But this is because

$$S_x^+ := \sum_{u:x(u)>0} \mu(u)|x(u)| = \frac{1}{2} = \sum_{u:x(u)<0} \mu(u)|x(u)| =: S_x^-$$

by the constraints on x, and either $S_h^+ \ge S_x^+$ or $S_h^- \ge S_x^-$ depending on the sign of c, so

$$\sum_{u\in V} \mu(u)|h(u)| = S_h^+ + S_h^- \ge \min(S_x^+, S_x^-) = \frac{1}{2}.$$

Therefore, the hard direction follows from step 2 in the proof of Theorem 6.2.1. □

Now we introduce the reweighted ℓ_1 program. The Lovász extension takes the form

$$\tilde{F}_e(x) = \max_{y\in\mathcal{B}(F_e)} \langle y, x\rangle,$$

which we replace by

$$\tilde{F}_e(x) = \max_{\{\lambda_{e,i}\}} \lambda_{e,i} \langle y_{e,i}, x \rangle$$

where $\{y_{e,i} : i \in \mathcal{I}_e\}$ is the set of extreme points of $\mathcal{B}(F_e)$, and $\lambda_{e,i}$ is the weight assigned to the extreme point $y_{e,i}$, so that $\lambda_{e,i} \geq 0$ and $\sum_{i \in \mathcal{I}_e} \lambda_{e,i} \leq 1$.

To make it possible to lift to ℓ_2^2 later, we should have $|\langle y_{e,i}, x \rangle|$ instead of the non-symmetric $\langle y_{e,i}, x \rangle$ (in the sense that replacing x by $-x$ changes the result). Adding absolute value blows up the objective value, however. Our solution is to add a global "balance" constraint that

$$\sum_{(e,i) \in \mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0},$$

where $\mathcal{I} := \{(e,i) \,|\, e \in E, i \in \mathcal{I}_e\}$. Thus we arrive at the following reweighted program.

Definition 6.3.3 (Reweighted ℓ_1 Program)**.** *Let $F = \{F_e\}_{e \in E}$ be a nonnegative, normalized submodular transformation on (V, E) and μ be a measure on V. Let $\{y_{e,i}\}_{i \in \mathcal{I}_e}$ be an enumeration of the extreme points of $\mathcal{B}(F_e)$, and write $\mathcal{I} := \{(e,i) \,|\, e \in E, i \in \mathcal{I}_e\}$. Define the $\eta^*_\mu(F)$ program as*

$$\begin{aligned}
\min_{x \in \mathbb{R}^V} \max_{\{\lambda_{e,i}\}} \quad & \sum_{(e,i) \in \mathcal{I}} \lambda_{e,i} \, |\langle y_{e,i}, x \rangle| \\
\text{subject to} \quad & \sum_{u \in V} \mu(u) x(u) = 0 \\
& \sum_{u \in V} \mu(u) |x(u)| = 1. \\
& \lambda_{e,i} \geq 0 \qquad \forall (e,i) \in \mathcal{I} \\
& \sum_{i \in \mathcal{I}_e} \lambda_{e,i} \leq 1 \qquad \forall e \in E \\
& \sum_{(e,i) \in \mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0}.
\end{aligned}$$

In other words, $\{\lambda_{e,i}\}$ is a distribution over the extreme points $y_{e,i} \in \mathcal{E}(B_e)$ for all $e \in E$.

Example 6.3.4 (Balance Constraints for Directed Graphs)**.** *Let $G = (V, E)$ be a directed graph, and let F_e be the directed edge cut function in Proposition 6.1.1(iii) for $e \in E$. We have established in Example 6.1.7 that the extreme points of $\mathcal{B}(F_e)$ are $\mathbb{1}_u - \mathbb{1}_v$ and $\vec{0}$. Verify that the balance constraints*

$$\sum_{(e,i) \in \mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0}$$

is equivalent to assigning weights in $[0, 1]$ *to each edge, so that in the reweighted graph, the total incoming weight is the same as the total outgoing weight at each vertex, i.e. that the reweighted graph is Eulerian.*

We now show that $\eta^*_\mu(F)$ and $\phi_\mu(F)$ are equivalent up to a constant factor.

Proposition 6.3.5 (Easy direction for $\eta^*_\mu(F)$)**.** *Let* $F = \{F_e\}_{e\in E}$ *be a nonnegative, normalized submodular transformation on* (V, E) *and* μ *be a measure on* V*. Then,*

$$\eta^*_\mu(F) \le 2\phi_\mu(F).$$

Proof. We will prove the stronger assertion that $\eta^*_\mu(F) \le 2\eta_\mu(F)$, and then the proposition follows from Lemma 6.3.2. This amounts to showing that

$$\frac{1}{2}\max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}|\langle y_{e,i}, x\rangle| \le \max_{\{\lambda'_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda'_{e,i}\langle y_{e,i}, x\rangle$$

for any feasible x, where $\{\lambda_{e,i}\}$ are subject to the constraints of the $\eta^*_\mu(F)$ program and $\{\lambda'_{e,i}\}$ are subject to the constraints of the $\eta_\mu(F)$ program, i.e. without the balance constraints.

Let $\{\lambda_{e,i}\}$ be optimal for LHS for a particular feasible x. Our goal is to find a feasible $\{\lambda'_{e,i}\}$ for RHS with a comparable objective value. Writing

$$H^+_x := \{y \in \mathbb{R}^V : \langle y, x\rangle > 0\} \qquad \text{and} \qquad H^-_x := \{y \in \mathbb{R}^V : \langle y, x\rangle \le 0\},$$

we simply take $\lambda'_{e,i} = \lambda_{e,i}$ if $y_{e,i} \in H^+_x$ and $\lambda'_{e,i} = 0$ otherwise. Such choice of $\lambda'_{e,i}$ is easily seen to be feasible. we can bound the objective value on LHS as follows:

$$\begin{aligned}
\frac{1}{2}\sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}|\langle y_{e,i}, x\rangle| &= \frac{1}{2}\left[\sum_{y_{e,i}\in H^+_x} \lambda_{e,i}\langle y_{e,i}, x\rangle + \sum_{y_{e,i}\in H^-_x} \lambda_{e,i}\langle y_{e,i}, -x\rangle\right] \\
&\overset{(*)}{=} \sum_{y_{e,i}\in H^+_x} \lambda'_{e,i}\langle y_{e,i}, x\rangle \\
&= \sum_{(e,i)\in\mathcal{I}} \lambda'_{e,i}\langle y_{e,i}, x\rangle \qquad (\because \lambda'_{e,i} = 0 \text{ for } y_{e,i} \notin H^+_x)
\end{aligned}$$

where $(*)$ holds because the balance constraint implies that

$$\left\langle \sum_{y_{e,i}\in H^+_x} \lambda_{e,i}y_{e,i}, x\right\rangle = \left\langle \sum_{y_{e,i}\in H^-_x} \lambda_{e,i}y_{e,i}, -x\right\rangle.$$

Combining this and Lemma 6.3.2, we are done. ∎

To prove the "hard direction", we construct the dual program of $\eta^*_\mu(F)$ and utilize the extra information from the dual variables for rounding. The same approach was used in Section 5.7 for rounding ℓ_1 programs for directed expansions.

Given a feasible $x : V \to \mathbb{R}$ to the $\eta^*_\mu(F)$ program, the inner maximization program is

$$\begin{aligned} \eta^*_\mu(x) := \quad & \max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \left|\langle y_{e,i}, x\rangle\right| \\ \text{subject to} \quad & \lambda_{e,i} \geq 0 \qquad \forall (e,i) \in \mathcal{I} \\ & \sum_{i\in\mathcal{I}_e} \lambda_{e,i} \leq 1 \qquad \forall e \in E \\ & \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0}. \end{aligned}$$

By taking the LP dual of the above, we obtain the following dual program.

Proposition 6.3.6 (Dual Program)**.** *Let $F = \{F_e\}_{e\in E}$ and $\mu : V \to \mathbb{R}^+$ be as before, and let $x : V \to \mathbb{R}$ be a feasible solution to the $\eta^*_\mu(F)$ program. The following program has objective $\eta^*_\mu(x)$:*

$$\begin{aligned} \xi^*_\mu(x) := \quad & \min_{q\in\mathbb{R}^E_{\geq 0}, r\in\mathbb{R}^V} \quad \sum_e q(e) \\ & \text{subject to} \quad \left|\langle y_{e,i}, x\rangle\right| \leq q(e) + \langle r, y_{e,i}\rangle \qquad \forall e \in E, y_{e,i} \in \mathcal{E}(B_e). \end{aligned}$$

Proof. This is by standard LP duality, where $q(e)$ is the dual variable for the constraint that $\sum_{i\in\mathcal{I}_e} \lambda_{e,i} \leq 1$, and $r(u)$ for $u \in V$ is the dual variable for the constraint that the u-coordinate of $\sum_{e,i} \lambda_{e,i} y_{e,i}$ equals zero. □

Remark 6.3.7. *By appropriately defining F, the dual program here recovers the inner minimization programs in Lemma 5.7.5 and Lemma 5.7.6.*

Proposition 6.3.8 (Hard direction for $\eta^*_\mu(F)$)**.** *Let $F = \{F_e\}_{e\in E}$ and $\mu : V \to \mathbb{R}^+$ be as before. Then,*

$$\frac{\phi_\mu(F)}{8} \leq \eta^*_\mu(F).$$

Proof. Let $x^* \in \mathbb{R}^V$ be an optimal solution to the $\eta^*_\mu(F)$ program in Definition 6.3.3 that $\eta^*_\mu(x^*) = \eta^*_\mu(F)$. Then, by Proposition 6.3.6 the objective value of the following dual program is $\eta^*_\mu(F)$:

$$\begin{array}{lll} \min_{q\in\mathbb{R}^E_{\geq 0}, r\in\mathbb{R}^V} & \sum_e q(e) & \\ \text{subject to} & |\langle y_{e,i}, x^*\rangle| \leq q(e) + \langle y_{e,i}, r\rangle & \forall e\in E, y_{e,i}\in\mathcal{E}(B_e). \end{array}$$

Note that regardless of r, the optimal value for $q(e)$ would be

$$q(e) = \max\left(0, \max_{i\in\mathcal{I}_e}(|\langle y_{e,i}, x^*\rangle| - \langle y_{e,i}, r\rangle)\right).$$

Let (q^*, r^*) be the optimal solution to the dual program above. Following the proof in Section 5.7.3 for directed graphs, define

- $g_1(u) := \max\{0, x^*(u) + r^*(u) - c_1\}$
- $g_2(u) := \max\{0, x^*(u) - r^*(u) - c_2\}$
- $g_3(u) := \max\{0, -x^*(u) + r^*(u) + c_2\}$
- $g_4(u) := \max\{0, -x^*(u) - r^*(u) + c_1\}$,

where c_1 is a μ-weighted median of $x^* + r^*$ and c_2 is a μ-weighted median of $x^* - r^*$, so the support of each g_i has μ-measure at most $\mu(V)/2$. For $t\in\mathbb{R}$ define the threshold sets

$$S_t(z) := \{u\in V : z(u) > t\}.$$

To bound the numerator, first consider g_1. Order the elements in V such that

$$(-x^* - r^*)(u_1) \geq (-x^* - r^*)(u_2) \geq \cdots \geq (-x^* - r^*)(u_{|V|}).$$

Writing y_{e,i^*} for the extreme point in $\mathcal{B}(F_e)$ so that $\langle y_{e,i}, -x^* - r^*\rangle$ is maximized, we have

$$\begin{aligned} \int_0^\infty F_e(S_t(g_1))\,dt &= \sum_{i<|V|} F_e(\{u_1,\ldots,u_i\})(g_1(u_i) - g_1(u_{i+1})) \\ &\leq \sum_{i<|V|} F_e(\{u_1,\ldots,u_i\})(r^*(u_{i+1}) - r^*(u_i) + x^*(u_{i+1}) - x^*(u_i)) \\ &= \langle y_{e,i^*}, -x^*\rangle - \langle y_{e,i^*}, r^*\rangle \qquad \text{(by Proposition 6.1.5)} \\ &\leq |\langle y_{e,i^*}, x^*\rangle| - \langle y_{e,i^*}, r^*\rangle \leq q^*(e). \end{aligned}$$

As for g_4, we have

$$\begin{aligned}\int_0^\infty F_e(S_t^c(g_4))\,dt &= \sum_{i<|V|} F_e(\{u_{i+1},\dots,u_{|V|}\}^c)(g_4(u_{i+1})-g_4(u_i))\\ &\le \sum_{i<|V|} F_e(\{u_1,\dots,u_i\})(r^*(u_{i+1})-r^*(u_i)+x^*(u_{i+1})-x^*(u_i))\\ &\le q^*(e).\end{aligned}$$

Similar for g_2 and g_3 with the vertex ordering satisfying

$$(x^*-r^*)(u_1)\ge(x^*-r^*)(u_2)\ge\cdots\ge(x^*-r^*)(u_{|V|}).$$

The denominator bound is proven the same way as in Section 5.7.3. As $\mathbb{1}\perp y_{e,i}$ for all $y_{e,i}\in\mathcal{E}(B_e)$, we may shift r^* by an appropriate multiple of $\mathbb{1}$ so that the dual solution is still feasible with the same objective value, and $\sum_{u\in V}\mu(u)r(u)=0$. Following the argument in Section 5.7.3, we conclude that

$$\sum_{i=1}^4\sum_{u\in V}\mu(u)g_i(u)\ge\frac{1}{2}.$$

Therefore, there exists $g=g_i$ for some $1\le i\le 4$, such that if we use this function for threshold rounding,

- $\int_0^\infty \min\left\{F(S_t),F(S_t^c)\right\}dt\le\sum_{e\in E}q^*(e)=\eta_\mu^*(F)$;
- $\int_0^\infty \mu(S_t)\,dt\ge 1/8$;
- $\mu(S_t)\le\mu(V)/2$ always.

By the averaging argument, we can return some $S=S_t$ or S_t^c, where

$$\min\left(\phi_\mu(S),\phi_\mu(S^c)\right)=\frac{\min\left\{F(S),F(S^c)\right\}}{\min\left\{\mu(S),\mu(S^c)\right\}}\le 8\eta_\mu^*(F).$$

This finishes the proof of the hard direction. □

Combining Lemma 6.3.2, Proposition 6.3.5, and Proposition 6.3.8, we are able to draw a neat conclusion about the three main ℓ_1 quantities in this section.

Theorem 6.3.9 (All ℓ_1 quantities are equivalent)**.** *Let F and μ be as before. Then,*

$$\phi_\mu(F) \asymp \eta_\mu(F) \asymp \eta^*_\mu(F)$$

Question 6.3.10. *Is there a direct proof showing that $\eta_\mu(F) \asymp \eta^*_\mu(F)$?*

In view of Theorem 6.3.9, if we can find either $\eta_\mu(F)$ or $\eta^*_\mu(F)$ in polynomial time, we would be able to obtain a constant-factor approximation to $\phi_\mu(F)$ in polynomial time. However, this is NP-hard assuming the small-set expansion hypothesis, for example because of the hardness result in the special case of vertex expansion [LRV13], which the current setting captures.

This wraps up our discussion about the ℓ_1 programs. To summarize, we wrote a natural fractional relaxation $\eta_\mu(F)$ of the combinatorial quantity $\phi_\mu(F)$ that we are interested in. The Lovász extension $\tilde{F}_e(x)$ is interpreted as choosing a vector in $\mathcal{B}(F_e)$ that maximizes inner product with x. To derive the reweighted program $\eta^*_\mu(F)$, we impose the constraint that the vectors chosen from each $\mathcal{B}(F_e)$ sum to $\vec{0}$, among other modifications. We were able to show that the three quantities are within a constant factor of each other; see Theorem 6.3.9. Unfortunately, it does not really improve our situation, as we do not know how to solve these programs efficiently. Nevertheless, we will show that a morally identical procedure when applied to "ℓ_2^2 programs" will allow us to obtain SDP's and in particular recover the λ^*_2 programs in the previous chapters.

6.3.2 ℓ_2^2 Program and Reweighting

We now design a reweighted ℓ_2^2 program by lifting the reweighted ℓ_1 program to ℓ_2^2.

Definition 6.3.11 (Reweighted ℓ_2 program)**.** *Let $F = \{F_e\}_{e\in E}$ be a nonnegative, normalized submodular transformation on (V, E) and μ be a measure on V. Let $\{y_{e,i}\}_{i\in\mathcal{I}_e}$ be an enumeration of the extreme points of $\mathcal{B}(F_e)$, and write $\mathcal{I} := \{(e, i) \,|\, e \in E, i \in \mathcal{I}_e\}$. Define* *the $\lambda^*_\mu(F)$ program as*

$$\begin{aligned}\lambda^*_\mu(F) :=\quad & \min_{X\succeq 0}\max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}\langle X, y_{e,i}y_{e,i}^T\rangle \\ \text{subject to}\quad & \langle X, \mu\mu^T\rangle = 0 \\ & \langle X, \operatorname{diag}(\mu)\rangle = 1 \\ & \lambda_{e,i} \geq 0 \qquad \forall (e,i)\in\mathcal{I} \\ & \sum_{i\in\mathcal{I}_e}\lambda_{e,i} \leq 1 \qquad \forall e\in E \\ & \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}y_{e,i} = \vec{0}.\end{aligned}$$

Here is how to lift the ℓ_1 reweighted program in Definition 6.3.3 to ℓ_2^2:

1. Keep all linear constraints as-is and replace the terms inside absolute values by their squares, i.e. $|\langle y_{e,i}, x\rangle|$ becomes $\langle y_{e,i}, x\rangle^2$ and $|x(u)|$ becomes $x(u)^2$.
2. Express everything about $x \in \mathbb{R}^V$ in terms of $X = xx^T \in \mathbb{R}^{V\times V}$. Concretely, $\langle y_{e,i}, x\rangle^2$ becomes $\langle X, y_{e,i}y_{e,i}^T\rangle$, and the constraints about x becomes the constraints about X, both of which are linear in X.
3. The current domain of X is the set of rank-1 PSD matrices, which is not convex. To remedy this, relax it to the set of all PSD matrices.

According to Proposition 6.1.1, submodular transformations capture all the cut functions in generalized graphs. We shall see that the reweighted ℓ_2^2 program in the preceding definition captures the reweighted eigenvalues for these cut functions as special cases, for an appropriately defined volume measure.

Example 6.3.12 (Reweighted Eigenvalue for Undirected Vertex Expansion)**.** *Let $G = (V,E)$ be an undirected graph. Let $F = \{F_v\}_{v\in V}$ be the submodular transformation so that each F_v is the vertex cut function in Proposition 6.1.1(ii). Let $\mu(u) \equiv 1$. We shall show that $\lambda^*_\mu(G)$ is equivalent to the reweighted second eigenvalue for vertex expansion of G in Definition 1.1.1.*

As a special case of Example 6.1.7, it can be seen that the extreme points of F_v are $y_{u,u} := \vec{0}$ and $y_{v,u} := \mathbb{1}_u - \mathbb{1}_v$ for $u \in \partial v$, and so the Lovász extension is

$$\tilde{F}_v(x) = \max_{u\in\{v\}\cup\partial v}(x(u) - x(v)).$$

Since the objective of $\lambda^*_\mu(F)$ is linear in both X and $\{\lambda_{e,i}\}$, and the domains are convex and compact, we can swap the min and the max thanks to von Neumann's minimax theorem Theorem 2.8.1. Regarding X as the Gram matrix of $f : V \to \mathbb{R}^n$ (see Section 2.4), so that $X(u,v) = \langle f(u), f(v)\rangle$, we have

$$\begin{aligned}\min_{X\succeq 0}\max_{\{\lambda_{v,u}\}}\sum_{(v,u)\in\mathcal{I}}\lambda_{v,u}\langle X, y_{v,u}y_{v,u}^T\rangle &= \max_{\{\lambda_{v,u}\}}\min_{X\succeq 0}\sum_{(v,u)\in\mathcal{I}}\lambda_{v,u}\langle X, y_{v,u}y_{v,u}^T\rangle \\ &= \max_{\{\lambda_{v,u}\}}\min_{f:V\to\mathbb{R}^n}\sum_{(v,u)\in\mathcal{I}}\lambda_{v,u}\left\|f(u)-f(v)\right\|_2^2.\end{aligned}$$

The first two constraints of $\lambda^*_\mu(F)$ are equivalent to

$$\sum_{u\in V} f(u) = \vec{0} \qquad \textit{and} \qquad \sum_{u\in V}\|f(u)\|_2^2 = 1, \qquad (*)$$

and the remaining constraints of $\lambda^*_\mu(F)$ on $\lambda_{v,u}$ are equivalent to the degree constraints that

$$1 = \sum_{u:uv\in E}\lambda_{v,u} = \sum_{u:uv\in E}\lambda_{u,v} \quad \forall v\in V.$$

(We can assume without loss that $\sum_{u:uv\in E}\lambda_{v,u} = 1$ since $\vec{0}$ is one of the extreme points of $\mathcal{B}(F_v)$. This is akin to adding self loops to the graph as was done in Chapter 4.) Then, the program in Definition 1.1.1 is obtained by setting $P(u,v) = (\lambda_{u,v} + \lambda_{v,u})/2$ and rewriting the maximum spectral gap objective using

$$\max_P 1 - \alpha_2(P) = \max_P \min_{f:V\to\mathbb{R}^n}\sum_{u,v\in V} P(u,v)\left\|f(u)-f(v)\right\|^2,$$

with f subject to the constraints in $(*)$ and P is subject to the degree constraints that

$$\sum_{u:uv\in E} P(u,v) = 1$$

and the reversibility constraints that $P(u,v) = P(v,u)$. That $P(u,v) = 0$ if $uv \notin E$ is implied, because there is no $\lambda_{u,v}$ for $uv \notin E$. Therefore, the objectives and the constraints of the two programs are equivalent.

Example 6.3.13 (Reweighted Eigenvalue for Directed Edge Conductance)**.** *Suppose that a directed graph $G = (V,E)$ is given. Consider the submodular transformation $\{F_e\}_{e\in E}$ where each F_e is the directed edge cut function in Proposition 6.1.1(iii), and $\mu(u)$ equals* *the total degree of u in G. We shall show that $\lambda_\mu^*(F)$ is equivalent to the reweighted second eigenvalue for directed edge conductance of G in Definition 5.1.7. Again, by Example 6.1.7 the extreme points of $\mathcal{B}(F_e)$ are $\vec{0}$ and $y_{u,v} := \mathbb{1}_u - \mathbb{1}_v$ for $e = uv$, and the Lovász extension of F_e is*

$$\tilde{F}_e(x) = \max_{y \in \mathcal{B}(F_e)} \langle y, x \rangle = \max(0, x(u) - x(v)).$$

The first two constraints of $\lambda_\mu^*(F)$ are equivalent to

$$\sum_{u \in V} \mu(u) f(u) = \vec{0} \qquad \textit{and} \qquad \sum_{u \in V} \mu(u) \left\| f(u) \right\|_2^2 = 1, \qquad (*)$$

and the balance constraint that

$$\sum_{uv \in E} \lambda_{u,v} y_{u,v} = \vec{0}$$

is equivalent to the Eulerian constraints on the $\lambda_{u,v}$-reweighted graph that

$$\sum_{u: uv \in E} \lambda_{u,v} = \sum_{u: vu \in E} \lambda_{v,u} \qquad \forall v \in V.$$

Then, the program in Definition 5.1.7 is obtained via the equivalent formulation in Proposition 5.4.2 by setting $A(u,v) = \lambda_{u,v}$.

Analogous results hold for ordinary graph conductance, vertex expansion in directed graphs, and conductance in undirected hypergraphs Proposition 6.1.1 that we have seen in previous chapters, with appropriately defined measures μ. We omit the derivations for brevity. We also remark that, by appropriately modifying the definitions of the expansion $\phi_\mu(F)$ and the reweighted eigenvalue $\lambda_\mu^*(F)$, we can deal with weighted graphs as well, but we choose not to do so here for simplicity.

Unlike in [Yos19] where the SDP is indeed a relaxation of $\lambda_\mu(F)$, for which a Cheeger inequality is known, we do not know of any general relation between the reweighted eigenvalue $\lambda_\mu^*(F)$ and the nontrivial eigenvalue $\lambda_\mu(F)$. This is because $\lambda_\mu^*(F)$ is constructed from the reweighted ℓ_1 program instead of from $\lambda_\mu(F)$. Even in the special case where F is symmetric, $\lambda_\mu^*(F)$ and $\lambda_\mu(F)$ are not the same.

Remark 6.3.14 (Reweighting in Symmetric Case)**.** *Consider the reweighted ℓ_2^2 program $\lambda_\mu^*(F)$ where F is symmetric. By Proposition 6.1.5 the base polytopes $\mathcal{B}(F_e)$ are symmetric as well. Write $y_{e,-i}$ for $-y_{e,i}$, which is also an extreme point in $\mathcal{B}(F_e)$. We can mandate* *that the weights $\lambda_{e,i}$ satisfy the stronger constraints that $\lambda_{e,i} = \lambda_{e,-i}$ for all $(e,i) \in \mathcal{I}$, so that the balance constraints*

$$\sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0}$$

are satisfied. Then, we are free from the global balance constraints in the reweighted program, and we can directly maximize the energy of each submodular function by choosing $\lambda_{e,i^*} = \lambda_{e,-i^*} = 1/2$ where y_{e,i^*} is the maximizer for $\langle y_{e,i}, x\rangle$. We establish that

$$\begin{aligned}
\lambda^*_\mu(F) &= \min_{f:V\to\mathbb{R}^n} \max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \left\| \sum_{u\in V} y_{e,i}(u) f(u) \right\|^2 \\
&= \max_{\{\lambda_{e,i}\}} \min_{f:V\to\mathbb{R}^n} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \left\| \sum_{u\in V} y_{e,i}(u) f(u) \right\|^2 \\
&= \max_{\{\lambda_{e,i}\}} \min_{x:V\to\mathbb{R}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \sum_{j\in[n]} \langle y_{e,i}, x\rangle^2,
\end{aligned}$$

where $\lambda_{e,i}$ are subject to the balance constraints of $\lambda^*_\mu(F)$, and f and x are subject to their usual normalization constraints. The first equality is by definition, the second equality is by von Neumann's minimax theorem (Theorem 2.8.1), and the third equality is by observing that the inner minimization problem is independent for each coordinate of $f : V \to \mathbb{R}$.

If we could swap the max and the min one more time and apply the above argument about the choice of $\lambda_{e,i}$, then $\lambda^*_\mu(F)$ and $\lambda_\mu(F)$ would be equal. However, the domain for x corresponds to an additional constraint on the Gram matrix $X(u,v) := x(u)x(v)$ that it be of rank one, which makes the domain non-convex. Therefore, even in the case where F is symmetric, the difference between the max-min program and the min-max program explains the difference between the reweighted eigenvalue $\lambda^*_\mu(F)$ and the nontrivial eigenvalue $\lambda_\mu(F)$.

In the next few subsections, we attempt to derive a Cheeger inequality relating the expansion $\phi_\mu(F)$ and the reweighted eigenvalue $\lambda^*_\mu(F)$ of a submodular transformation F. The easy direction consists of showing that $\lambda^*_\mu(F)$ is still a relaxation of the expansion $\phi_\mu(F)$, and the hard direction amounts to exhibiting and analyzing a rounding algorithm from a solution to $\lambda^*_\mu(F)$ to a small-expansion subset.

6.3.3 Easy Direction

We first prove the following easy direction that the expansion of a submodular transformation upper bounds its reweighted eigenvalue.

Proposition 6.3.15. *Let $F = \{F_e\}$ be a nonnegative, normalized submodular transformation on (V, E) so that each F_e has range $\subseteq [0,1]$. Let $\mu : V \to \mathbb{R}^+$ be a measure on V. Then,*

$$\lambda^*_\mu(F) \leq 2\phi_\mu(F).$$

Proof. Let $S \subseteq V$ be such that

$$\phi_\mu(S) = \frac{F(S)}{\min(\mu(S), \mu(S^c))} = \phi_\mu(F).$$

Consider the following vector solution to $\lambda^*_\mu(F)$ where $X(u, v) = \langle f(u), f(v)\rangle$:

$$f(v) := \begin{cases} \frac{1}{\mu(S)}, & \text{if } v \in S \\ \frac{-1}{\mu(S^c)}, & \text{otherwise.} \end{cases}$$

Note that $f(v)$ are one-dimensional and X is of rank one. Moreover,

$$\langle y_{e,i}, f\rangle^2 \leq |\langle y_{e,i}, f\rangle| \cdot \|y_{e,i}\|_1 \cdot \|f\|_\infty \leq |\langle y_{e,i}, f\rangle| \cdot \frac{2}{\min(\mu(S), \mu(S^c))},$$

where the last inequality uses the definition of f and the maximum 1-norm bound on the base polytope in Proposition 6.1.5. Now we may use the balance constraint to infer that

$$\begin{aligned} \frac{1}{2} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} |\langle y_{e,i}, f\rangle| &= \frac{1}{2} \sum_{\substack{(e,i)\in\mathcal{I} \\ \langle y_{e,i}, f\rangle > 0}} \lambda_{e,i} \langle y_{e,i}, f\rangle + \frac{1}{2} \sum_{\substack{(e,i)\in\mathcal{I} \\ \langle y_{e,i}, f\rangle \leq 0}} \lambda_{e,i} \langle -y_{e,i}, f\rangle \\ &= \sum_{\substack{(e,i)\in\mathcal{I} \\ \langle y_{e,i}, f\rangle > 0}} \lambda_{e,i} \langle y_{e,i}, f\rangle \\ &\leq \sum_{e\in E} \tilde{F}_e(f) = \left(\frac{1}{\mu(S)} + \frac{1}{\mu(S^c)}\right) \cdot F(S) \leq \frac{2F(S)}{\min(\mu(S), \mu(S^c))}. \end{aligned}$$

This completes the proof. ∎

6.3.4 Hard Direction: Overview

Let $X(u, v) = \langle f(u), f(v)\rangle$ be a solution to the $\lambda^*_\mu(F)$ program, where $f : V \to \mathbb{R}^n$. Following previous proofs in Chapter 4 and Chapter 5, the first step is to project the n-dimensional solution to a one-dimensional solution, and the second step is to round the one-dimensional ℓ_2^2 solution to an ℓ_1 solution. After that, we obtain a solution to the ℓ_1 reweighted program $\eta^*_\mu(F)$, to which a threshold rounding procedure per Proposition 6.3.8 yields a set with small expansion.

6.3.5 Projection to One-Dimensional Program

As before, given an n-dimensional solution $f : V \to \mathbb{R}^n$ to the $\lambda_\mu^*(F)$ program, we project to a one-dimensional solution using random Gaussian projection by setting $x(v) = \langle f(v), z\rangle$ where $z \sim N(0,1)^n$ is a random Gaussian vector. We do not have a simple condition that bounds the projection loss in the general case. Define the one-dimensional program $\lambda_\mu^{(1)}(F)$ as the n-dimensional program in Definition 6.3.11, but where X is required to be of rank one. In terms of embedding, $\lambda_\mu^{(1)}(F)$ may be formulated as

$$\begin{aligned}
\lambda_\mu^*(F) :=\qquad & \min_{x:V\to\mathbb{R}} \max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}\langle x, y_{e,i}\rangle^2 \\
\text{subject to}\qquad & \sum_{u\in V} \mu(u)x(u) = 0 \\
& \sum_{u\in V} \mu(u)x(u)^2 = 1,
\end{aligned}$$

and where $\lambda_{e,i}$ are subject to the same constraints as in Definition 6.3.11.

Proposition 6.3.16 (Projection Loss)**.** *Let $F = \{F_e\}_{e\in E}$ be a nonnegative, normalized submodular transformation on (V,E). Suppose there is a number $M(F)$ such that*

$$\mathbb{E}_x \left[\max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}(x^T y_{e,i})^2 \right] \le M(F) \cdot \max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \left\| \sum_{u\in V} y_{e,i}(u) f(u) \right\|^2$$

for any $X = (\langle f(u), f(v)\rangle)_{u,v\in V}$ satisfying $\langle X, \mu\mu^T\rangle = 0$ and $\langle X, \operatorname{diag}(\mu)\rangle = 1$, and where $x(v) = \langle f(v), z\rangle$ is the one-dimensional projected solution with $z \sim N(0,1)^n$. Then,

$$\lambda_\mu^{(1)}(F) \lesssim M(F) \cdot \lambda_\mu^*(F).$$

Proof. Let $f : V \to \mathbb{R}^n$ be the n-dimensional solution and X is its Gram matrix. By Fact 2.10.7 and Markov's inequality, with probability at least $1 - 1/24 - 11/12 \ge 1/24$ the projected solution $x : V \to \mathbb{R}$ satisfies

$$\max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}(x^T y_{e,i})^2 \le 24 \cdot M(F) \cdot \max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \left\| \sum_{u\in V} y_{e,i}(u) f(u) \right\|^2$$

and $\sum_{u\in V} \mu(u)x(u)^2 \ge 1/2$. The constraint $\sum_{u\in V} \mu(u)x(u) = 0$ remains true since Π is a linear operator. □

Perhaps the most powerful tool that one might apply to bound the expected maximum of subgaussian random variables is generic chaining à la Talagrand [Tal05], but it is not so easy to use. In some settings, we can again use the Large Optimal Property in Lemma 5.5.5 to find an admissible value for $M(F)$. Using the same proof as in Theorem 5.6.4, one can show that, if F is such that

$$\max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}\langle f, y_{e,i}\rangle^2 \geq \frac{1}{K(F)} \cdot \sum_{(e,i)\in\mathcal{I}} \langle f, y_{e,i}\rangle^2 \tag{6.5}$$

for any $f : V \to \mathbb{R}$ satisfying $f \perp \mu$ and $\sum_{v\in v} \mu(v)f(v)^2 = 1$, then Gaussian projection to $d = O(\log K(F))$ dimensions increases the objective value by at most a constant factor, after which we can select the best coordinate for an additional factor d loss. This shows that we can take $M(F) = O(\log K(F))$.

Example 6.3.17 (Some Symmetric Cases)**.** *In the case where F is symmetric, then by Proposition 6.1.5, if $y_{e,i}$ is an extreme point of $\mathcal{B}(F_e)$ then so is $y_{e,-i} := -y_{e,i}$. Therefore, by taking $\lambda_{e,i} = 1/\max_e |\mathcal{I}_e|$, the balance constraints are satisfied, and we obtain a feasible reweighting. Therefore, we can take $K(F) = \max_e |\mathcal{I}_e|$ and $M(F) = O(\log \max_e |\mathcal{I}_e|)$. This captures the $O(\log r)$ loss for edge conductance in hypergraphs and also, trivially, the $O(1)$ projection loss for the classical setting of edge conductance in undirected graphs.*

Note that, however, the number of extreme points can be exponential in $|V|$. As an example, let f be the function $f(S) := \min(|S|, |S^c|)$ for $S \subseteq V$, which can be verified to be submodular. Suppose $|V|$ is even. The extreme points of $\mathcal{B}(f)$ are then all the vectors in $\{\pm 1\}^V$ whose entries sum to zero. There are $\binom{|V|}{|V|/2} \asymp 2^{|V|}/|V|$ many of them, and so the projection loss upper bound using the number of extreme points is only $O(|V|)$.

Below we demonstrate using one more example from Chapter 5.

Example 6.3.18 (Directed Edge Conductance)**.** *Let $G = (V, E)$ be a directed graph. Consider the submodular transformation $F = \{F_e\}_{e\in E}$ on (V, E) where F_e is the directed edge cut function as in Proposition 6.1.1(iii) for $e \in E$. From Example 6.1.11 we know that the extreme points for $\mathcal{B}(F_e)$ where $e = uv$ are $y_{e,1} := \mathbb{1}_u - \mathbb{1}_v$ and $y_{e,0} := \vec{0}$. By Hoffman's circulation lemma in Lemma 5.5.4, there exist $\alpha_e \in [1, \alpha(G)]$ such that the reweighted graph with arc weights $w'(e) = \alpha_e$ is Eulerian. Here, $\alpha(G)$ is the asymmetric ratio of G as defined in Definition 5.2.1. That means by taking $\lambda_{e,1} = \alpha_e/\alpha(G)$ and $\lambda_{e,0} = 0$ we have*

$$\sum_{i\in I_e} \lambda_{e,i} \leq 1 \;\; \forall e \in E \quad \textit{and} \quad \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0},$$

so that $\{\lambda_{e,i}\}$ *is a feasible solution to the* λ_F^* *program, and*

$$\sum_{(e,i)\in\mathcal{I}} \lambda_{e,i}\langle f, y_{e,i}\rangle^2 \geq \frac{1}{\alpha(G)} \cdot \sum_{(e,i)\in\mathcal{I}} \langle f, y_{e,i}\rangle^2.$$

Therefore, we can take $K(F) = \alpha(G)$ *and the projection loss is thus* $O(\log \alpha(G))$.

We prove in the following a new projection result for directed hypergraphs. The small projection loss corroborates with the earlier remark that directed hypergraph cut functions are "simple".

Proposition 6.3.19 (Projection Loss for Directed Hypergraphs)**.** *Let* $H = (V, E)$ *be a directed hypergraph of rank* r*, and let* $\{F_e\}_{e\in E}$ *be a submodular transformation on* (V, E)*, so that each* F_e *is the hyperedge cut function defined in Proposition 6.1.1(vi). Then,*

$$\lambda_\mu^{(1)}(F) \lesssim \log(r \cdot \alpha(G_H)) \cdot \lambda_\mu^*(F),$$

where $\alpha(G_H)$ *is the asymmetric ratio of the weighted directed graph* $G_H = (V, E_H, w_H)$ *defined as* $E_H = V \times V$ *and* $w_H(uv) = |\{e \in E : u \in e^-, v \in e^+\}|$.

Proof. Referring to (6.5), it suffices to prove that $K(F) \leq \text{poly}(r) \cdot \alpha(G_H)$. Consider the graph G_H. By Hoffman's circulation lemma in Lemma 5.5.4, there exists an assignment $\alpha_{uv} \in [1, \alpha(G_H)]$ of arc weights, such that the reweighted graph with arc weights $w'(uv) = \alpha_{uv} \cdot w_H(uv)$ is Eulerian. By Example 6.1.11, each nonzero extreme point of $\mathcal{B}(F_e)$ can be written as $y_{e,(u,v)} := \mathbb{1}_u - \mathbb{1}_v$ where $u \in e^-$ and $v \in e^+$. Therefore, there are at most $1+|e^-|\cdot|e^+| \leq 1+r^2$ extreme points for each $\mathcal{B}(F_e)$. Setting $\lambda_{e,(u,v)} = \alpha_{uv}/((1+r^2)\cdot\alpha(G_H))$, we have that $\{\lambda_{e,(u,v)}\}$ is a feasible solution to the $\lambda_\mu^*(F)$ program, and

$$\sum_{(e,(u,v))\in\mathcal{I}} \lambda_{e,(u,v)}\langle f, y_{e,(u,v)}\rangle^2 \geq \frac{1}{(1+r^2)\cdot\alpha(G_H)} \cdot \sum_{(e,(u,v))\in\mathcal{I}} \langle f, y_{e,(u,v)}\rangle^2.$$

So, $K(F) \leq \text{poly}(r) \cdot \alpha(G_H)$, and this finishes the proof. □

Open Problem 6.3.20. *In a similar spirit to Lemma 5.5.2 relating vertex expansion and asymmetric ratio of directed graphs, provide an upper bound to the projection loss for directed hypergraphs using a more fundamental quantity than the asymmetric ratio of* G_H.

6.3.6 The ℓ_2^2 to ℓ_1 Step

After projecting to a one-dimensional vector solution, the next step is the ℓ_2^2 to ℓ_1 step. It turns out that the ℓ_2^2 to ℓ_1 step only goes through when F is *simple* (defined in the paragraph before Proposition 6.1.6), and the proof faces a major obstacle in the general case. The most general non-trivial setting that the current proof can hope to capture is directed hypergraph expansion, which does correspond to a simple submodular transformation by Proposition 6.1.6). We first state and prove the positive result, then discuss the difficulties in more general settings.

Lemma 6.3.21 (ℓ_2^2 to ℓ_1 Step for Simple Submodular Transformations)**.** *Let $F = \{F_e\}_{e\in E}$ be a simple, normalized submodular transformation on V. Let $\mu : V \to \mathbb{R}^+$ be a measure on V, and let $x : V \to \mathbb{R}$ be such that $\sum_{v\in V} \mu(v)x(v) = 0$. Suppose that μ satisfies that $\mu(v)$ is at least the number of F_e's influenced by v. Then, there exists $h : V \to \mathbb{R}$ such that $\sum_{v\in V} \mu(v)h(v) = 0$ and*

$$\frac{\max_{\{\lambda_{e,i}\}} \sum_{(e,i)} \lambda_{e,i}|\langle h, y_{e,i}\rangle|}{\sum_{v\in V} \mu(v)|h(v)|} \lesssim \sqrt{\frac{\max_{\{\lambda_{e,i}\}} \sum_{(e,i)} \lambda_{e,i}\langle f, y_{e,i}\rangle^2}{\sum_{v\in V} \mu(v)f(v)^2}},$$

where $\lambda_{e,i}$ on both sides are subject to the constraints that $\lambda_{e,i} \geq 0$, $\sum_{i\in I_e} \lambda_{e,i} \leq 1$, and $\sum_{(e,i)} \lambda_{e,i}y_{e,i} = \vec{0}$.

Proof. The proof basically follows previous ℓ_2^2 to ℓ_1 proofs. Given $f : V \to \mathbb{R}$ such that $\sum_{v\in V} \mu(v)f(v) = 0$, let $c \in \mathbb{R}$ be a μ-weighted median of f and set h in the same way as in (3.1). Then, 0 is again a μ-weighted median of h, and the familiar denominator bound

$$\sum_{v\in V} \mu(v)|h(v)| = \sum_{v\in V} \mu(v)(f(v) - c)^2 \geq \sum_{v\in V} \mu(v)f(v)^2$$

holds. For the numerator, for any $y_{e,i}$, we know by Example 6.1.7 that it is either the zero vector or of the form $\mathbb{1}_u - \mathbb{1}_v$. If it is the zero vector, then $|\langle h, y_{e,i}\rangle| = 0$. Otherwise, assuming without loss of generality that $f(u) \leq f(v)$ we have

$$\begin{aligned} |\langle h, y_{e,i}\rangle| &= |h(u) - h(v)| \\ &\leq (f(u) - f(v))^2 + 2|f(u) - f(v)| \cdot \min_{f(u)\leq t\leq f(v)} |t - c| \\ &= |\langle f, y_{e,i}\rangle|^2 + 2|\langle f, y_{e,i}\rangle| \cdot \min_{f(u)\leq t\leq f(v)} |t - c|, \end{aligned} \tag{6.6}$$

where the inequality is a strengthening of (3.2) and can again be verified by checking the cases where $h(u)$ and $h(v)$ have the same signs and where they have different signs. Let

$$M := \max_{\{\lambda_{e,i}\}} \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \langle f, y_{e,i}\rangle^2$$

and write $u_{e,i}$ and $v_{e,i}$ for u and v for a specific extreme point $y_{e,i}$. For any feasible reweighting $\{\lambda_{e,i}\}$, we have

$$\begin{aligned}
& \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} |\langle h, y_{e,i}\rangle| \\
\leq\ & \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \langle f, y_{e,i}\rangle^2 + 2 \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} |\langle f, y_{e,i}\rangle| \cdot \min_{f(u_{e,i})\leq t\leq f(v_{e,i})} |t-c| \\
\overset{(*)}{\leq}\ & \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \langle f, y_{e,i}\rangle^2 + 2 \sqrt{\sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \langle f, y_{e,i}\rangle^2 \cdot \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \left(\min_{f(u_{e,i})\leq t\leq f(v_{e,i})} |t-c| \right)^2} \\
\leq\ & M + 2\sqrt{M \cdot \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} (f(u_{e,i}) - c)^2} \\
\leq\ & M + 2\sqrt{M \cdot \sum_{u\in V} \mu(u)(f(u)-c)^2},
\end{aligned}$$

where we used the Cauchy-Schwarz inequality in $(*)$ and the definition of μ in the final inequality.

Since F is simple and normalized, by the definition of μ we have

$$F(\emptyset) = \mu(\emptyset) = 0 \qquad \text{and} \qquad F(S) - F(S \setminus \{u\}) \leq \mu(u)$$

for all $S \subseteq V$ and $u \in V \setminus S$. This shows that $\phi_\mu(F) \leq 1$ and by the easy direction in Proposition 6.3.15 we have

$$\frac{M}{\sum_{u\in V} \mu(u) f(u)^2} = \frac{\max_{\{\lambda_{e,i}\}} \sum_{(e,i)} \lambda_{e,i} \langle f, y_{e,i}\rangle^2}{\sum_{u\in V} \mu(u) f(u)^2} \leq O(1).$$

Combining these inequalities,

$$\begin{aligned}
\frac{\max_{\{\lambda_{e,i}\}} \sum_{(e,i)} \lambda_{e,i} |\langle h, y_{e,i}\rangle|}{\sum_{v\in V} \mu(v)|h(v)|} &\le \frac{M}{\sum_{v\in V} \mu(v)|h(v)|} + 2\sqrt{\frac{M}{\sum_{v\in V} \mu(v)|h(v)|}} \\
&\le \frac{M}{\sum_{v\in V} \mu(v) f(v)^2} + 2\sqrt{\frac{M}{\sum_{v\in V} \mu(v) f(v)^2}} \\
&\lesssim \sqrt{\frac{M}{\sum_{v\in V} \mu(v) f(v)^2}},
\end{aligned}$$

from which the result follows. □

Remark 6.3.22 (ℓ_2^2 to ℓ_1 Argument in General Case)**.** *In general, we do not know of a way to construct the ℓ_1 solution h from f, so there there is a relation like (6.6) between $|\langle h, y_{e,i}\rangle|$ and $|\langle f, y_{e,i}\rangle|$, where $y_{e,i}$ is an extreme point of $\mathcal{B}(F_e)$. The high-level explanation is that $y_{e,i}$ may no longer be the extreme point that maximizes $\langle f, y_{e,i}\rangle$, and so cancellations may occur when evaluating the inner product, which do not easily carry from ℓ_2^2 to ℓ_1.*

To give a concrete example, let $y = (1, 1, 1, -1, -1, -1)$ and $f = (3, -1, -2, 1, 1, -2)$. Then $\sum_{u\in V} f(u) = 0$ and $\langle f, y\rangle = 0$. Using the mapping in (3.1) to construct h, we have $h = (9, -1, -4, 1, 1, -4)$, and clearly $|\langle h, y\rangle| > 0$.

Therefore, it seems that for the ℓ_2^2 to ℓ_1 step to go through, we cannot hope to control each individual extreme point, and a more global argument is needed.

6.3.7 Cheeger Inequality for Directed Hypergraph Expansion

We are now ready to prove the first main result of the chapter, which is a Cheeger inequality for directed hypergraphs.

Theorem 6.3.23 (Cheeger Inequality for Directed Hypergraphs)**.** *Let $H = (V, E)$ be a directed hypergraph. Consider its edge conductance which is defined in Section 2.3.3 as*

$$\vec{\phi}_\mu(S) := \frac{\min(|\delta^+(S)|, |\delta^+(S^c)|)}{\min\{\mu(S), \mu(S^c)\}} \quad \textit{and} \quad \vec{\phi}_\mu(H) := \min_{\emptyset \ne S \subset V} \vec{\phi}_\mu(S),$$

where μ is the total degree measure. Let F_e be the submodular cut function defined in Proposition 6.1.1(vi) for each directed hyperedge $e \in E$, and let $F_H = \{F_e\}_{e\in E}$. Then,

$$\frac{\vec{\phi}_\mu(H)^2}{\log(r \cdot \alpha(G_H))} \lesssim \lambda_\mu^*(F_H) \lesssim \vec{\phi}_\mu(H),$$

where $\alpha(G_H)$ is the asymmetric ratio of the weighted directed graph $G_H = (V, E_H, w_H)$ defined as $E_H = V \times V$ and $w_H(uv) = |\{e \in E : u \in e^-, v \in e^+\}|$. Moreover, a cut $S \subseteq V$ with this expansion guarantee can be produced in polynomial time.

Proof. First, we reformulate the conductance problem as the μ-expansion of F_H. Indeed, for

$$F_e(S) := \mathbb{1}[e^- \cap S \neq \emptyset \wedge e^+ \cap S^c \neq \emptyset]$$

as defined in Proposition 6.1.1(vi), we have $F_H(S) = |\delta^+(S)|$, $\vec{\phi}_\mu(S) = \min(\phi_\mu(S), \phi_\mu(S^c))$, and consequently,

$$\phi_\mu(F_H) = \vec{\phi}_\mu(H).$$

Therefore, we need to prove that

$$\frac{\phi_\mu(F_H)^2}{\log(r \cdot \alpha(G_H))} \lesssim \lambda_\mu^*(F_H) \lesssim \phi_\mu(F_H).$$

It is easy to verify that F_H is nonnegative and normalized, and so the easy direction is just Proposition 6.3.15. For the hard direction, given an optimal solution $X(u,v) = \langle f(u), f(v)\rangle$ to $\lambda_\mu^*(F_H)$, we project to a one-dimensional ℓ_2^2 solution $x : V \to \mathbb{R}$ using Proposition 6.3.19 and construct a one-dimensional ℓ_1 solution $h : V \to \mathbb{R}$ from x using Lemma 6.3.21. h satisfies the guarantees that $\sum_{u \in V} \mu(u)h(u) = 0$ and

$$\frac{\max_{\{\lambda_{e,i}\}} \sum_{(e,i)} \lambda_{e,i} |\langle h, y_{e,i}\rangle|}{\sum_{v \in V} \mu(v)|h(v)|} \lesssim \sqrt{\log(r \cdot \alpha(G_H)) \cdot \lambda_\mu^*(F_H)}.$$

Therefore, an appropriately scaled version of h is a feasible solution to the ℓ_1 reweighted program $\eta_\mu^*(F_H)$ in Definition 6.3.3, with objective value $O\left(\sqrt{\log(r \cdot \alpha(G_H)) \cdot \lambda_\mu^*(F_H)}\right)$. Then, the rounding algorithm in Proposition 6.3.8 produces a set $S \subseteq V$ such that

$$\phi_\mu(S) \lesssim \sqrt{\log(r \cdot \alpha(G_H)) \cdot \lambda_\mu^*(F_H)}.$$

Squaring and rearranging, we obtain the hard direction.

Since $\lambda_\mu^*(F_H)$ is an SDP with polynomially many constraints, hence solvable in polynomial time, and that the projection step, the ℓ_2^2 to ℓ_1 step, and the threshold rounding step all take polynomial time, the proof of the hard direction informs a polynomial-time algorithm to produce a cut $S \subseteq V$ with expansion guaranteed by our Cheeger inequality. $\square$

6.3.8 Generalizations of Cheeger Inequality for Directed Hypergraph Expansion

In this section, we discuss generalizations of Cheeger inequality for directed hypergraphs. Since edge conductance in directed graphs is a special case, we do not expect an analogue of bipartite Cheeger inequality nor higher order Cheeger inequality. However, it is indeed possible to prove an analogue of the improved Cheeger inequality with an appropriate definition of higher eigenvalues.

First, we define the k-th reweighted eigenvalue of a submodular transformation, which is defined by replacing the normalization constraints on X with spreading constraints. This is a generalization of the definitions of k-th reweighted eigenvalues in Chapter 4 and Chapter 5.

Definition 6.3.24 (k-th Reweighted Eigenvalue for Submodular Transformations)**.** *Let $F = \{F_e\}_{e\in E}$ be a nonnegative, normalized submodular transformation on (V, E) and μ be a measure on V. Let $\{y_{e,i}\}_{i\in\mathcal{I}_e}$ be an enumeration of the extreme points of $\mathcal{B}(F_e)$, and write $\mathcal{I} := \{(e, i) \,|\, e \in E, i \in \mathcal{I}_e\}$. Define the $(\lambda^*_\mu)_k(F)$ program as*

$$\begin{aligned} \max_{\{\lambda_{e,i}\}} \quad & \lambda_k\left(\sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} y_{e,i} y_{e,i}^T\right) \\ \text{subject to} \quad & \lambda_{e,i} \geq 0 \qquad \forall (e,i) \in \mathcal{I} \\ & \sum_{i\in\mathcal{I}_e} \lambda_{e,i} \leq 1 \qquad \forall e \in E \\ & \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0}, \end{aligned}$$

We leave it as an exercise to verify that $(\lambda^*_\mu)_2(F)$ is equivalent to $\lambda^*_\mu(F)$ defined in Definition 6.3.11.

Below, we prove an improved Cheeger inequality for simple submodular transformations/directed hypergraphs.

Theorem 6.3.25 (Improved Cheeger Inequality for Directed Hypergraphs)**.** *Let $H = (V, E)$ be a directed hypergraph. Let μ be the total degree measure on V. Let F_e be the submodular cut function defined in Proposition 6.1.1(vi) for each directed hyperedge $e \in E$, and $F_H = \{F_e\}_{e\in E}$. Then, for any $2 \leq k \leq n/2$,*

$$\lambda^*_\mu(F_H) \lesssim \vec{\phi}_\mu(H) \lesssim \frac{k \cdot \log(r \cdot \alpha(G_H)) \cdot \lambda^*_\mu(F_H)}{\sqrt{(\lambda^*_\mu)_k(F_H)}},$$

where $\alpha(G_H)$ is the asymmetric ratio of the weighted directed graph $G_H = (V, E_H, w_H)$ defined as $E_H = V \times V$ and $w_H(uv) = |\{e \in E : u \in e^-, v \in e^+\}|$. Moreover, a cut $S \subseteq V$ with this expansion guarantee can be produced in polynomial time.

Proof. As before, $\vec{\phi}_\mu(H) = \phi_\mu(F_H)$. The easy direction is directly from Theorem 6.3.23. For the hard direction, following Section 4.6, we first define a convex relaxation of $(\lambda^*_\mu)_k(F_H)$, which we call $(\sigma^*_\mu)_k(F_H)$ and may be interpreted as the sum of the k-smallest reweighted eigenvalues:

$$\begin{aligned}
\min_{X \succeq 0} \max_{\{\lambda_{e,i}\}} \quad & \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} \langle X, y_{e,i} y_{e,i}^T \rangle \\
\text{subject to} \quad & \operatorname{diag}(\mu)^{\frac{1}{2}} X \operatorname{diag}(\mu)^{\frac{1}{2}} \preceq I_n \\
& \operatorname{tr}\left(\operatorname{diag}(\mu)^{\frac{1}{2}} X \operatorname{diag}(\mu)^{\frac{1}{2}}\right) = k \\
& \lambda_{e,i} \geq 0 \qquad \forall (e,i) \in \mathcal{I} \\
& \sum_{i \in \mathcal{I}_e} \lambda_{e,i} \leq 1 \qquad \forall e \in E \\
& \sum_{(e,i)\in\mathcal{I}} \lambda_{e,i} y_{e,i} = \vec{0},
\end{aligned}$$

We shall then show an improved threshold rounding guarantee as in the proof of Proposition 4.6.1 that:

Proposition 6.3.26. *For any given one-dimensional solution $X(u,v) = \langle x(u), x(v) \rangle$ to the $(\sigma^*_\mu)_k(F_H)$ program with inner maximization objective $M(X)$, where $x : V \to \mathbb{R}$, and a k-step function $y_x : V \to \mathbb{R}$ approximating x (see Definition 3.1.17), we can construct from it a one-dimensional ℓ_1 solution $h : V \to \mathbb{R}$, such that*

$$\frac{\max_{\{\lambda_{e,i}\}} \sum_{(e,i)} \lambda_{e,i} |\langle h, y_{e,i} \rangle|}{\sum_{v \in V} \mu(v) |h(v)|} \lesssim k \cdot M(X) + k \cdot \|x - y_x\|_\mu \cdot \sqrt{M(X)}$$

and $\sum_{u \in V} \mu(u) h(u) = 0$.

The proof of Proposition 6.3.26 is deferred to Appendix B. Next, as in Proposition 4.6.2 we control the quality of the k-step approximation using $(\sigma^*_\mu)_k(F)$:

Proposition 6.3.27. *For any one-dimensional feasible solution X to the $(\sigma^*_\mu)_k(F_H)$ program with inner maximization objective $M(X)$, where $X(u,v) = \langle x(u), x(v) \rangle$ for $x : V \to \mathbb{R}$, there exists a k-step function y approximating x, with*

$$\|x - y\|_\mu^2 \lesssim \frac{k \cdot M(X)}{(\sigma^*_\mu)_k(F_H)}.$$

The proof of Proposition 6.3.27 is also deferred to Appendix B.

Assuming these two propositions, we may prove the hard direction as follows. Let X be an optimal solution to the $(\sigma_\mu^*)_{2k}(F)$ program, where $X(u,v) = \langle f(u), f(v)\rangle$ with $f : V \to \mathbb{R}^n$. Project f to a one-dimensional solution $x : V \to \mathbb{R}$ using Proposition 6.3.19. Construct the $(2k)$-step approximation y_x to x using Proposition 6.3.27 and round to an ℓ_1 solution $h : V \to \mathbb{R}$ using Proposition 6.3.26, which satisfies

$$\begin{aligned}
&\frac{\max_{\{\lambda_{e,i}\}} \sum_{(e,i)} \lambda_{e,i} |\langle h, y_{e,i}\rangle|}{\sum_{v\in V} \mu(v)|h(v)|} \\
\lesssim\ & (2k)\cdot M(X) + (2k)\cdot \|x - y_x\|_\mu \cdot \sqrt{M(X)} \qquad \text{(by Proposition 6.3.26)} \\
\lesssim\ & \frac{k^{\frac{3}{2}} \cdot M(X)}{\sqrt{(\sigma_\mu^*)_{2k}(F_H)}} \qquad \text{(by Proposition 6.3.27)} \\
\lesssim\ & \frac{k^{\frac{3}{2}} \cdot \log(r\cdot \alpha(G_H)) \cdot \lambda_\mu^*(F_H)}{\sqrt{(\sigma_\mu^*)_{2k}(F_H)}}, \qquad \text{(by Proposition 6.3.19)}
\end{aligned}$$

and the proof of the hard direction is complete using the simple fact that the sum of the smallest $2k$ eigenvalues of the reweighting is at least k times the k-th smallest eigenvalue of the reweighting, i.e.

$$(\sigma_\mu^*)_{2k}(F_H) \geq k \cdot (\lambda_\mu^*)_k(F_H);$$

see Section 4.6 for reference.

Finally, it is straightforward to verify that all of the steps can be implemented in polynomial time. □

6.4 Second Approach: Energy using Flows

In this section, we introduce informally a potential second approach for defining reweighted eigenvalues on submodular transformations. The motivation is to overcome the obstacle of the ℓ_2^2 to ℓ_1 step that bad cancellations might occur when evaluating the inner product $\langle y, x\rangle$ for extreme points y of the base polytopes $\mathcal{B}(F_e)$.

As we have seen in Lemma 6.3.21 that the ℓ_2^2 to ℓ_1 step works when F is simple, i.e. the extreme points y only take the form $y_{u,v} := \mathbb{1}_u - \mathbb{1}_v$. For more complex points $y \in \mathcal{B}(F_e)$, since $y \perp \mathbb{1}_V$, we may try to express y as a convex combination of the simple extreme points $\mathbb{1}_u - \mathbb{1}_v$, so that

$$\langle y, x \rangle = \sum_{u,v \in V} \rho_{u,v} \langle y_{u,v}, x \rangle = \sum_{u,v \in V} \rho_{u,v}(x(u) - x(v)),$$

where $\rho_{u,v}$ are subject to constraints induced by the constraints governing y. The optimization problem from the Lovász extension

$$\max_{y \in B(F_e)} \langle y, x \rangle$$

can then be regarded as a kind of multicommodity flow problem, where the goal is to find a flow ρ, so that the induced demand y respects submodular constraints, and that the total ℓ_1 "energy" is maximized.

The goal of the reweighting is to replace $x(u)-x(v)$ by $|x(u)-x(v)|$, so that the program can be lifted to ℓ_2^2. One issue arises: while the original flow problem is conservative, in the sense that the energy incurred by a flow path only depends on the start and end points, the same can no longer be said when the edge energy becomes $|x(u) - x(v)|$. For example, if $x = (1, 0, 1 - \varepsilon)$ where $\varepsilon \in (0, 1)$, then in the original setting, the unit flow paths $1 \to 3$ and $1 \to 2 \to 3$ have the same energy ε, but in the reweighted setting, $1 \to 3$ has energy ε whereas $1 \to 2 \to 3$ has energy $2 - \varepsilon$. Therefore, additional constraints on $\rho_{u,v}$ are needed to ensure that the energy does not blow up.

We say that a submodular transformation $F = \{F_e\}_{e \in E}$ is *positively edge decomposable* if there exists a polyhedral constraint set

$$\mathcal{F}(e) := \left\{ \rho_{u,v}^{(e)} \in \mathbb{R}_{\geq 0}^{V \times V} : \sum_{u,v} c_{u,v}^{(e)} \rho_{u,v}^{(e)} \leq b^{(e)} \right\}$$

for each $e \in E$ where $c_{u,v}^{(e)}$ are vectors with nonnegative entries, such that

$$\max_{y_e \in \mathcal{B}(F_e)} \sum_{e \in E} \langle y_e, x \rangle = \max_{\substack{\rho_{u,v}^{(e)} \in \mathcal{F}(e) \\ \rho_{u,v} \text{ "Eulerian"}}} \sum_{e \in E} \sum_{u,v \in V} \rho_{u,v}^{(e)} |x(u) - x(v)| \tag{6.7}$$

for all $x \in \mathbb{R}^V$, and where $\rho_{u,v}^{(e)}$ is subject to an additional Eulerian balance constraint that

$$\sum_{e \in E} \left[\sum_{v \in V} \rho_{u,v}^{(e)} - \sum_{v \in V} \rho_{v,u}^{(e)} \right] = 0$$

for all $u \in V$. A submodular transformation $F = \{F_e\}_{e \in E}$ is *approximately positively edge decomposable* if, instead of both sides being equal in (6.7), they are within a constant factor of one another.

This definition is motivated by reweighted eigenvalues for directed hypergraphs. In that case, our constraints here are that $\rho_{u,v}^{(e)} \leq 0$ for $u \notin e^-$ or $v \notin e^+$, and that

$$\sum_{u \in e^-, v \in e^+} \rho_{u,v}^{(e)} \leq 1,$$

and that the ρ-reweighted graph is Eulerian. The corresponding reweighting in Definition 6.3.3 is simply by taking $\lambda_{e,(u,v)} = \rho_{u,v}^{(e)}$ for the extreme point $y_{e,(u,v)} := \mathbb{1}_u - \mathbb{1}_v \in \mathcal{B}(F_e)$.

Example 6.4.1 (Cardinality-Based Submodular Function)**.** *Another example is the following cardinality-based submodular function $f(S) := \min(|S|, |V \setminus S|)$. Assuming that $|V|$ is even, the extreme points are all vectors in $\{\pm 1\}^V$ whose entries sum to zero. Therefore, $\tilde{f}(x)$ is equal to the sum of the largest $|V|/2$ entries in x minus the sum of the smallest $|V|/2$ entries in x.*

One can check that this is equivalent to finding the maximum fractional matching on the graph $(V, V \times V)$ where the edge weight $w(uv) = |x(u) - x(v)|$ and the fractional degree of each $u \in V$ is at most one. This can be translated into the constraints that $\rho_{u,v} \geq 0$ and $\sum_{v \in V} \rho_{u,v} + \sum_{v \in V} \rho_{v,u} \leq 1$ for all $u \in v$, and so f is positively edge decomposable.

Geneally, we define the reweighted ℓ_1 program as follows:

$$\begin{aligned}
\min_{x : V \to \mathbb{R}} \max_{\rho_{u,v}^{(e)} \in \mathcal{F}(e)} \quad & \sum_{e \in E} \sum_{u,v \in V} \rho_{u,v}^{(e)} |x(u) - x(v)| \\
\text{subject to} \quad & \sum_{u \in V} \mu(u) x(u) = 0 \\
& \sum_{u \in V} \mu(u) |x(u)| = 1 \\
& \sum_{e \in E} \left[\sum_{v \in V} \rho_{u,v}^{(e)} - \sum_{v \in V} \rho_{v,u}^{(e)} \right] = 0 \qquad \forall u \in V.
\end{aligned}$$

By (6.7), this is equivalent to the fractional ℓ_1 program $\eta_\mu(F)$ defined in Definition 6.3.1. By Theorem 6.3.9, the latter is in turn equal, up to multiplicative constant, to the expansion $\phi_\mu(F)$.

We can then define a reweighted eigenvalue by lifting this reweighted ℓ_1 program to ℓ_2^2, replacing $|x(u) - x(v)|$ with $\|f(u) - f(v)\|^2$ and the ℓ_1 normalization constraints on x

with ℓ_2^2 normalization constraints on f. The easy direction that the reweighted eigenvalue is upper bounded by $\phi_\mu(F)$ still holds. As for the hard direction, the projection step loss again can be analyzed using large optimal property, the ℓ_2^2 to ℓ_1 step goes through assuming an appropriate choice of vertex measure μ, and the threshold rounding step follows from the ℓ_1 result. We omit the details here.

One important question here is the applicability of this approach. While it works for directed hypergraph expansion and the counting-based cut function, it would be useful to identify a subclass of interesting submodular transformations that are *approximately positively edge decomposable* and for which a spectral theory via this approach is possible.

6.5 Concluding Remarks

In this chapter, we lifted the framework of reweighted eigenvalues to submodular transformations. We proposed a definition of reweighted eigenvalue for submodular transformations and explored the possibility of a Cheeger inequality relating reweighted eigenvalue and expansion. We observe that the definition of reweighted eigenvalue deviates from past approaches: instead of designing an SDP relaxation for the natural nontrivial eigenvalue program $\lambda_\mu(F)$, reweighted eigenvalue is lifted from a reweighted ℓ_1 program, which is formed by "symmetrizing" the Lovász extension and imposing a global balance constraint to control the objective value.

Our investigation identifies the ℓ_2^2 to ℓ_1 step as the bottleneck towards this goal. This step only goes through for simple submodular transformations, which correspond exactly to edge conductance in directed hypergraphs. Therefore, while reweighted eigenvalue yields a comprehensive spectral theory for generalized graphs, it does not appear to be a viable route towards a spectral theory for more complex settings.

In terms of positive results, we proved a Cheeger inequality for edge conductance in directed hypergraphs. With an appropriate definition of k-th reweighted eigenvalue, we proved an improved Cheeger inequality for directed hypergraphs. We remark that these results hold for weighted directed hypergraphs as well, with suitable modifications to the definitions and the proofs.

There are some concrete open problems. The first is to either find an alternative proof for the ℓ_2^2 to ℓ_1 step or exhibit a counterexample. The second is to give a more explicit analysis of the projection loss in Section 6.3.5, especially for directed hypergraphs. The third is to identify an interesting subclass of submodular transformations for which the second approach in Section 6.4 works.

Chapter 7

Reweighted Eigenvalues for Special Graphs

In this chapter, we investigate reweighted eigenvalues for special classes of undirected graphs. First, we prove that classes of graphs such as planar graphs, bounded genus graphs, and H-minor free graphs have small reweighted second eigenvalue. Using the Cheeger inequality established in Chapter 4 relating the second reweighted eigenvalue $\lambda_2^*(G)$ and vertex expansion $\psi(G)$ of a graph G, our result implies that the spectral rounding algorithm using reweighted eigenvalue finds small balanced separators in these special graphs, with a better upper bound on separator size than obtained in previous work [ST96, Kel06, BLR10].

We then prove that the same classes of graphs have small k-th reweighted eigenvalue $\lambda_k^*(G)$, which completes the picture and provides a performance guarantee for the spectral algorithm in Chapter 4 for finding k-way vertex cuts in these special graphs.

7.1 Our Results

Let $G = (V, E)$ be an undirected graph. Before stating the results, let us recall the formal definition of reweighted eigenvalues.

Definition 7.1.1 (Reweighted Eigenvalues (restatement of Definition 4.1.9))**.** *Let $G = (V, E)$ be a graph and let $\pi : V \to \mathbb{R}^+$ be a distribution on V. The maximum reweighted k-th smallest eigenvalue of the normalized Laplacian matrix of G is defined as $\lambda_k^*(G) := \max_{P \geq 0} \lambda_k(I - P)$, where P is subject to the following constraints: $P(u, v) = 0$ if $uv \notin E$, $\sum_{v \in V} P(u, v) = 1$ for all $u \in V$, and $\pi(u)P(u, v) = \pi(v)P(v, u)$ for all $u, v \in V$.*

7.1.1 Spectral Algorithm for Balanced Separators in Planar Graphs

Our first result is an upper bound on the second reweighted eigenvalue $\lambda_2^*(G)$ when G is a planar graph.

Theorem 7.1.2 (Reweighted Eigenvalue Upper Bound for Planar Graphs)**.** *Let $G = (V, E)$ be a planar graph, and π be the uniform distribution on V. Then, $\lambda_2^*(G) \leq O(1/n)$.*

Note that the eigenvalue upper bound has no dependence on the maximum degree Δ. A recursive application of the vertex cut-finding algorithm using reweighted eigenvalues (see Section 4.3 and Section 3.5.2) yields a balanced separator[1] of size $O(\sqrt{(\log \Delta) \cdot n})$. This is already an improvement over the spectral algorithm by Spielman and Teng [ST96] (see Section 3.5.3) which produced balanced separators of size $O(\sqrt{\Delta \cdot n})$.

To match the Lipton-Tarjan optimal separator size of $O(\sqrt{n})$ [LT79], we instead start with an appropriate vector solution to a low-dimensional dual program $\gamma^{(d)}(G)$ (which is defined the same as $\gamma(G)$ in Proposition 3.2.3 but with $f : V \to \mathbb{R}^d$ instead of $\mathbb{R}^n$) and bypass the projection step. The result is summarized below.

Theorem 7.1.3 (Objective Value of Low-Dimensional Dual Program)**.** *Let $G = (V, E)$ be a planar graph, and π be the uniform distribution. Define the d-dimensional dual program as*

$$\begin{aligned}
\gamma^{(d)}(G) := \min_{f: V \to \mathbb{R}^d,\ g: V \to \mathbb{R}_{\geq 0}} \quad & \frac{1}{n} \sum_{v \in V} g(v) \\
\text{subject to} \quad & \frac{1}{n} \sum_{v \in V} \|f(v)\|^2 = 1 \\
& \sum_{v \in V} f(v) = \vec{0} \\
& g(u) + g(v) \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E.
\end{aligned}$$

Then, there is a polynomial-time algorithm that computes a solution (f, g) to $\gamma^{(3)}(G)$ with objective value $O(1/n)$.

As a result, by selecting the best coordinate to be the one-dimensional solution and following the algorithmic proof of Theorem 4.3.7 that $\psi(G)^2 \lesssim \gamma^{(1)}(G)$, we obtain a vertex

[1] Recall from Section 3.5.2 that a balanced separator is defined as a set of vertices, whose removal will break the graph into connected components of size at most $2n/3$.

subset $S \subseteq V$ such that $0 < |S| \leq n/2$ and

$$\psi(S) \lesssim \sqrt{\gamma^{(1)}(G)} \leq \sqrt{3 \cdot \gamma^{(3)}(G)} \lesssim \frac{1}{\sqrt{n}},$$

and recursive application of this algorithm (again, see Section 3.5.2) yields a balanced separator of size $O(\sqrt{n})$, which is optimal up to constants. We give a precise runtime guarantee for the resulting spectral partitioning algorithm below.

Corollary 7.1.4 (Spectral Partitioning using Low-Dimensional Embedding)**.** *Let $G = (V, E)$ be a planar graph and π be the uniform distribution. Then, there is a polynomial-time algorithm that extracts a balanced separator of size $O(\sqrt{n})$ using $\gamma^{(3)}(G)$ defined in Theorem 7.1.3. Furthermore, the runtime of the algorithm is $\tilde{O}(n^2)$.*

We shall prove Theorem 7.1.2, Theorem 7.1.3, and Corollary 7.1.4 in Section 7.2.

7.1.2 Second Eigenvalue Bounds for Other Special Graphs

Our second result is an adaptation of the work by Biswal, Lee, and Rao [BLR10] to upper bound the second smallest reweighted eigenvalue $\lambda_2^*(G)$ of two special classes of graphs: graphs of bounded genus g, and graphs that are K_h-minor free. A simple modification of their proof yields the following analogous result for reweighted eigenvalues under uniform vertex distribution.

Theorem 7.1.5 ($\lambda_2^*(G)$ Upper Bound for Special Graphs)**.** *Let $G = (V, E)$ be a graph, and let $\pi = \mathbb{1}/n$ be the uniform distribution on V. Then the following upper bounds on $\lambda_2^*(G)$ hold:*

- *If G is of genus g, then $\lambda_2^*(G) \leq O(g \log^2 g/n)$.*
- *If G is K_h minor-free, then $\lambda_2^*(G) \leq O(h^6 \log h/n)$.*

As a result of this theorem and the vertex Cheeger inequality (Theorem 4.1.3), the spectral partitioning algorithm using reweighted eigenvalues yields a balanced separator of size $O(\sqrt{g \log^2 g \cdot \log \Delta \cdot n})$ if G is of genus g, and size $O(\sqrt{h^6 \log h \cdot \log \Delta \cdot n})$ if G is K_h-minor free. To the best of our knowledge, Theorem 3.5.5 by [BLR10] was the previous best-known spectral algorithm for balanced separators in these special classes of graphs, and our results improve the dependence of the separator size on the maximum degree Δ from $\sqrt{\Delta}$ to $\sqrt{\log \Delta}$.

Furthermore, a careful adaptation of their proof yields the following generalization for reweighted eigenvalues under arbitrary distribution on V.

Theorem 7.1.6 ($\lambda_2^*(G)$ Upper Bound for Special Graphs, Arbitrary Distribution)**.** *Let $G = (V, E)$ be a graph, and let $\pi : V \to \mathbb{R}^+$ be any distribution on V that satisfies $\pi_{\max} := \max_{v \in V} \pi(v) \leq 1/2$. Then the following upper bounds on $\lambda_2^*(G)$ hold:*

- *If G is of genus g, then $\lambda_2^*(G) \leq O(\pi_{\max} \cdot g \log^2 g)$.*
- *If G is K_h minor-free, then $\lambda_2^*(G) \leq O(\pi_{\max} \cdot h^6 \log h)$.*

We shall prove Theorem 7.1.5 and Theorem 7.1.6 in Section 7.3.

7.1.3 Higher Eigenvalue Bounds for Special Graphs

Our third result is an adaptation of the work by Kelner, Lee, Price, and Teng [KLPT11] to upper bound the k-th smallest reweighted eigenvalue $\lambda_k^*(G)$ of special classes of graphs.

Theorem 7.1.7 ($\lambda_k^*(G)$ Upper Bound for Special Graphs)**.** *Let $G = (V, E)$ be a graph, and let $\pi = \mathbb{1}/n$ be the uniform distribution on V. Let $2 \leq k \leq n$. Then the following upper bounds on $\lambda_k^*(G)$ hold:*

- *If G is planar, then $\lambda_k^*(G) \leq O(\frac{k}{n})$.*
- *If G is of genus g, then $\lambda_k^*(G) \leq O(\frac{k}{n} \cdot g \log^2 g)$.*
- *If G is K_h minor-free, then $\lambda_k^*(G) \leq O(\frac{k}{n} \cdot h^6 \log h)$.*

As a result of Theorem 7.1.7, the spectral partitioning algorithm from the higher-order Cheeger inequality for reweighted eigenvalues (see Theorem 4.1.10) using $\lambda_{2k}^*(G)$ yields k disjoint subsets $S_1, \ldots, S_k$, such that each S_i has vertex expansion at most:

- $O(k^{\frac{3}{2}} \log k \sqrt{(\log \Delta)/n})$ for planar graphs,
- $O(k^{\frac{3}{2}} \log k \sqrt{g \log^2 g \cdot (\log \Delta)/n})$ for genus g graphs, and
- $O(k^{\frac{3}{2}} \log k \sqrt{h^6 \log h \cdot (\log \Delta)/n})$ for K_h-minor free graphs.

Furthermore, a careful adaptation of their proof yields the following generalization for reweighted eigenvalues under arbitrary distribution on V.

Theorem 7.1.8 ($\lambda_k^*(G)$ Upper Bound for Special Graphs, Arbitrary Distribution)**.** *Let $G = (V, E)$ be a graph, and let $2 \le k \le n$. Let $\pi : V \to \mathbb{R}^+$ be any distribution on V that satisfies $\pi_{\max} := \max_{v \in V} \pi(v) \le 1/k$. Then the following upper bounds on $\lambda_k^*(G)$ hold:*

- *If G is planar, then $\lambda_k^*(G) \le O(k\pi_{\max})$.*
- *If G is of genus g, then $\lambda_k^*(G) \le O(k\pi_{\max} \cdot g \log^2 g)$.*
- *If G is K_h minor-free, then $\lambda_k^*(G) \le O(k\pi_{\max} \cdot h^6 \log h)$.*

We shall prove Theorem 7.1.8 and Theorem 7.1.7 in Section 7.4.

7.2 Recovering the Planar Separator Theorem

In this section, we prove Theorem 7.1.3 that, for any planar graph $G = (V, E)$, a solution (f, g) to the three-dimensional program $\gamma^{(3)}(G)$ with objective value $O(1/n)$ exists and can be computed in polynomial time. This immediately implies Theorem 7.1.2. We then analyze the runtime of the algorithm in finding $O(\sqrt{n})$-sized balanced separators to obtain Corollary 7.1.4.

Following Spielman and Teng [ST96], we start with the Koebe-Andreev-Thurston "kissing disks" embedding on the sphere $\mathbb{S}^2$; see Section 3.5.3 for a review. Spielman and Teng considered a family of circle-preserving maps on $\mathbb{S}^2$ and used the Brouwer fixed point theorem to establish the existence of a map that make the average position of the disk centers equal to the origin, thus giving a feasible solution to the minimization problem for $\lambda_2'(G)$. The objective value of the solution is upper bounded using a geometric argument.

To provide an upper bound for $\gamma^{(3)}(G)$, we show how to construct a feasible solution (f, g) to the program from the "kissing disks" embedding. To make the proof algorithmic, we must construct the solution explicitly, instead of applying Brouwer fixed point theorem which merely establishes existence. The key observation is that it suffices, as an intermediary step, to ensure that the centers are not too concentrated in a small area. This is a significantly easier task, and there is a simple procedure that does so whilst preserving the "kissing disks" configuration. This is the content of Section 7.2.1. Then, we upper bound the objective value of the obtained solution using a similar geometric argument to [ST96]. This is the content of Section 7.2.2.

7.2.1 Balancing the Centers of Kissing Disks

Suppose that the following "kissing disks" embedding $v \mapsto D_v$ is computed, where D_v is a spherical cap with center $z(v)$ and geodesic radius $r(v) > 0$. The goal of this subsection is to construct explicitly a circle-preserving map like the $\Phi_{w,\rho}$ map defined in Section 3.5.3, so that if $\bar{z}(v)$ is the center of the mapped disk $\bar{D}_v$, then the average position

$$\bar{z}_{\text{avg}} := \frac{1}{n} \sum_{v \in V} \bar{z}(v)$$

has ℓ_2 distance at most $1 - c$ from the origin for some absolute constant $c > 0$. This is enough for the construction of a feasible solution to $\gamma^{(3)}(G)$ with objective value $O(1/n)$ in the next subsection.

Lemma 7.2.1 (Balancing Kissing Disks)**.** *Let $G = (V, E)$ be a graph and suppose that a "kissing disks" embedding $v \mapsto D_v$ is given; more precisely, $D_v \subseteq \mathbb{S}^2$ are geodesic disks in $\mathbb{S}^2 := \{(x, y, w) \in \mathbb{R}^3 : x^2 + y^2 + w^2 = 1\}$ given in center-radius form, with centers $z(v)$ and radii $r(v)$, such that their interiors are pairwise disjoint, and D_u and D_v touch if and only if $uv \in E$. Then, for some universal constant $0 < c < 1$, we can compute in linear time another "kissing disks" embedding $v \mapsto \bar{D}_v$, such that if $\bar{z}(v) \in \mathbb{S}^2$ is the center of the mapped disk $\bar{D}_v$, then*

$$\left\| \frac{1}{n} \sum_{v \in V} \bar{z}(v) \right\| \leq 1 - c. \tag{7.1}$$

Our strategy is to apply a series of disk-preserving mappings to the original embedding, informally described below.

1. Take $w \in \mathbb{S}^2$ to be any point so that w is not in the closure of any disk, and stereographically project $\mathbb{S}^2$ to the plane tangent to $\mathbb{S}^2$ at $-w$.

2. Translate the disks on the plane so that $(0, 0)$ is the "median" of their centers.

3. Scale the disks to control the distances from the disks to $(0, 0)$.

4. Apply inverse stereographic projection back to $\mathbb{S}^2$.

Intuitively, the second and third steps ensure that, after the inverse stereographic projection in the last step, the centers of the disks will be reasonably "well-spread" in both longitudinal and latitudinal directions.

Proof. First, since stereographic projection and its inverse, as well as translation and dilation on the plane are all disk-preserving transformations that preserve the "kissing disks" property, the outcome of the transformation is a "kissing disks" embedding on $\mathbb{S}^2$.

Then, we shall prove (7.1) for the centers $\bar{z}(v)$ of the mapped disks. It is easier to prove the equivalent "balanced" condition holds for $\bar{z}(v)$ that

$$\sum_{u,v\in V} \|\bar{z}(u) - \bar{z}(v)\|^2 \geq \Omega(n^2). \tag{7.2}$$

Note that $\|\bar{z}(u)\| = 1$ since $\bar{z}(u)$ is on the unit sphere, and since

$$\sum_{u,v\in V} \|\bar{z}(u) - \bar{z}(v)\|^2 = 2n \sum_{v\in V} \|\bar{z}(u)\|^2 - 2 \sum_{u,v\in V} \langle \bar{z}(u), \bar{z}(v)\rangle = 2n^2 \left(1 - \left\|\frac{1}{n}\sum_{v\in V} \bar{z}(v)\right\|^2\right),$$

the above condition implies the desired condition that

$$\left\|\frac{1}{n}\sum_{v\in V} \bar{z}(v)\right\| \leq 1 - c$$

for some universal constant $c \in (0, 1)$.

We now describe our transformation in full detail.

1. First, take $w \in \mathbb{S}^2$ to be any point so that w is not in the closure of any disk, and stereographically project[2] $\mathbb{S}^2$ to the plane tangent to $\mathbb{S}^2$ at $-w$. To find such w, one way is to choose an arbitrary disk and choose a point close to its boundary that is not in the closure of any other disks.

   After the stereographic projection step, identify the plane with the standard Euclidean plane $\mathbb{R}^2$ and assign (x, y) coordinates to points, and call the normal direction the w-axis. Refer to Figure 7.1 for a visualization of the coordinate system. Write D_v^Π as the image of the disk D_v under this projection, and $z^\Pi(v) = (x^\Pi(v), y^\Pi(v))$ as the center of D_v^Π. Note that in general, this is not the image of the original disk center $z(v)$ under the projection.

2. The translation step follows. We translate the disks so that the median of $x^\Pi(v)$ is zero and the median of $y^\Pi(v)$ is zero. While this does not give precise control of the

[2]See Figure 3.3 for an illustration.

location of the disk centers after lifting back to the sphere, it ensures that the "left" and right" half ($Q_3 \cup Q_4$ and $Q_1 \cup Q_2$ respectively, in Figure 7.1) each contains at most $n/2$ centers, and the "top" and "bottom" half ($Q_2 \cup Q_3$ and $Q_1 \cup Q_4$ respectively, in Figure 7.1) each contains at most $n/2$ centers.

3. Then comes the scaling step. We dilate the disks about the plane origin by the smallest possible factor so that, after the inverse stereographic projection in the final step, at least $1/10$ of the disk centers $\bar{z}(v)$[3] lie in the southern hemisphere, i.e. the lower half of $\mathbb{S}^2$ in the w-direction. This is possible since, for every disk D_v^{Π} except at most one disk containing the plane origin, the distance from the center $\bar{z}(v)$ of the lifted disk to the south pole increases monotonically from 0 to π (distance between south pole and north pole) as the scaling factor increases from 0 to ∞. We may find the desired scaling factor by computing, for each disk, the threshold scaling factor that makes the distance from $\bar{z}(v)$ to the south pole equal to $\pi/2$, and sorting the scaling factors.

4. At last, we apply inverse stereographic projection to obtain disks $\bar{D}_v$ on the sphere with centers $\bar{z}(v)$.

We verify (7.2) for the mapped centers $\bar{z}(v)$. There are two cases to consider.

In the first case, at least $1/10$ of the centers $\bar{z}(v)$ have distance at least $5\pi/6$ from the south pole. Let's call this set of points $\mathcal{N}$, and the set of points in the southern hemisphere $\mathcal{S}$. The geodesic distance between any point in $\mathcal{N}$ and any point in $\mathcal{S}$ is $\Omega(1)$, and so the Euclidean distance is at least a universal constant $\delta_1 > 0$, from which it follows that

$$\sum_{u,v \in V} \|\bar{z}(u) - \bar{z}(v)\|^2 \geq \sum_{u \in \mathcal{N}} \sum_{v \in \mathcal{S}} \|\bar{z}(u) - \bar{z}(v)\|^2 \geq \delta_1 \cdot |\mathcal{N}| \cdot |\mathcal{S}| \geq \frac{\delta_1 n^2}{100},$$

so (7.2) is true in this case.

In the second case, at most $1/10$ of the centers $\bar{z}(v)$ have distance at least $5\pi/6$ from the south pole. Then, there are at least $n(1 - 1/10 - 1/10) = 4n/5$ centers in the strip region denoted $\mathcal{R}$, which is the set of points in the sphere whose geodesic distance to the south pole is between $\pi/2$ and $5\pi/6$. Let $\mathcal{R}_i := \hat{Q}_i \cap \mathcal{R}$ where $\hat{Q}_i := Q_i \times \mathbb{R}$ and Q_i are the quadrants in Figure 7.1. Take the region $\mathcal{R}_{i^*}$ containing the fewest centers, which will contain *at most* $n/4$ centers. For now, assume that the following claim is true:

[3]Again, we caution that this is in general not the image of $z^{\Pi}(v)$ (after translation and scaling) under the inverse stereographic projection.

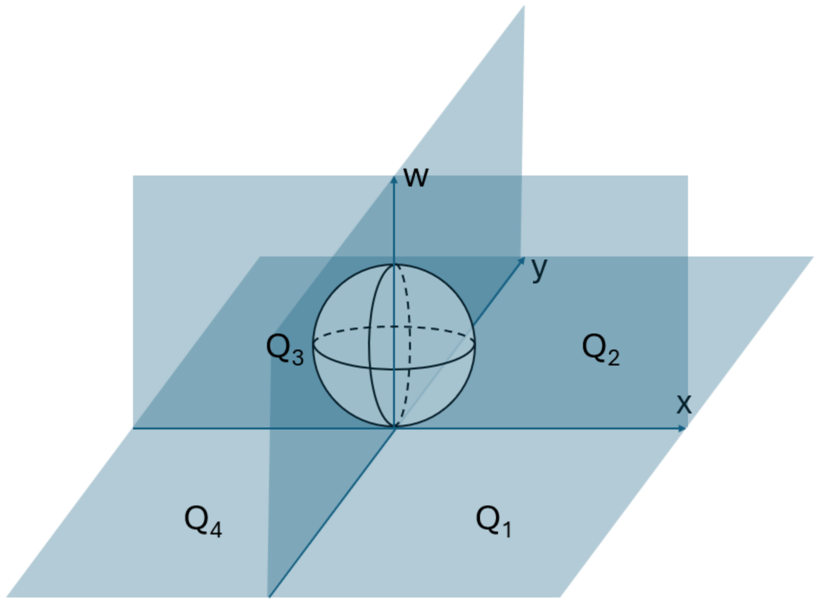


Figure 7.1: The quadrants Q_i of the plane. Boundaries are not included. For example, Q_1 is the set of points on the plane whose x-coordinates are positive and y-coordinates are negative. We use $\hat{Q}_i$ to denote $Q_i \times \mathbb{R}$, which is the set of points in the xyw-space whose projection to the plane lies in Q_i. In this xyw-coordinate system, the south pole and the north pole of $\mathbb{S}^2$ are the points $(0, 0, 0)$ and $(0, 0, 2)$ respectively.

Claim 7.2.2. *Any two adjacent regions $\mathcal{R}_i$ and $\mathcal{R}_j$ contain at least $3n/10$ centers in total in the closure.*

Then, the two regions $\mathcal{R}_j$ and $\mathcal{R}_k$ adjacent to $\mathcal{R}_{i^*}$ each contain at least $3n/10 - n/4 = n/20$ centers. As $\mathcal{R}_j$ and $\mathcal{R}_k$ are not adjacent, the geodesic distance between any point in $\mathcal{R}_j$ and any point in $\mathcal{R}_k$ is $\Omega(1)$, and so the Euclidean distance is at least a universal constant $\delta_2 > 0$. By the same argument as the first case, we have

$$\sum_{u,v \in V} \|\bar{z}(u) - \bar{z}(v)\|^2 \geq \frac{\delta_2 n^2}{400},$$

again verifying (7.2).

So, conditioned on Claim 7.2.2, (7.2) holds for the centers $\bar{z}(v)$ of the mapped disks. Proving Claim 7.2.2 amounts to verifying that, if the center of the disk is in the quadrant Q_i after the translation step and that the disk does not contain $(0, 0)$, then the center

$\bar{z}(v)$ of the final mapped disk $\bar{D}_v$ is in $\hat{Q}_i \cap \mathbb{S}^2$. This is because both dilation and inverse stereographic projection preserve the longitude of the disk center. Thus, for any two adjacent regions $\mathcal{R}_i$ and $\mathcal{R}_j$, the other two regions contain at most $n/2$ points in total by the translation step, and so $\mathcal{R}_i$ and $\mathcal{R}_j$ must together contain at least $4n/5 - n/2 = 3n/10$ points in total in the closure.

Finally, to see that this transformation can be computed in linear time, note that:

- Both stereographic projection and inverse stereographic projection can be computed in linear time, or constant time per disk. By choosing three points on the disk boundary and computing their images. The center of the mapped disk can be computed from the images of these three points.
- The translation step can be computed in linear time, as it amounts to finding the median of n numbers.
- The scaling step can also be computed in linear time, or constant time per disk. For each disk that does not contain $(0, 0)$, take the two intersection points of its boundary with the ray from $(0, 0)$ through its center. By a rotation about the w-axis, suppose that the ray is the positive x-axis and the two intersection points are $(x_1, 0)$ and $(x_2, 0)$, where $0 < x_1 < x_2$. Then, the required threshold scaling factor for the disk is $2/\sqrt{x_1 x_2}$, per the elementary geometric argument in Figure 7.2.

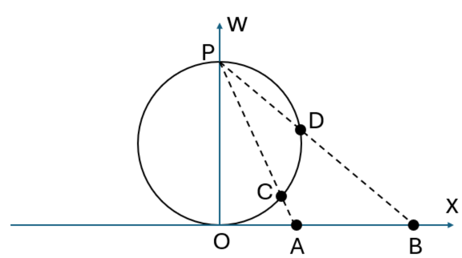


Figure 7.2: The $y = 0$ section of Figure 7.1. In xw-coordinates, $O = (0, 0)$, $P = (0, 2)$, $A = (\beta x_1, 0)$ and $B = (\beta x_2, 0)$ where $\beta > 0$ is the scaling factor. The inverse stereographic projection maps A to C and B to D. Let $\theta_1 := \angle CPO$ and $\theta_2 := \angle DPO$. The threshold scaling factor β should make $\theta_1 + \theta_2 = \pi/2$, which is equivalent to $\tan(\theta_1) = \cot(\theta_2)$, or $OP^2 = AO \cdot BO$. But this is just $2^2 = (\beta x_1) \cdot (\beta x_2)$, which rearranges to $\beta = 2/\sqrt{x_1 x_2}$.

□

7.2.2 Dual Solution from Balanced Kissing Disks

In this subsection, we explain how to extract from the "kissing disks" embedding of Lemma 7.2.1 a feasible solution to the dual program $\gamma^{(3)}(G)$ with small objective value. We prove Theorem 7.1.3 and, as a corollary, Theorem 7.1.2.

Proof of Theorem 7.1.3. Given a planar graph $G = (V, E)$ we start with a "kissing disk" embedding $v \mapsto D_v$, which can be computed in polynomial time due to Mohar [Moh93]. From this, we apply Lemma 7.2.1 to compute a good "kissing disk" embedding $v \mapsto \bar{D}_v$, whose average position of the centers $\bar{z}(v)$ satisfy

$$\|\bar{z}_{\text{avg}}\| = \left\| \frac{1}{n} \sum_{v \in V} \bar{z}(v) \right\| \leq 1 - c$$

for some universal constant $c \in (0, 1)$. Define the following solution (f, g) to the $\gamma^{(3)}(G)$ program:

$$f(v) := \beta \cdot (\bar{z}(v) - \bar{z}_{\text{avg}}) \qquad \text{and} \qquad g(v) := 2\beta^2 \bar{r}(v)^2$$

for $v \in V$, where $\bar{r}(v) > 0$ is the radius of the disk $\bar{D}_v$ and $\beta > 0$ is the unique scaling factor so that

$$\frac{1}{n} \sum_{v \in V} \|f(v)\|^2 = 1.$$

Computing (f, g) can be done in polynomial time. Immediately by the definition of $\bar{z}_{\text{avg}}$, we have $\sum_{v \in V} f(v) = \vec{0}$. It remains to check that

$$g(u) + g(v) \geq \|f(u) - f(v)\|^2 \qquad \forall uv \in E$$

so that (f, g) is feasible, and to upper bound the objective.

For the first part, note that for $uv \in E$, the two disks $\bar{D}_u$ and $\bar{D}_v$ are tangent to one another. So, the *geodesic* distance between $\bar{z}(u)$ and $\bar{z}(v)$ is exactly $\bar{r}(u) + \bar{r}(v)$. Since Euclidean distance is upper bounded by geodesic distance on the sphere, we have

$$\|\bar{z}(u) - \bar{z}(v)\|^2 \leq (\bar{r}(u) + \bar{r}(v))^2 \leq 2(\bar{r}(u)^2 + \bar{r}(v)^2),$$

and using this we can infer that $g(u) + g(v) \geq \|f(u) - f(v)\|^2$, and so (f, g) is feasible.

To upper bound the objective value of this solution, note that

$$\frac{1}{n} \sum_{v \in V} g(v) = \frac{1}{n} \sum_{v \in V} 2\beta^2 \bar{r}(v)^2 \lesssim \frac{\beta^2}{n},$$

which is $O(1/n)$ if we can show that $\beta = O(1)$. In the last inequality, we used the same observation as in Section 3.5.3 that $\sum_{v \in V} \bar{r}(v)^2$ is upper bounded by a constant times the area of $\mathbb{S}^2$, which is $O(1)$. Indeed, for each $v \in V$, since $\|\bar{z}(v)\| = 1$ and $\|\bar{z}_{\text{avg}}\| \leq 1 - c$ for some positive universal constant c, we have

$$\frac{1}{n} \sum_{v \in V} \|\bar{z}(v) - \bar{z}_{\text{avg}}\|^2 \geq \frac{1}{n} \sum_{v \in V} (1 - (1 - c))^2 \geq c^2.$$

Therefore, the scaling factor β satisfies $\beta \leq 1/c \leq O(1)$. We have shown that the solution is feasible and has objective value at most $O(1/n)$. □

Proof of Theorem 7.1.2. Note that $\gamma^{(d)}(G)$ for any $d \leq n$ is a restriction of the $\gamma(G)$ program which is strongly dual to $\lambda_2^*(G)$. Therefore, we have

$$\lambda_2^*(G) = \gamma(G) \leq \gamma^{(3)}(G) \lesssim \frac{1}{n}$$

where the last inequality is by Theorem 7.1.3. □

7.2.3 Efficient "Optimal" Spectral Partitioning for Planar Graphs

We design a spectral algorithm, Algorithm 3 for finding balanced separators in planar graphs based on Theorem 7.1.3 and the Cheeger rounding algorithm in Chapter 4 for vertex expansion. We demonstrate Corollary 7.1.4 that the runtime of Algorithm 3 is $\tilde{O}(n^2)$. The runtime bottleneck is the computation of an initial "kissing circles" embedding in the first step, and the remaining steps of the algorithm runs in time $\tilde{O}(n^{1.5})$.

Proof of Corollary 7.1.4. First, we prove the correctness of Algorithm 3 by demonstrating that it outputs a balanced separator of size $O(\sqrt{n})$. As shown in the proof of Theorem 7.1.3, the objective value of the solution (f, g) computed in line 5 is $O(1/|V'|) = O(1/n)$. The objective value of the one-dimensional coordinate solution is at most three times that which is also $O(1/n)$. So, using the Cheeger rounding algorithm in Theorem 4.3.7, line 6 produces a cut A whose vertex expansion is $O(1/\sqrt{n})$.

Let U be the union of the cuts A found in the while loop. Then, $\partial U = S$, so removing S disconnects U from V'. Moreover, since each iteration we remove at most $(1/2 + o(1))|V'|$ vertices from V', at the end of the algorithm V' has size at least $n/3 - o(n)$ and at most $2n/3$, and we can conclude the same about U. So, S is a balanced separator, and its size is at most $O(1/\sqrt{n}) \cdot |U|$ which is $O(\sqrt{n})$.

Algorithm 3 Spectral Partitioning Algorithm for Planar Graphs

Input: Planar graph $G = (V, E)$

Output: A vertex separator $S \subseteq V$

1: Compute a "kissing circles" embedding $v \mapsto D_v$ of the graph G per [DLQ20]
2: Let $V' := V$ be the set of remaining vertices, and $S := \emptyset$
3: **while** $|V'| \geq 2n/3$ **do**
4: Balance the embedding $v \mapsto D_v$ for $G[V']$ per Lemma 7.2.1 to obtain $v \mapsto \bar{D}_v$
5: Compute feasible solution (f, g) to $\gamma^{(3)}(G[V'])$ as described in Section 7.2.2
6: Choose the best coordinate and apply Theorem 4.3.7 to round the one-dimensional solution, obtaining a cut $A \subseteq V$ with $|A| \leq |V'|/2$ and $\psi(A)$ small
7: $S \leftarrow S \cup \partial A$, $V' \leftarrow V' \setminus (A \cup \partial A)$
8: **end while**
9: **return** S

Next, we prove that the runtime of the algorithm is $\tilde{O}(n^2)$. Line 1 takes time $\tilde{O}(n^2)$ by [DLQ20]. The while loop only repeats $O(\sqrt{n})$ times, since $\psi(A) \leq O(1/\sqrt{n})$ implies that at least $|A| \geq \Omega(\sqrt{n})$ vertices are removed from V' every iteration. Inside the while loop, line 4 takes $O(|V'|) = O(n)$ time by Lemma 7.2.1, and lines 5-7 can be implemented to run in $O((|V'| \log |V'| + |E(G[V'])|)) = \tilde{O}(n)$ time. Therefore, the overall runtime is dominated by the first step and is $\tilde{O}(n^2)$. □

7.3 Bounding Reweighted Second Eigenvalue

In this section, we prove Theorem 7.1.5 and Theorem 7.1.6 which upper bounds $\lambda_2^*(G)$ for graphs G that are either of bounded genus g, or K_h-minor free. First, we show that a small tweak of the main proof in Biswal, Lee, and Rao [BLR10] suffices to prove the reweighted eigenvalue bound for the uniform distribution case. Then, we detail the adaptations needed to generalize the result to arbitrary distributions. The reader is referred to Section 3.5.5 for a review of [BLR10].

7.3.1 The Uniform Distribution Case

In order to prove the reweighted eigenvalue upper bounds in Theorem 7.1.5 for the uniform distribution, it suffices to derive a result analogous to Lemma 3.5.8 that upper bounds $\lambda_2^*(G)$ using distortion and metric parameters. This is the content of the following lemma.

Lemma 7.3.1. *Let $G = (V, E)$ be a graph and $\pi = \mathbb{1}/n$. Then,*

$$\lambda_2^*(G) \lesssim n^3 \cdot \alpha(G)^2 \cdot \left[\max_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{u,v\in V} d_s(u,v)}{\sqrt{\sum_{v\in V} s(v)^2}} \right]^{-2},$$

where $\alpha(G)$ is the average distortion parameter in Lemma 3.5.7 and $d_s(u,v)$ is the shortest path metric defined in (3.15).

Proof. By Definition 7.1.1, to upper bound $\lambda_2^*(G)$, it suffices to prove an upper bound on

$$\lambda_2(I - P) = \min_{\substack{f:V\to\mathbb{R} \\ f\perp\mathbb{1}}} \frac{\sum_{uv\in E} P(u,v)(f(u) - f(v))^2}{\sum_{v\in V} f(v)^2}$$

for every reweighting P that satisfies the constraints of $\lambda_2^*(G)$ under uniform distribution: namely, that $P \geq 0$, $P(u,v) = 0$ for $uv \notin E$, $\sum_{v\in V} P(u,v) = 1$ for all $u \in V$, and $P(u,v) = P(v,u)$ for all $u, v \in V$.

Using (3.14) and Lemma 3.5.7, we can rewrite the denominator of the minimization problem, and then relate ℓ_2^2 quasimetrics induced by f to shortest path metrics d_s induced by $s : V \to \mathbb{R}_{\geq 0}$:

$$\begin{aligned}
& \min_{\substack{f:V\to\mathbb{R} \\ f\perp\mathbb{1}}} \frac{\sum_{uv\in E} P(u,v)(f(u) - f(v))^2}{\sum_{v\in V} f(v)^2} \\
= \ & 2n \min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E} P(u,v)(f(u) - f(v))^2}{\sum_{u,v\in V} (f(u) - f(v))^2} \\
\lesssim \ & 2n \cdot \alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{uv\in E} P(u,v) d_s(u,v)^2}{\sum_{u,v\in V} d_s(u,v)^2} \\
\leq \ & 4n \cdot \alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{uv\in E} P(u,v)(s(u)^2 + s(v)^2)}{\sum_{u,v\in V} d_s(u,v)^2} \\
= \ & 4n \cdot \alpha(G)^2 \cdot \min_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{v\in V} s(v)^2}{\sum_{u,v\in V} d_s(u,v)^2},
\end{aligned}$$

where the last equality uses the constraints that $P(u,v) = P(v,u)$ and $\sum_{v\in V} P(u,v) = 1$ for all $u \in V$. From here, proceeding as in the proof of Lemma 3.5.8, we arrive at the desired result. □

As in Section 3.5.5, combining Lemma 7.3.1 with Lemma 3.5.9, Lemma 3.5.7 and Lemma 3.5.10 yields Theorem 7.1.5. Compared to Lemma 3.5.8, there is no dependence on Δ in the upper bound, which explains the factor Δ improvement in Theorem 7.1.5.

7.3.2 The General Case

To generalize Theorem 7.1.5 to arbitrary distributions on V, our strategy is again to introduce π-weighted variants for certain quantities considered in [BLR10]. Below, we use the proof of Theorem 3.5.5 as a blueprint, and explain the modifications needed for proving Theorem 7.1.6.

Rayleigh Quotient to Shortest Path Metric

We prove the following generalization of Lemma 7.3.1.

Lemma 7.3.2. *Let $G = (V, E)$ be a graph and π be any distribution on V. Then,*

$$\lambda_2^*(G) \lesssim \alpha(G)^2 \cdot \left[\max_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{u,v\in V} \pi(u)\pi(v) d_s(u,v)}{\sqrt{\sum_{v\in V} \pi(v) s(v)^2}} \right]^{-2},$$

where $\alpha(G)$ is the average distortion parameter in Lemma 3.5.7 and $d_s(u, v)$ is the shortest path metric defined in (3.15).

Note that the disappearance of n^3 is of no concern: by introducing the π-weights, in the uniform distribution case where $\pi = 1/n$, the maximization program incurs an additional factor of $(\sqrt{n}/n^2)^{-2} = n^3$ compared to Lemma 7.3.1.

Proof. Again, it suffices to prove an upper bound on

$$\lambda_2(I - P) = \min_{\substack{f:V\to\mathbb{R} \\ f\perp\pi}} \frac{\sum_{uv\in E} \pi(u)P(u,v)(f(u) - f(v))^2}{\sum_{v\in V} \pi(v)f(v)^2} \tag{7.3}$$

for every reweighting P that satisfies that constraints of $\lambda_2^*(G)$: namely, that $P \geq 0$, $P(u, v) = 0$ for $uv \notin E$, $\sum_{v\in V} P(u, v) = 1$ for all $u \in V$, and $\pi(u)P(u, v) = \pi(v)P(v, u)$ for all $u, v \in V$. We follow the proof of Lemma 3.5.8, introducing the necessary modifications along the way.

To handle the denominator as before, we use Fact 2.10.4 to infer (note that $\pi(V) = 1$)

$$\sum_{v\in V} \pi(v)f(v)^2 = \frac{1}{2} \sum_{u,v\in V} \pi(u)\pi(v)(f(u) - f(v))^2 \tag{7.4}$$

for any $f \perp \pi$. After substituting (7.4) in the denominator of (7.3), the objective of the minimization becomes invariant to translation (by $c \cdot \mathbb{1}$). As any $f : V \to \mathbb{R}$ can be translated so that $f \perp \pi$, the constraint $f \perp \pi$ can be dropped without loss. We are thus left with the task of upper bounding the objective of the minimization problem

$$\min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E} \pi(u)P(u,v)(f(u)-f(v))^2}{\sum_{u,v\in V} \pi(u)\pi(v)(f(u)-f(v))^2}.$$

Next, to go from ℓ_2^2 quasimetric to the shortest path metric, a straightfoward π-weighted generalization of Lemma 3.5.7 is needed. We include it in Appendix C for completeness.

Lemma 7.3.3 (Average Distortion [BLR10])**.** *For any graph $G = (V, E)$ and any distribution π on V, there exists $\alpha(G) > 0$ such that the following holds: for any shortest path metric d_s, there is a function $f : V \to \mathbb{R}$ such that if we write $d_f(u,v) := |f(u) - f(v)|$ then $d_f(u,v) \le d_s(u,v)$ for all $u, v \in V$ and*

$$\sum_{u,v\in V} \pi(u)\pi(v)d_s(u,v)^2 \lesssim \alpha(G)^2 \cdot \sum_{u,v\in V} \pi(u)\pi(v)d_f(u,v)^2.$$

Furthermore, the same upper bounds on $\alpha(G)$ in Lemma 3.5.7 hold here as well.

With this, we can lower bound the denominator and upper bound the numerator of the minimization objective, giving

$$\min_{f:V\to\mathbb{R}} \frac{\sum_{uv\in E} \pi(u)P(u,v)(f(u)-f(v))^2}{\sum_{u,v\in V} \pi(u)\pi(v)(f(u)-f(v))^2} \le \alpha(G)^2 \min_{s:V\to\mathbb{R}_{\ge 0}} \frac{\sum_{uv\in E} \pi(u)P(u,v)d_s(u,v)^2}{\sum_{u,v\in V} \pi(u)\pi(v)d_s(u,v)^2}.$$

Now, for any $s : V \to \mathbb{R}^+$ with induced shortest path metric d_s, we have

$$\begin{aligned} \frac{\sum_{uv\in E} \pi(u)P(u,v)d_s(u,v)^2}{\sum_{u,v\in V} \pi(u)\pi(v)d_s(u,v)^2} &\le \frac{2\sum_{uv\in E} \pi(u)P(u,v)(s(u)^2+s(v)^2)}{\sum_{u,v\in V} \pi(u)\pi(v)d_s(u,v)^2} \\ &= \frac{2\sum_{v\in V} \pi(v)s(v)^2}{\sum_{u,v\in V} \pi(u)\pi(v)d_s(u,v)^2}, \end{aligned}$$

where the inequality is from $d_s(u,v)^2 = (s(u)+s(v))^2 \le 2(s(u)^2+s(v)^2)$ and the equality follows from the reweighted eigenvalue constraints that $\pi(u)P(u,v) = \pi(v)P(v,u)$ and $\sum_{v\in V} P(u,v) = 1$ for all $u \in V$.

Finally, we use Cauchy-Schwarz inequality to get to the desired form in the statement:

$$\begin{aligned}
\min_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{v\in V}\pi(v)s(v)^2}{\sum_{u,v\in V}\pi(u)\pi(v)d_s(u,v)^2} &\leq \min_{s:V\to\mathbb{R}_{\geq 0}} \frac{\left(\sum_{u,v\in V}\pi(u)\pi(v)\right)\cdot\sum_{v\in V}\pi(v)s(v)^2}{(\sum_{u,v\in V}\pi(u)\pi(v)d_s(u,v))^2} \\
&= \left[\max_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{u,v\in V}\pi(u)\pi(v)d_s(u,v)}{\sqrt{\sum_{v\in V}\pi(v)s(v)^2}}\right]^{-2},
\end{aligned}$$

and the proof is complete. □

Metric Spread and Flow Congestion

The next step is to relate the objective of the maximization program

$$\max_{s:V\to\mathbb{R}_{\geq 0}} \frac{\sum_{u,v\in V}\pi(u)\pi(v)d_s(u,v)}{\sqrt{\sum_{v\in V}\pi(v)s(v)^2}} =: \max_{s:V\to\mathbb{R}_{\geq 0}} \Lambda_s(G)$$

to a multicommodity flow problem, where the goal is to minimize flow congestion. We incorporate π-weights in the definition of flow congestion, defining

$$\text{con}_\pi(F) := \left(\sum_{v\in V}\frac{c_F(v)^2}{\pi(v)}\right)^{1/2}.$$

The intuition is that the congestion at a vertex should be measured relative to the amount of flow it is required to send, which in this case it is $\pi(v)$. Practically the same proof as for Lemma 3.5.9 yields the following duality result, which is deferred to Appendix C.

Lemma 7.3.4 (Flow/Metric Duality, π-Weighted Version)**.** *For any graph $G = (V, E)$ and distribution $\pi : V \to \mathbb{R}^+$,*

$$\min_{F\in\mathcal{F}_\pi(G)} \text{con}_\pi(F) = \max_{s:V\to\mathbb{R}_{\geq 0}} \Lambda_s(G),$$

where the minimum is taken over $\mathcal{F}_\pi(G)$, the set of all solutions for the multicommodity flow problem with product demand $D(u,v) = \pi(u)\pi(v)$ for all $u \neq v \in V$.

Flow Congestion Lower Bound

In this step, we derive lower bounds on flow congestion in special graphs. The current setting differs from the unweighted setting in two ways. First, the demand graph is no longer uniform demand, but product demand with $D(u,v) = \pi(u)\pi(v)$ for $u \neq v$. Second, the definition of congestion has changed.

We first use a trivial relation to bridge the second difference. Recall that the unweighted congestion is defined as

$$\operatorname{con}(F) := \left(\sum_{v \in V} c_F(v)^2 \right)^{1/2}.$$

Then, letting $\pi_{\max} := \max_{u \in V} \pi(u)$ we have $\operatorname{con}_\pi(F) \geq \pi_{\max}^{-1/2} \operatorname{con}(F)$.

Now we can focus on lower bounding $\operatorname{con}(F)$. The strategy is to construct an auxiliary graph G', so that G' is itself a special graph — either bounded genus or K_h-minor free — and product-demand flow congestion on G correspond to uniform-demand flow congestion on G', so that we can lower bound the former using known lower bounds on the latter from Lemma 3.5.10.

On a high level, we want to represent vertex v with a cluster of $n_v \propto \pi(v)$ vertices, so that the uniform-demand multicommodity flow has $n_u n_v \propto \pi(u)\pi(v)$ demand between cluster u and cluster v.

Now we describe our construction. In the first step, given a distribution $\pi : V \to \mathbb{R}^+$, turn it into an integral weight function by finding positive integers $M, \{n_v\}_{v \in V}$ such that

$$\frac{n_v}{M} \leq \pi(v) \leq 2\frac{n_v}{M}$$

for all $v \in V$. That it is possible is proven below.

Proposition 7.3.5. *Given n positive real numbers $a_1, \ldots, a_n$ such that $\max_i a_i \leq (\sum_i a_i)/2$, there exists positive integers $M, b_1, \ldots, b_n$ such that $b_i/M \leq a_i \leq 2b_i/M$ for all $i \in [n]$, and moreover that $\max_i b_i \leq 3(\sum_i b_i)/4$.*

Proof. First choose $M \in \mathbb{N}$ to be such that $Ma_i \geq 2$ for all $i \in [n]$. Then, take $b_i := \lfloor Ma_i \rfloor$. This satisfies the first condition since $b_i \leq Ma_i$ and $2b_i \geq b_i + 1 \geq Ma_i$. This also satisfies the second condition because

$$b_j \leq Ma_j \leq \frac{1}{2} \sum_i (Ma_i) \leq \frac{3}{4} \sum_i \lfloor Ma_i \rfloor = \frac{3}{4} \sum_i b_i$$

for any b_j. ☐

For any solution F to the multicommodity flow problem with product demand $D(u,v) = \pi(u)\pi(v)$, there is a solution to the multicommodity flow problem with product demand $D^*(u,v) = n_u n_v$ that is upper bounded by $4M^2 \cdot F$, and so a lower bound on the congestion of the latter implies a lower on the congestion of the former.

The reason we pass to integral product demand is so that discrete constructions are possible. Given $G = (V,E)$ and n_u from above, construct the following auxiliary graph:

Definition 7.3.6 (Constellation Graph)**.** *Let $G = (V,E)$ be a graph and let $\{n_u\}_{u \in V}$ be given positive integers. The constellation graph*[4] *of G with respect to n_u is the graph $G' = (V', E')$ with*

- *For each $u \in V$, there are n_u copies of u in V': $u_0, u_1, \ldots, u_{n_u - 1}$. u_0 can be considered the original/central vertex;*
- *For each $(u,v) \in E$, there is an edge $(u_0, v_0) \in E'$;*
- *For each $u \in V$ and $1 \leq i < n_u$, there is an edge $(u_0, u_i) \in E'$.*

Refer to Figure 7.3 for an example.

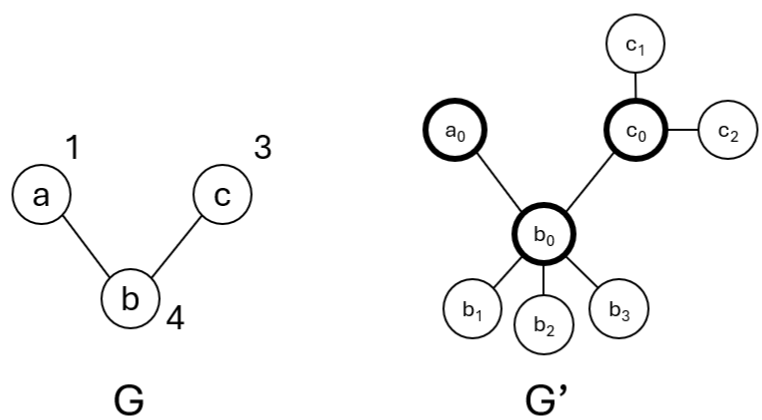


Figure 7.3: On the left is a graph $G = (V,E)$ equipped with an integral weight function n_v on V, where $n_a = 1, n_b = 4, n_c = 3$. On the right is the constellation graph G'.

The goal now is to relate (i) the multicommodity flow problem on G' with uniform demand and (ii) the multicommodity flow problem on G with product demand $D^*(u,v) = n_u n_v$, and then to show that special properties of G are inherited by G'. This is the content of the following two lemmata.

[4]The graph is an interconnected collection of star graphs, hence the name.

Lemma 7.3.7 (Constellation Graph Congestion)**.** *Let $G = (V, E)$ be a graph with distribution π on V, $\{n_u\}_{u \in V}$ be positive integers satisfying $\max_u n_u \le 2(\sum_u n_u)/3$, and G' be the constellation graph constructed from G and $\{n_u\}_{u \in V}$. Then, there exists universal constants $A, B > 0$ such that*

$$A \cdot \min_{F'} \operatorname{con}(F') \le \min_{F^*} \operatorname{con}(F^*) \le B \cdot \min_{F'} \operatorname{con}(F'),$$

where the minimum of F^* is taken over all solutions F^* of the multicommodity flow problem on G with product demand $D^*(u, v) = n_u n_v$, and the minimum of F' is taken over all flow solutions F' of the multicommodity flow problem on G' with uniform demand.

Lemma 7.3.8 (Constellation Graph Topology)**.** *Under the same setting as Lemma 7.3.7, if G has genus g then G' has genus g, and if G is K_h-minor free where $h \ge 3$ then G' is K_h-minor free.*

We first prove the two lemmata, then finish the proof of Theorem 7.1.6.

Proof of Lemma 7.3.7. First, we prove that there exists $B > 0$ such that

$$\min_{F^*} \operatorname{con}(F^*) \le B \min_{F'} \operatorname{con}(F').$$

We take $B = 2$. For any solution F' to the multicommodity flow problem on G' with uniform demand, construct a solution F^* to the multicommodity flow problem on G with product demand $D^*(u, v) = n_u n_v$, by "trimming" F': for any flow path p from u_i to v_j in F' carrying a positive amount of flow, if p is a degenerate flow path from u_i to u_i, we add that amount of flow to the degenerate flow path from u to u; otherwise, we take the flow path p' by removing $u_i \to u_0$ (whenever $i \ne 0$) and $v_0 \to v_j$ (whenever $j \ne 0$) from p, so that p' uses only edges from G and has a natural image in G. We add that amount of flow to the flow path p' in G. Add these flow paths up to form F^*.

For every unit of flow in F' sent from u_i to v_j, there is a unit of flow in F^* sent from u to v. Therefore, the total amount of flow from u to v is $n_u n_v$, and F^* is feasible. Other than the handling of the degenerate flows, F^* is formed from F' by deleting and shortening existing flow paths. The degenerate flows contribute at most n_u to the congestion of F^* at u, where the original congestion of F' at u_0 is at least n_u due to flows of the form $u_0 \to u_i$. Therefore, the congestion at every vertex at most doubles after the procedure, and so

$$\min_{F^*} \operatorname{con}(F^*) \le 2 \min_{F'} \operatorname{con}(F')$$

as required.

Next, we prove that there exists $A > 0$ such that

$$A \min_{F'} \operatorname{con}(F') \le \min_{F^*} \operatorname{con}(F^*).$$

We take $A = 1/3$. For any solution F^* to the multicommodity flow problem on G with product demand $D^*(u, v) = n_u n_v$, we reverse the trimming procedure to form a solution F' to the multicommodity flow problem on G' with uniform demand, by "branching" the flow paths. For any flow path p from u to v carrying a positive amount of flow in F^*, subdivide that into $n_u n_v$ equal parts, and add that amount of flow to the flow path $u_i \to p \to v_j$ for all $0 \le i < n_u$, $0 \le j < n_v$. Add up these flows to form F'.

F' is feasible because the total amount of flow from u to v in F^* is $n_u n_v$, and so there is one unit of flow from u_i to v_j in F' for each (u_i, v_j). Let $c_{F'}(u_i)$ denote the congestion of vertex u_i in G' with respect to flow solution F'. First, we show that the congestion at u_0 dominates the total congestion at the other vertices u_i for $i \ne 0$:

$$\sum_{i \ne 0} c_{F'}(u_i) \le 2 c_{F'}(u_0).$$

This is because any flow path p from u_i to v_j for some $v \ne u$ includes u_i once but must also include u_0 once, and the total amount of flow through u_0 due to the flow paths $u_i \to u_j$ is $n_u^2 - (n_u - 1)$, whereas the amount of flow through each u_i for $i > 0$ due to the same flow paths is at most $2n_u$, netting a total of at most $2(n_u^2 - n_u)$.

Next, we show that the flow branching procedure does not increase the congestion at u_0 significantly. In fact, the congestion at u_0 only decreases due to some degenerate $u \to u$ paths in F^* becoming $u_i \to u_i$ paths in F', and is unaffected by other changes. Therefore,

$$\operatorname{con}(F^*)^2 = \sum_{u \in V} c_{F^*}(u)^2 \ge \sum_{u \in V} c_{F'}(u_0)^2 \ge \sum_{u \in V} \left(\frac{1}{3} \sum_{0 \le i < n_u} c_{F'}(u_i) \right)^2 \ge \left(\frac{\operatorname{con}(F')}{3} \right)^2,$$

so that

$$\frac{1}{3} \min_{F'} \operatorname{con}(F') \le \min_{F^*} \operatorname{con}(F^*).$$

The proof is complete. □

Proof of Lemma 7.3.8. For the first part, given a genus g graph $G = (V, E)$, we can embed G in a genus g surface. We can then embed the spike graph G' in the same surface by attaching degree-one vertices to G.

For the second part, we use a contrapositive argument. Suppose the constellation graph G' contains K_h as a minor for some $h \geq 3$. Then, we can obtain K_h from G' by some sequence of edge deletion, vertex deletion, and edge contraction. The claim is that all vertices u_i, $i \geq 1$, eventually gets deleted. This is because contraction of edge (u_0, u_i) is the same as deleting u_i, deletion of edge (u_0, u_i) must be followed by deleting u_i (otherwise it becomes an isolated vertex), and K_h cannot contain any degree-one vertex as $h \geq 3$. Furthermore, deletion of vertices u_i, $i \neq 0$, can be done in the very beginning of the sequence of operations. Therefore, applying the same sequence of operations on G' with the deletion of u_i, $i \neq 0$ moved to the beginning, at some point we obtain G, then we end up with K_h. This implies that G contains K_h as a minor. □

Proof of Theorem 7.1.6. Recall from Lemma 7.3.2 and Lemma 7.3.4 that

$$\lambda_2^*(G) \lesssim \alpha(G)^2 \cdot \left[\max_{s: V \to \mathbb{R}_{\geq 0}} \Lambda_s(G)\right]^{-2} = \alpha(G)^2 \cdot \left[\min_{F \in \mathcal{F}_\pi(G)} \mathrm{con}_\pi(F)\right]^{-2}. \tag{7.5}$$

Now, choose positive integers $M, \{n_u\}_{u \in V}$ using Proposition 7.3.5 so that $n_u/M \leq \pi(u) \leq 2n_u/M$ for all $u \in V$ and $\max_u n_u \leq 3(\sum_u n_u)/4$. By the trivial bound relating $\mathrm{con}_\pi(F)$ and $\mathrm{con}(F)$ as well as the explanation following Proposition 7.3.5, we have

$$\min_{F \in \mathcal{F}_\pi(G)} \mathrm{con}_\pi(F) \geq \pi_{\max}^{-1/2} \min_{F \in \mathcal{F}_\pi(G)} \mathrm{con}(F) \geq \pi_{\max}^{-1/2} \cdot \frac{1}{4M^2} \min_{F^*} \mathrm{con}(F^*).$$

where the minimum of F^* is taken overall multicommodity flows on G with product demand $D^*(u,v) = n_u n_v$. Now by Lemma 7.3.7, RHS is further lower bounded by a constant factor times

$$\pi_{\max}^{-1/2} \cdot \frac{1}{4M^2} \min_{F'} \mathrm{con}(F')$$

where the minimum of F' is taken over all multicommodity flows on the constellation graph G' with uniform demand.

We can now apply the congestion lower bounds in Lemma 3.5.10 to flows on G' to finish the proof. Let n' be the number of vertices in G'. Clearly, $n' \geq n$. We also have

$$n' = \sum_{u \in V} n_u \geq \sum_{u \in V} (M\pi(u) - 1) = M - n \geq M/2.$$

- If G is of genus $g \geq 1$, we first consider the case where $n \gtrsim \sqrt{g}$. Then, G' is of genus $g \geq 1$ by Lemma 7.3.8 with $n' \gtrsim \sqrt{g}$, so that $\mathrm{con}(F') \gtrsim (n')^2/\sqrt{g}$ for any F'. By

Lemma 7.3.3, $\alpha(G) \lesssim \log g$, and so

$$\begin{aligned}\lambda_2^*(G) &\lesssim \alpha(G)^2 \cdot \left[\min_F \mathrm{con}_\pi(F)\right]^{-2}\\ &\lesssim \alpha(G)^2 \pi_{\max} \cdot M^4 \left[\min_{F'} \mathrm{con}(F')\right]^{-2}\\ &\lesssim \alpha(G)^2 \pi_{\max} \cdot M^4 \cdot (g/(n')^4)\\ &\lesssim g \log^2 g \cdot \pi_{\max}.\end{aligned}$$

In the remaining case where $n < o(\sqrt{g})$, we use the trivial upper bound that $\lambda_2^*(G) \leq O(1)$[5] to deduce

$$\lambda_2^*(G) \lesssim 1 \lesssim \sqrt{g}/n \lesssim g \log^2 g \cdot \pi_{\max},$$

since $\pi_{\max} \geq 1/n$.

- If G is K_h-minor free for some $h \geq 3$, we first consider the case when $n \gtrsim h\sqrt{\log h}$. Then, G' is K_h-minor free by Lemma 7.3.8 with $n' \gtrsim h\sqrt{\log h}$, so that $\mathrm{con}(F') \gtrsim (n')^2/(h\sqrt{\log h})$ for any F'. By Lemma 7.3.3, $\alpha(G) \lesssim h^2$, and so

$$\begin{aligned}\lambda_2^*(G) &\lesssim \alpha(G)^2 \cdot \left[\min_F \mathrm{con}_\pi(F)\right]^{-2}\\ &\lesssim \alpha(G)^2 \pi_{\max} \cdot M^4 \left[\min_{F'} \mathrm{con}(F')\right]^{-2}\\ &\lesssim \alpha(G)^2 \pi_{\max} \cdot M^4 \cdot (h^2 \log h/(n')^4)\\ &\lesssim h^6 \log h \cdot \pi_{\max}.\end{aligned}$$

In the remaining case where $n < o(h\sqrt{\log h})$, we again use $\lambda_2^*(G) \leq O(1)$ and $\pi_{\max} \geq 1/n$ to deduce that

$$\lambda_2^*(G) \lesssim 1 \lesssim h\sqrt{\log h}/n \lesssim h^6 \log h \cdot \pi_{\max}.$$

This completes the proof of Theorem 7.1.6. ☐

Remark 7.3.9 (Vertex Expansion Bounds)**.** *In the uniform distribution case, Theorem 7.1.6 combined with the vertex Cheeger inequality implies that $\psi(G) \leq O(\sqrt{g \log^2 g \cdot \log \Delta / n})$ if G is of genus g, and $\psi(G) \leq O(\sqrt{h^6 \log h \cdot \log \Delta / n})$ if G is K_h-minor free. As remarked in the end of Section 3.5.5, Biswal, Lee, and Rao [BLR10] obtained a better vertex expansion bound with one less $\sqrt{\log \Delta}$ factor. However, their result does not come with a polynomial-time algorithm to obtain cuts $S \subseteq V$ that satisfy the same vertex expansion upper bound, while our bound is attainable via the Cheeger rounding algorithm for reweighted eigenvalues.*

[5]This follows, for example, from the easy direction of Theorem 4.1.3 and that $\psi(G) \leq 1$.

7.4 Bounding Higher Eigenvalues

In this section, we prove Theorem 7.1.7 and Theorem 7.1.8 which upper bound $\lambda_k^*(G)$ for special classes of graphs G. Similar to the $\lambda_2^*(G)$, in the uniform distribution case a small tweak of the proof in [KLPT11] suffices, whereas in the general distribution case more work is needed. We make frequent references to the review of [KLPT11] in Section 3.5.6 and the reader is advised to read that subsection before proceeding.

7.4.1 The Uniform Distribution Case

We modify the proof of Lemma 3.5.15 to upper bound $\lambda_k^*(G)$ using two metric parameters: metric padding and metric spreading. The two parameters are controlled respectively using Lemma 3.5.16 and Lemma 3.5.18, and Theorem 7.1.7 follows after combining everything. The definitions of the metric parameters are due to [KLPT11] and we copy here for convenience.

Definition 7.4.1 (Metric Padding Parameter (restatement of Definition 3.5.13)). *Let (X, d_X) be a finite metric space. For any partitioning P of X, using $P(x)$ to denote the partition that contains $x \in X$, define the padding parameter $\beta(P, \gamma)$ to be the infimal value of $\beta \geq 1$ such that*

$$|\{x \in X : B(x, \gamma/\beta) \subseteq P(x)\}| \geq \frac{|X|}{2}.$$

In other words, at least half of the points in X satisfy that all points (γ/β)-close to it are in the same partition.

Further, let

$$\beta_\gamma(X, d_X) := \inf_P \beta(P, \gamma),$$

where the infimum is taken over all parititoning P of (X, d_X) where each partition has diameter at most γ.

Definition 7.4.2 (Metric Spreading Parameter (restatement of Definition 3.5.14)). *Let $G = (V, E)$ be a graph. Let $s : V \to \mathbb{R}_{\geq 0}$ be a weight function on the vertices, and d_s be the vertex-weighted shortest path metric on V (c.f. (3.15)). For $\varepsilon > 0$ and a collection Ψ of nonempty subsets of V, say that s is (Ψ, ε)-spreading if, for every $S \in \Psi$ one has*

$$\frac{1}{|S|^2} \sum_{u,v \in S} d_s(u, v) \geq \varepsilon \cdot \sqrt{\sum_{u \in V} s(u)^2}.$$

Write $\varepsilon_\Psi(G, s)$ to be the maximal value of ε such that s is (Ψ, ε)-spreading.

Lemma 7.4.3. *Let $G = (V, E)$ be a graph and let $\pi = \mathbb{1}/n$ be the uniform distribution on V. For any $2 \le k \le n$, the following holds. For any weight function $s : V \to \mathbb{R}_{\ge 0}$ satisfying*

$$\sum_{u \in V} s(u)^2 = 1,$$

we have

$$\lambda_k^*(G) \lesssim \frac{1}{\varepsilon^2 n} \left(\beta_{\varepsilon/2}(V, d_s)\right)^2,$$

where $\varepsilon = \varepsilon_{\lfloor n/4k \rfloor}(G, s)$.

Proof. Let $Q \in \mathbb{R}_{\ge 0}^{V \times V}$ be a feasible reweighting for the $\lambda_k^*(G)$ program in Definition 7.1.1. That means Q is supported on E, $\sum_{v \in V} Q(u, v) = 1$ for all $u \in V$, and $Q(u, v) = Q(v, u)$ for all $u, v \in V$. We would like to prove that

$$\lambda_k(I - Q) \lesssim \frac{1}{\varepsilon^2 n} \left(\beta_{\varepsilon/2}(V, d_s)\right)^2.$$

Write $\beta = \beta_{\varepsilon/2}(V, d_s)$. By Definition 7.4.1, there exists a partitioning P of V such that:

- Each partition $P(v)$ has diameter $\le \varepsilon/2$;
- At least half of the vertices $v \in V$ satisfy $B(v, \varepsilon/(2\beta)) \subseteq P(v)$.

As in Lemma 3.5.15, from P we can extract $q \ge 2k$ disjoint subsets $T_1, \dots, T_q$ and their "cores" $T_1', \dots, T_q'$, so that $B(u, \varepsilon/(2\beta)) \subseteq T_i$ for each $u \in T_i'$, and that each T_i' has size between $r/2$ and r, where $r = \lfloor n/4k \rfloor$.

Now we use these sets to define disjointly supported vectors with small Rayleigh quotient (energy divided by mass) with respect to $I - Q$. Consider the following "smooth localization" $f_1, \dots, f_q$ of the subsets:

$$f_i(u) := \max\left(0, \frac{\varepsilon}{2\beta} - d_s(u, T_i')\right).$$

First, note the following about the mass of each f_i:

$$\|f_i\|^2 = \sum_{u \in V} f_i(u)^2 \ge \left(\frac{\varepsilon}{2\beta}\right)^2 \cdot |T_i'| \ge \frac{\varepsilon^2 n}{32\beta^2 k}.$$

Next, observing that each f_i is supported on T_i, the total energy with respect to $I - Q$ is

$$\begin{aligned}\sum_{i \in [q]} \mathcal{E}(f_i) &= \sum_{i \in [q]} f_i^T (I - Q) f_i \\ &= \sum_{i \in [q]} \sum_{uv \in E} Q(u,v)(f_i(u) - f_i(v))^2 \\ &\leq \sum_{i \in [p]} \sum_{u \in T_i} \sum_{v : uv \in E} Q(u,v)(f_i(u) - f_i(v))^2 \\ &\leq \sum_{i \in [p]} \sum_{u \in T_i} \sum_{v : uv \in E} Q(u,v) d_s(u,v)^2 \\ &\leq 2 \sum_{i \in [p]} \sum_{u \in T_i} \sum_{v : uv \in E} Q(u,v)(s(u)^2 + s(v)^2) \\ &\leq 4 \sum_{u \in V} s(u)^2 = 4.\end{aligned}$$

The last inequality is because the coefficient of $s(u)^2$ is at most

$$2 \left(\sum_{v : uv \in E} Q(u,v) + \sum_{v : vu \in E} Q(v,u) \right) \leq 4,$$

and the other inequalities proceed as in Lemma 3.5.15. Since the number of functions is at least $2k$, there exists k such f_i's with disjoint support, such that $\mathcal{E}(f_i) \leq 4/k$. Therefore,

$$\frac{\mathcal{E}(f_i)}{\|f_i\|^2} \leq \frac{128}{\varepsilon^2 n} \beta^2$$

for the chosen f_i's, and applying Proposition 2.5.4 to the Q-weighted graph we obtain the desired upper bound on $\lambda_k(I - Q)$. □

As in Section 3.5.6, combining Lemma 7.4.3 with Lemma 3.5.16, Lemma 3.5.17, and Lemma 3.5.18 yields Theorem 7.1.7. Compared to Lemma 3.5.15, there is no dependence on Δ in the upper bound, which explains the factor Δ improvement in Theorem 7.1.7.

7.4.2 The General Case

To generalize Theorem 7.1.7 to arbitrary distributions on V, we introduce π-weighted variants of metric padding and metric spreading parameters, and use them to control $\lambda_k^*(G)$.

We show that the π-weighted metric padding parameter satisfies essentially the same upper bound as the unweighted counterpart. As for the π-weighted metric spreading parameter, we reduce to the unweighted case using, again, the constellation graph in Definition 7.3.6. Below, we use the proof of Theorem 3.5.12 as a blueprint, and explain the modifications needed for proving Theorem 7.1.8.

Rayleigh Quotient to Metric Parameters

The first component of the proof is to relate $\lambda_k^*(G)$ to the π-weighted version of metric padding parameter in Definition 7.4.1 and metric spreading parameter in Definition 7.4.2. We overload the unweighted notations as the π-weight will be clear from context.

Definition 7.4.4 (Metric Padding Parameter, π-Weighted Version)**.** *Let (X, d_X) be a finite metric space, and let $\pi : X \to \mathbb{R}^+$ be a distribution on X. For any partitioning P of X, using $P(x)$ to denote the partition that contains $x \in X$, define the π-weighted padding parameter $\beta(P, \gamma)$ to be the infimal value of $\beta \geq 1$ such that*

$$\pi\left(\{x \in X : B(x, \gamma/\beta) \subseteq P(x)\}\right) \geq \frac{1}{2} - o(1).$$

In other words, by π-weight at least almost half of the points in X satisfy that its (γ/β)-neighborhood is in the same partition. Further, let

$$\beta_\gamma(X, d_X) := \inf_P \beta(P, \gamma),$$

where the infimum is taken over all parititoning P of (X, d_X) where each partition has diameter at most γ.

For metric spreading, we overload the old notation, as the π-weight will be clear from context.

Definition 7.4.5 (Metric Spreading Parameter, π-Weighted Version)**.** *Let $G = (V, E)$ be a graph, and let $\pi : V \to \mathbb{R}^+$ be a distribution on the vertices. Let $s : V \to \mathbb{R}_{\geq 0}$ be a weight function on the vertices, and d_s be the vertex-weighted shortest path metric on V. For $\varepsilon > 0$ and a collection Ψ of nonempty subsets of V, say that s is (Ψ, ε)-spreading if, for every $S \in \Psi$ one has*

$$\frac{1}{\pi(S)^2} \sum_{u,v \in S} \pi(u)\pi(v) d_s(u, v) \geq \varepsilon \cdot \sqrt{\sum_{u \in V} \pi(u) s(u)^2}.$$

Write $\varepsilon_\Psi(G, s)$ to be the maximal value of ε such that s is (Ψ, ε)-spreading.

We are now ready to state and prove an upper bound on $\lambda_k^*(G)$ for general π weights. We impose a mild condition on π to make the proof go through.

Lemma 7.4.6. *Let $G = (V, E)$ be a graph. For any $2 \le k \le n$, the following holds. For any distribution π on V such that $\pi_{\max} := \max_{u \in V} \pi(u) \le 1/k$, and for any weight function $s : V \to \mathbb{R}_{\ge 0}$ satisfying*

$$\sum_{u \in V} \pi(u) s(u)^2 = 1,$$

we have

$$\lambda_k^*(G) \lesssim \frac{1}{\varepsilon^2} \left(\beta_{\varepsilon/2}(V, d_s)\right)^2,$$

where $\varepsilon = \varepsilon_{\Psi_{r,5r}}(G, s)$, $r = 1/(4k)$, and $\Psi_{a,b} := \{S \subseteq V : \pi(S) \in [a, b]\}$ is the collection of subsets of V whose π-weight is between a and b.

Proof. Let Q be a feasible reweighting for the $\lambda_k^*(G)$ program in Definition 7.1.1. That means Q is supported on E, $\sum_{u \in V} \pi(u) Q(u, v) = \pi(v)$ for all $v \in V$, and $\pi(u) Q(u, v) = \pi(v) Q(v, u)$ for all $u, v \in V$. The last constraint can be written as $\Pi Q = Q^T \Pi$ where $\Pi := \operatorname{diag}(\pi)$. This means that ΠQ is symmetric, and so is $\Pi^{1/2} Q \Pi^{-1/2}$. As $I - Q$ and $I - \Pi^{1/2} Q \Pi^{-1/2}$ are similar, by Fact 2.4.1 our task becomes proving that

$$\lambda_k(I - \Pi^{1/2} Q \Pi^{-1/2}) \lesssim \frac{1}{\varepsilon^2 n} \left(\beta_{\varepsilon/2}(V, d_s)\right)^2.$$

Write $\beta = \beta_{\varepsilon/2}(V, d_s)$. By Definition 7.4.4, there exists a partitioning P of V such that:

- Each partition has diameter $\le \varepsilon/2$;
- Vertices $v \in V$ carrying at least half π-weight satisfy $B(v, \varepsilon/(2\beta)) \subseteq P(v)$, with $P(v)$ denoting the partition containing v.

Let $S_1, \ldots, S_\ell$ be the partitions in P. We use the definition of ε to show that the π-weight of each S_i is at most $r = 1/(4k)$. For if not, then since each vertex has π-weight at most $1/k = 4r$, by possibly removing some vertices from S_i we obtain a subset S^* of V whose π-weight is between r and $r + 4r = 5r$, and since its diameter is $\le \varepsilon/2$ we have

$$\frac{1}{\pi(S^*)^2} \sum_{u,v \in S^*} \pi(u)\pi(v) d_s(u, v) \le \frac{\varepsilon}{2} = \frac{\varepsilon}{2} \cdot \sqrt{\sum_{u \in V} \pi(u) s(u)^2},$$

which contradicts the definition of ε, as $S^* \in \Psi_{r,5r}$.

Therefore, the π-weight of each S_i is at most r. As in Lemma 3.5.15, we can then extract $q \geq 2k$ disjoint subsets $T_1, \ldots, T_q$ and their "cores" $T'_1, \ldots, T'_q$, so that $B(u, \varepsilon/(2\beta)) \subseteq T_i$ for each $u \in T'_i$, and that each T'_i has π-weight between $r/2$ and r.

Now we use these sets to define disjointly supported vectors with small Rayleigh quotient (energy divided by mass) with respect to $(I - \Pi^{1/2}Q\Pi^{-1/2})$. Consider the following "smooth localization" $f_1, \ldots, f_q$ of the subsets:

$$f_i(u) := \max\left(0, \frac{\varepsilon}{2\beta} - d_s(u, T'_i)\right),$$

and define $\tilde{f}_i := \Pi^{1/2} f_i$. First, note the following about the mass of each $\tilde{f}_i$:

$$\left\|\tilde{f}_i\right\|^2 = \sum_{u \in V} \pi(u) f_i(u)^2 \geq \left(\frac{\varepsilon}{2\beta}\right)^2 \cdot \pi(T'_i) \geq \frac{\varepsilon^2}{32\beta^2 k}.$$

Observe that since f_i is supported on T_i and that Π is diagonal, $\tilde{f}_i$ is also supported on T_i. The total energy with respect to $(I - \Pi^{1/2}Q\Pi^{-1/2})$ is then

$$\begin{aligned}
\sum_{i \in [q]} \mathcal{E}(\tilde{f}_i) &= \sum_{i \in [q]} \tilde{f}_i^T (I - \Pi^{1/2}Q\Pi^{-1/2}) \tilde{f}_i \\
&= \sum_{i \in [q]} f_i^T (\Pi - \Pi Q) f_i \\
&= \sum_{i \in [q]} \sum_{uv \in E} \pi(u) Q(u,v) (f_i(u) - f_i(v))^2 \\
&\leq \sum_{i \in [p]} \sum_{u \in T_i} \sum_{v : uv \in E} \pi(u) Q(u,v) (f_i(u) - f_i(v))^2 \\
&\leq \sum_{i \in [p]} \sum_{u \in T_i} \sum_{v : uv \in E} \pi(u) Q(u,v) d_s(u,v)^2 \\
&\leq 2 \sum_{i \in [p]} \sum_{u \in T_i} \sum_{v : uv \in E} \pi(u) Q(u,v) (s(u)^2 + s(v)^2) \\
&\leq 4 \sum_{u \in V} \pi(u) s(u)^2 = 4.
\end{aligned}$$

The last inequality is because the coefficient of $s(u)^2$ is at most

$$2\left(\sum_{v : uv \in E} \pi(u) Q(u,v) + \sum_{v : vu \in E} \pi(v) Q(v,u)\right) \leq 4\pi(u),$$

and the other inequalities go through as in Lemma 3.5.15. Since the number of functions is at least $2k$, there exists k such $\tilde{f}_i$'s with disjoint support, such that $\mathcal{E}(\tilde{f}_i) \leq 4/k$. Therefore,

$$\frac{\mathcal{E}(\tilde{f}_i)}{\left\|\tilde{f}_i\right\|^2} \leq \frac{128}{\varepsilon^2}\beta^2$$

for the chosen $\tilde{f}_i$'s, and Proposition 2.5.4 yields the desired upper bound on $\lambda_k(I-Q)$. □

Bounds on Metric Padding

The second component of the proof is to upper bound the π-weighted metric padding parameter in Definition 7.4.4. The strategy is simply to reduce to the unweighted case using the constellation graph in Definition 7.3.6.

To prepare for the reduction, we prove a variant of Proposition 7.3.5 that allows us to "discretize" the distribution π with arbitrarily small error.

Proposition 7.4.7. *Given $c > 0$ and n positive real numbers $a_1, \ldots, a_n$ summing to 1, there exists positive integers $M, b_1, \ldots, b_n$ such that $b_i/M \leq a_i \leq b_i/M + c$ for all $i \in [n]$.*

Proof. Let Z be a positive integer such that $1/Z < \zeta$, and let $M \in \mathbb{N}$ be such that $Ma_i \geq Z$ for all $i \in [n]$. Note that $M \geq Z$ and so $1/M \leq 1/Z$. Let $b_i := \lfloor Ma_i \rfloor$. Then, clearly $b_i/M \leq a_i$, and

$$a_i \leq \frac{b_i + 1}{M} = \frac{b_i}{M} + \frac{1}{M} \leq \frac{b_i}{M} + \zeta,$$

as desired. □

Lemma 7.4.8 (Upper Bounding Metric Padding Parameter, π-Weighted Version)**.** *Let $G = (V, E)$ be a graph, and $\pi : V \to \mathbb{R}^+$ be any distribution on V. Let d_s be a shortest path metric corresponding to a vertex weight function $s : V \to \mathbb{R}_{\geq 0}$. Let $\gamma > 0$. Then,*

- *If G is planar, then $\beta_\gamma(V, d_s) \leq O(1)$.*
- *If G has genus $g \geq 1$, then $\beta_\gamma(V, d_s) \leq O(\log g)$.*
- *If G excludes K_h as a minor, then $\beta_\gamma(V, d_s) \leq O(h^2)$.*

So, in these cases, G satisfies the same metric padding upper bound as the unweighted case in Lemma 3.5.16.

Proof. First, apply Proposition 7.4.7 to find positive integers $\{n_u\}_{u \in V}$ and M, such that $M = \sum_{u \in V} n_u$ and n_u/M is arbitrarily close to $\pi(u)$ for all $u \in V$. Then, construct the constellation graph $G' = (V', E')$ for G and $\{n_u\}$ as in Definition 7.3.6 and define the following vertex weight function $s' : V' \to \mathbb{R}_{\geq 0}$ by

$$s'(u_i) := \begin{cases} s(u), & \text{if } i = 0; \\ 0, & \text{otherwise.} \end{cases}$$

Now, by Lemma 3.5.16, there exists a partitioning P' of the metric space $(V', d_{s'})$, such that $\text{diam}_{d_{s'}}(P') \leq \gamma$, and the *unweighted* metric padding parameter $\beta(P', \gamma)$ is upper bounded by $O(1)$ (resp. $O(\log g)$ and $O(h^2)$) in the planar (resp. bounded genus and K_h-minor free) case. That means for this specific value of $\beta = \beta(P', \gamma)$, at least half of the vertices of V' are in the set of "core" vertices

$$S'_{\text{core}} := \{u_i \in V' : B_{d_{s'}}(u_i, \gamma/\beta) \subseteq P'(u_i)\},$$

where again $P'(u_i)$ denotes the partition in P' that u_i belongs to. The plan, then, is to construct a partitioning P of (V, d_s), such that $\text{diam}_{d_s}(P) \leq \gamma$ and the π-weighted metric padding parameter $\beta(P, \gamma)$ is at most β.

It is clear that $d_{s'}(u_i, u_j) = s(u)$ and $d_{s'}(u_i, v_j) = d_s(u, v)$ for $u \neq v$. So, we may assume without loss of generality that, if $u_i \in S_{\text{core}}$ for some $0 \leq i < n_u$, then $u_0 \in S'_{\text{core}}$. Take P' and consider the partitioning P of V induced by the restriction of P' on $\{u_0\}_{u \in V}$. Formally, if $S'_1, \ldots, S'_\ell$ are the partitions in P', then the nonempty subsets among

$$S_i := \{u \in V : u_0 \in S'_i\}$$

are the partitions in P. Then, clearly $\text{diam}_{d_s}(S_i) \leq \gamma$ as well, and if $u_0 \in S'_{\text{core}}$, that means $v_0 \in P'(u_0)$ for all $v \in V$ such that $d_s(u, v) \leq \gamma/\beta$. This in turn implies that $B_{d_s}(u, \gamma/\beta) \subseteq P(u)$. Then, the total π-weight of all "core" vertices in V with respect to P is at least

$$\sum_{u : u_0 \in S'_{\text{core}}} \pi(u) \geq \left(\sum_{u : u_0 \in S'_{\text{core}}} \frac{n_u}{M} \right) - o(1) \geq \left(\sum_{u_i \in S'_{\text{core}}} \frac{1}{M} \right) - o(1) \geq \frac{1}{2} - o(1),$$

which implies that the π-weighted padding parameter $\beta(P, \gamma)$ is at most β. □

Bounds on Metric Spreading

The third and final component of the proof is to lower bound the π-weighted metric spreading parameter in Definition 7.4.5. Again, we start with a primal-dual result relating maximum metric spreading to minimum subset flow congestion. Then, we use the constellation graph in Definition 7.3.6 to reduce to the unweighted case, and use existing lower bound results in Section 3.5.6.

As in Section 7.3, for any multicommodity flow solution F we define the π-weighted congestion as

$$\mathrm{con}_\pi(F) := \sqrt{\sum_{u \in V} \frac{c_F(u)^2}{\pi(u)}},$$

where $c_F(u)$ is the total amount of flow in F passing through vertex u. Similar to how Lemma 7.3.4 generalizes Lemma 3.5.9, Lemma 7.4.9 in the following generalizes Lemma 3.5.17 that establishes strong duality between metric spreading parameter and subset flow congestion. The proof is deferred to Appendix C.

Lemma 7.4.9 (Flow/Metric Duality, π-Weighted Version)**.** *Let $G = (V, E)$ be a graph with vertex distribution $\pi : V \to \mathbb{R}^+$. Let Ψ be a collection of nonempty subsets of V. Then, the π-weighted metric spreading maximization problem is strongly dual to the "subset flow" minimum π-weighted congestion problem:*

$$\max_{s:V \to \mathbb{R}_{\geq 0}} \varepsilon_\Psi(G, s) = \min_{F \in \mathcal{F}_\pi^\Psi(G)} \mathrm{con}_\pi(F),$$

where $\mathcal{F}_\pi^\Psi(G)$ is the set of all multicommodity flows on G whose demand graph satisfies

$$D_h(u, v) = \pi(u)\pi(v) \cdot \sum_{A \in \Psi : \{u,v\} \subseteq A} \frac{h(A)}{\pi(A)^2} \qquad \forall u, v \in V$$

for some distribution h on Ψ, i.e. $h : \Psi \to \mathbb{R}_{\geq 0}$ with $\sum_{A \in \Psi} h(A) = 1$.

Having established the connection between metric spreading and subset flow congestion, we end up with a similar situation as Section 7.3, where we need to lower bound $\mathrm{con}_\pi(F)$ for some solution F to a multicommodity flow problem.

Again, the relation

$$\mathrm{con}_\pi(F) \geq \pi_{\max}^{-1/2} \left(\sum_{v \in V} c_F(v)^2 \right)^{1/2} = \pi_{\max}^{-1/2} \cdot \mathrm{con}(F) \tag{7.6}$$

allows us to reduce the task at hand to lower bounding the unweighted congestion $\mathrm{con}(F)$. Then, we relate the current multicommodity flow problem $\mathcal{F}_\pi^\Psi(G)$, where the demand graph is a weighted sum of subset π-product demands, to the multicommodity flow problem

$\mathcal{F}^{\Psi'}(G')$ (see Lemma 3.5.18 for definition) on an auxiliary graph $G' = (V', E')$ and an appropriate collection of subsets Ψ' of V', and use Lemma 3.5.18 to lower bound the congestion of F for $F \in \mathcal{F}^{\Psi'}(G')$, where the demand graph is a weighted sum of subset uniform demands.

A natural choice for the auxiliary graph is the constellation graph in Definition 7.3.6, since we already know from Lemma 7.3.8 that the special properties of G are preserved. The following lemma summarizes the key properties of the reduction. Properties (i) and (iii) ensure that we can apply the results in [KLPT11], and properties (ii) ensure that we obtain an explicit congestion lower bound depending on n, k, and other simple parameters.

Lemma 7.4.10 (Properties of Constellation Graph). *Let $G = (V, E)$ be a graph with distribution π on V. Let $\{n_u\}_{u\in V}, M$ be positive integers from Proposition 7.4.7 such that n_u/M is arbitrarily close to $\pi(u)$ for all $u \in V$, and G' be the constellation graph constructed from G and $\{n_u\}_{u\in V}$. Given $r \in (0, 1)$, let $\Psi := \Psi_{r,5r} = \{S \subseteq V : \pi(S) \in [r, 5r]\}$ and*

$$\Psi' := \Big\{\{u_i : u \in S, 0 \le i < n_u\} \,|\, S \in \Psi\Big\}.$$

Then, the following properties hold:

- *(i) If G is planar, so is G'. If G is of genus $g \ge 1$, so is G'. If G is K_h-minor free for $h \ge 3$, so is G'.*
- *(ii) $|S'| \in [rM/2, 6rM]$ for all $S' \in \Psi'$.*
- *(iii) For some absolute constants $A, B > 0$,*

$$A \cdot \min_{F\in\mathcal{F}^{\Psi}_{\pi}(G)} \mathrm{con}(F) \le \min_{F'\in\mathcal{F}^{\Psi'}(G')} \mathrm{con}(F') \le B \cdot \min_{F\in\mathcal{F}^{\Psi}_{\pi}(G)} \mathrm{con}(F).$$

Proof. We prove the three properties one by one. (i) follows directly from Lemma 7.3.8. (ii) is because for all $S \in \Psi$ whose corresponding set in Ψ' is $S' := \{u_i : u \in S, 0 \le i < n_u\}$, one has $|S'|/M \approx \pi(S) \in [r, 5r]$ for all $S' \in \Psi'$. Since the approximation can be arbitrarily fine, we can make sure that $|S'| \in [rM/2, 6rM]$ for all $S' \in \Psi'$.

It remains to prove (iii). The proof is similar to Lemma 7.3.7. First, we prove that

$$\frac{1}{3} \min_{F\in\mathcal{F}^{\Psi}_{\pi}(G)} \mathrm{con}(F) \le \min_{F'\in\mathcal{F}^{\Psi'}(G')} \mathrm{con}(F').$$

Let $F' \in \mathcal{F}^{\Psi'}(G')$ be a solution to the subset flow problem $\mathcal{F}^{\Psi'}(G')$. That means there is a distribution $h' : \Psi' \to \mathbb{R}_{\geq 0}$ such that the demand graph $D_{h'}$ satisfies

$$D_{h'}(u_i, v_j) = \sum_{S' \in \Psi' : \{u_i, v_j\} \subseteq S'} \frac{h'(S')}{|S'|^2} \qquad \forall u_i, v_j \in V'. \tag{7.7}$$

Define $h : \Psi \to \mathbb{R}_{\geq 0}$ so that $h(S) = h'(S')$ if $S \in \Psi$ corresponds to $S' \in \Psi'$. We are looking for a flow solution F to $\mathcal{F}_{\pi}^{\Psi}(G)$ such that the demand graph D_h satisfies

$$D_h(u, v) = \pi(u)\pi(v) \sum_{S \in \Psi : \{u, v\} \subseteq S} \frac{h(S)}{\pi(S)^2} \qquad \forall u, v \in V. \tag{7.8}$$

As in Lemma 7.3.7, fix F' and construct F by "trimming" F' as follows: for each flow path p' in F' from u_i to v_j carrying a positive amount of flow, we take the flow path p by removing $u_i \to u_0$ (whenever $i \neq 0$) and $v_0 \to v_j$ (whenever $j \neq 0$) from p', so that p only uses edges in G and has a natural image in G. Add that amount of flow that flow path in G. Add these flow paths up to form F.

We show that the congestion of F is at most $2 \cdot \mathrm{con}(F')$ and that F *approximately* satisfies the demand graph D_h. First, other than the addition of the degenerate flow paths $u \to u$ the congestion only decreases due to the flow trimming. As in Lemma 7.3.7, the contribution of the degenerate flow paths to the congestion at u is at most the original congestion at u_0, so the overall congestion at most doubles after the trimming. Next, the amount of flow sent between u and v is

$$\begin{aligned}
\sum_{\substack{0 \leq i < n_u \\ 0 \leq j < n_v}} D'_h(u_i, v_j) &= n_u n_v \cdot \sum_{S' \in \Psi' : \{u_0, v_0\} \subseteq S'} \frac{h'(S')}{|S'|^2} \\
&\approx M^2 \pi(u)\pi(v) \cdot \sum_{S \in \Psi : \{u, v\} \subseteq S} \frac{h(S)}{M^2 \pi(S)^2} = D_h(u, v).
\end{aligned}$$

Therefore, by scaling up F by a factor of at most $3/2$ and then removing some flow paths, we obtain an exact solution to the subset flow problem $\mathcal{F}_{\pi}^{\Psi}(G)$, with congestion at most $3 \cdot \mathrm{con}(F')$.

Next, we prove that

$$\min_{F' \in \mathcal{F}^{\Psi'}(G')} \mathrm{con}(F') \leq 4 \cdot \min_{F \in \mathcal{F}_{\pi}^{\Psi}(G)} \mathrm{con}(F).$$

The proof is by following the branching procedure in Lemma 7.3.7. Let $F \in \mathcal{F}_\pi^\Psi(G)$ be a solution to the subset flow problem $\mathcal{F}_\pi^\Psi(G)$. That means there is a distribution $h : \Psi \to \mathbb{R}_{\geq 0}$ such that the demand graph D_h satisfies (7.8). Define $h' : \Psi \to \mathbb{R}_{\geq 0}$ so that $h(S) = h'(S')$ if $S \in \Psi$ corresponds to $S' \in \Psi'$. We are looking for a flow solution F' to $\mathcal{F}^{\Psi'}(G')$ such that the demand graph $D_{h'}$ satisfies (7.7).

Construct F' as follows: for each flow path p in F from u to v carrying a positive amount of flow, send $(n_u n_v)^{-1}$ portion of that flow from u_i to v_j along the path $u_i \to p \to v_j$ for each $0 \leq i < n_u$ and $0 \leq j < n_v$.

We first show that the branching procedure at most triples the original congestion, and then show that F' is approximately feasible for $D_{h'}$. Following the same argument as Lemma 7.3.7, the congestion at u_0 dominates the total congestion at the other vertices u_i for $i \neq 0$:

$$\sum_{i \neq 0} c_{F'}(u_i) \leq 2 c_{F'}(u_0).$$

Furthermore, there is a net decrease in congestion at u_0 due to the branching procedure. Therefore, as in Lemma 7.3.7 we have $\text{con}(F') \leq 3 \cdot \text{con}(F)$. To show approximate feasibility, note that the total amount of flow from u_i to v_j is

$$\frac{1}{n_u n_v} D_h(u,v) \approx \frac{1}{M^2 \pi(u)\pi(v)} \cdot \pi(u)\pi(v) \cdot \sum_{S \in \Psi : \{u,v\} \subseteq S} \frac{M^2 \cdot h(S)}{|S|^2} = D_{h'}(u_i, v_j).$$

Therefore, by scaling up F' by a factor of at most $4/3$ and then removing some flow paths, we obtain an exact solution to the subset flow problem, with congestion at most $4 \cdot \text{con}(F)$.

This completes the proof of (iii) and hence of the lemma. □

Putting It All Together

Proof of Theorem 7.1.8. Start with the eigenvalue upper bound in Lemma 7.4.6 that

$$\lambda_k^*(G) \lesssim \frac{1}{\varepsilon^2} \left(\beta_{\varepsilon/2}(V, d_s) \right)^2, \tag{7.9}$$

where $\varepsilon = \varepsilon_\Psi(G, s)$ with $\Psi = \Psi_{r,5r} := \{S \subseteq V : \pi(S) \in [r, 5r]\}$ and $r = 1/(4k)$. Find positive integers $\{n_u\}$ and M so that n_u/M is arbitrarily close to $\pi(u)$ for all $u \in V$. Note that by the flow-metric duality in Lemma 7.4.9, property (iii) in Lemma 7.4.10, and (7.6),

$$\varepsilon = \min_{F \in \mathcal{F}_\pi^\Psi(G)} \text{con}_\pi(F) \geq \pi_{\max}^{-1/2} \cdot \min_{F \in \mathcal{F}_\pi^\Psi(G)} \text{con}(F) \gtrsim \pi_{\max}^{-1/2} \cdot \min_{F' \in \mathcal{F}^{\Psi'}(G')} \text{con}(F'),$$

where G' is the constellation graph given G and $\{n_u\}$, and Ψ' is defined as in Lemma 7.4.10.

We may now apply the congestion lower bound in Lemma 3.5.10 case by case. Let $n' := |V'|$ be the number of vertices of G'. Then,

$$1 = \sum_{u \in V} \pi(u) \approx \sum_{u \in V} \frac{n_u}{M} = \frac{n'}{M}.$$

Let $M^{\Psi'} := \sum_{S' \in \Psi'} h(S')/|S'|^2$ as in Lemma 3.5.10. By property (ii) in Lemma 7.4.10, $M^{\Psi'} = \Theta(1/(rM)^2) = \Theta((k/n')^2)$.

- If G is planar, then G' is also planar by property (i) in Lemma 7.4.10. There are two cases to consider per Lemma 3.5.10. If $M^{\Psi'} < o(1)$ then by Lemma 3.5.10 it is possible to take $\varepsilon = \Theta(\pi_{\max}^{-1/2}(n')^{-1/2}(M^{\Psi'})^{-1/4}) = \Theta(1/\sqrt{k \cdot \pi_{\max}})$. Now by Lemma 7.4.8, $\beta_{\varepsilon/2}(V, d_s) \leq O(1)$. Plugging all the values in (7.9) we get

  $$\lambda_k^*(G) \lesssim k\pi_{\max}.$$

  If $M^{\Psi'} \geq \Omega(1)$, that means $k \geq \Omega(n') \geq \Omega(n)$. Since $\pi_{\max} \geq 1/n$ and $\lambda_k^*(G) \leq O(1)$ by the trivial upper bound (see the proof of Theorem 7.1.6), we obtain the desired upper bound via

  $$\lambda_k^*(G) \lesssim 1 \lesssim \frac{k}{n} \lesssim k\pi_{\max}.$$

- If G has genus $g \geq 1$, then G' is also of genus $g \geq 1$ by property (i) in Lemma 7.4.10. Again there are two cases to consider. If $gM^{\Psi'} < o(1)$ then by Lemma 3.5.10 it is possible to take $\varepsilon = \Theta(\pi_{\max}^{-1/2}(gn')^{-1/2}(M^{\Psi'})^{-1/4}) = \Theta(1/\sqrt{gk \cdot \pi_{\max}})$. Now by Lemma 7.4.8, $\beta_{\varepsilon/2}(V, d_s) \leq O(\log g)$. Plugging all the values in (7.9) we get

  $$\lambda_k^*(G) \lesssim k\pi_{\max} \cdot g \log^2 g.$$

  If $gM^{\Psi'} \geq \Omega(1)$, then a similar argument as in the first case establishes the required upper bound on $\lambda_k^*(G)$.

- If G is K_h-minor free for $h \geq 3$, then G' is also K_h-minor free by property (i) in Lemma 7.4.10. Again there are two cases to consider. If $(h^2 \log h)M^{\Psi'} < o(1)$ then by Lemma 3.5.10 it is possible to take $\varepsilon = \Theta(\pi_{\max}^{-1/2}(h^2 \log h \cdot n')^{-1/2}(M^{\Psi'})^{-1/4}) = \Theta(1/\sqrt{(h^2 \log h)k \cdot \pi_{\max}})$. Now by Lemma 7.4.8, $\beta_{\varepsilon/2}(V, d_s) \leq O(h^2)$. Plugging all the values in (7.9) we get

  $$\lambda_k^*(G) \lesssim k\pi_{\max} \cdot h^6 \log h.$$

  If $(h^2 \log h)M^{\Psi'} \geq \Omega(1)$, then a similar argument as in the first case establishes the required upper bound on $\lambda_k^*(G)$.

This concludes the proof. □

7.5 Concluding Remarks

In this chapter, we have proved upper bounds on reweighted eigenvalues of special classes of graphs. These bounds lead to a new suite of spectral algorithms for finding small balanced separators, with improved bounds over past work [ST96, BLR10, KLPT11] when the graph is not constant-degree. In particular, for planar graphs we presented a fast spectral algorithm with optimal $O(\sqrt{n})$ separator size, based on adjusting the "kissing circle" embedding for Cheeger rounding. It is possible that similar techniques may be applied to bounded genus graphs as well.

While there is a fast and simple algorithm for computing eigenvalues and eigenvectors of the ordinary graph Laplacian, the current fastest algorithm in [LTW24] for computing reweighted eigenvalues, which runs in almost-linear time, is considerably more complicated than the classical counterpart. A fast and simple algorithm for computing reweighted eigenvalues is thus of interest here, for it would potentially make spectral partitioning using reweighted eigenvalues a practical alternative to existing algorithms.

One may wonder if there is an analogous upper bound on reweighted eigenvalues in the directed graph and hypergraph settings. While the claim that "topologically simple directed graphs and hypergraphs have small reweighted eigenvalues" is likely true, the concepts of planarity, genus, and graph minor do not generalize readily to these settings, so to make precise such a claim one needs to first find meaningful classes of generalized graphs to investigate.

Chapter 8

Tightening Reweighted Eigenvalues with Triangle Inequalities

In this chapter, we propose new semidefinite programming relaxations for expansion and multi-way expansion problems. They are derived from the reweighted eigenvalues formulation in Chapter 4 and Chapter 5 by adding ℓ_2^2 triangle inequality constraints to the programs.

For generalized graph expansion problems, we show that this achieves $O(\sqrt{\log n})$ integrality gap, providing a simple and unified approach to attain the best-known approximation. Solving the reweighted eigenvalues program with ℓ_2^2 triangle inequalities gives an embedding of the generalized graph satisfying the conditions of the ARV structure theorem [ARV09], and the traditional region growing argument [LR99, ARV09] is used to extract sparse cuts.

For k-way expansion problems, we adapt the approach of [BFK+14, LM14a, LM14b] using orthogonal separators to round the reweighted sum of first k eigenvalue programs plus ℓ_2^2 triangle inequalities and find $(1-\varepsilon)k$ sparse cuts of expansion $O_\varepsilon(k \log k \log\log k \sqrt{\log n})$ times the optimal. This works for undirected expansion problems, most generally multi-way hypergraph edge expansion.

The first half of the chapter about generalized graph expansion problems is based on the paper [LTW24]. In the paper, the main goal is to design almost linear-time algorithms to obtain low expansion cuts whose expansion is within $O(\sqrt{\log n})$ of the optimal. This is however out of the scope of this thesis, and we refer the interested reader to the paper.

8.1 Our Results

We present our results in this section. Results on expansion problems on generalized graphs are presented in Section 8.1.1, and results on k-way expansions are presented in Section 8.1.2.

8.1.1 Expansions and Reweighted Second Eigenvalues with Triangle Inequalities

We consider a new semidefinite program for generalized expansion quantities based on the reweighted eigenvalue formulation. To illustrate the idea and to better compare with previous work, we lay out our results for directed edge expansion (with arbitrary vertex weights). We first recall the definition of directed edge expansion.

Definition 8.1.1 (π-Weighted Directed Edge Expansion (from Section 2.3.2))**.** *Let $G = (V, E, w)$ be a directed graph with edge weights $w : E \to \mathbb{R}^+$, and let $\pi : V \to \mathbb{R}^+$ be a vertex measure. For $S \subseteq V$, let $\delta^+(S) := \{uv \in E : u \in S,\ v \notin S\}$ be the set of arcs going out of S. The π-weighted edge expansion of $S \subseteq V$ and of the graph G are defined as*

$$\vec{\phi}_\pi(S) := \frac{\min\{w(\delta^+(S)), w(\delta^+(S^c))\}}{\min\{\pi(S), \pi(S^c)\}} \quad \text{and} \quad \vec{\phi}_\pi(G) := \min_{\emptyset \neq S \subset V} \vec{\phi}_\pi(S).$$

As we have seen in Section 2.3.2, this is a general problem that encompasses various expansion problems studied in the literature. The directed edge expansion problem is when $\pi(u) = 1$ for all $u \in V$, and this is equivalent (up to a factor of $\Theta(n)$) to the directed sparsest cut of G, defined as

$$\min_{\emptyset \neq S \subset V} \frac{\min\{w(\delta^+(S)), w(\delta^+(S^c))\}}{|S| \cdot |S^c|}$$

and studied in [ACMM05]. The directed edge conductance in Definition 5.1.2 is when $\pi(u) = \deg_w(u) := \deg_w^+(u) + \deg_w^-(u)$, the weighted total degree of vertex u. Clearly, the corresponding problems in undirected graphs can be reduced to Definition 8.1.1 by bidirecting the edges in the undirected graph. Also, the undirected vertex expansion problem studied in [FHL08, LRV13][1] and the directed vertex expansion problem in Chapter 5 can

[1] To be precise, [FHL08] studies "minimum ratio vertex cuts", which is up to a constant equivalent to undirected vertex expansion as defined in Section 2.3.1 in the thesis. They use "vertex expansion" to refer to a different quantity incomparable to ours.

be reduced to Definition 8.1.1 through a standard reduction of splitting each vertex into two. Furthermore, the corresponding problems in undirected and directed hypergraphs can be reduced to Definition 8.1.1 through a reduction of replacing each hyperedge by a vertex as shown in [CS18] (see Definition 3.6.5).

We consider a new SDP relaxation for directed edge expansion in Definition 8.1.1. For undirected graphs, the SDP formulation in [ARV09] (c.f. (3.18)) can be understood as the spectral formulation for the second smallest Laplacian eigenvalue plus the ℓ_2^2 triangle inequalities [Tre16]. For directed graphs, our SDP formulation is the reweighted eigenvalue formulation in Definition 5.1.7 plus the ℓ_2^2 triangle inequalities.

Definition 8.1.2 (Reweighted Eigenvalue with Triangle Inequalities)**.** *Given an edge-weighted directed graph $G = (V, E, w)$, we say that $P : E \to \mathbb{R}_{\geq 0}$ is an Eulerian reweighting on G* [2] *if $\sum_{v:uv \in E} P(u, v) = \sum_{v:vu \in E} P(v, u)$ for all $u \in V$. We let $\mathcal{F}(G)$ be the set of all Eulerian reweightings on G that also satisfy the arc weight constraints that $P(u, v) \leq w(u, v)$ for all $uv \in E$. Given also vertex weights $\pi : V \to \mathbb{R}^+$, the $\lambda_2^{\triangle}(G)$ program for directed edge expansion is defined as*

$$\begin{aligned} \min_{f:V\to\mathbb{R}^n} \max_{P\in\mathcal{F}(G)} \quad & \sum_{u,v\in V} \frac{1}{2}\big(P(u,v)+P(v,u)\big)\cdot \|f(u)-f(v)\|^2 \\ \text{subject to} \quad & \sum_{u\in V} \pi(u)\cdot f(u) = \vec{0} \\ & \sum_{u\in V} \pi(u)\cdot \|f(u)\|^2 = 1 \\ & \|f(u)-f(v)\|^2 + \|f(v)-f(u')\|^2 \geq \|f(u)-f(u')\|^2 \qquad \forall u,v,u' \in V. \end{aligned} \tag{8.1}$$

Here we use the convention that $P(u, v) = 0$ if $uv \notin E$.

Using the semidefinite programming formulation for the second eigenvalue and the von Neumann min-max theorem, $\lambda_2^*(G)$ in Definition 5.1.7 can be rewritten as the form in Definition 8.1.2 without the triangle inequalities.

Just as the addition of triangle inequalities to the spectral formulation reduces the integrality gap of undirected edge expansion to $O(\sqrt{\log n})$ in [ARV09], we show the exact analog for directed edge expansion by using the spectral formulation for reweighted eigenvalues.

[2] In [LTW24], network flows is a main theme, and so P was referred to as a circulation. Here we revert to calling it an Eulerian reweighting for consistency with other chapters.

Theorem 8.1.3 ($O(\sqrt{\log n})$-Approximation for Directed Edge Expansion)**.** *For any edge-weighted directed graph $G = (V, E, w)$ with vertex weights $\pi : V \to \mathbb{R}^+$,*

$$\lambda_2^{\triangle}(G) \lesssim \vec{\phi}_\pi(G) \lesssim \sqrt{\log n} \cdot \lambda_2^{\triangle}(G).$$

Moreover, there is a randomized polynomial-time algorithm[3] *that, given G and π, computes a subset $S \subseteq V$ such that*

$$\vec{\phi}_\pi(S) \lesssim \sqrt{\log n} \cdot \vec{\phi}_\pi(G).$$

As we shall elaborate in Section 8.2, the proof is a simple adaptation of that in [ARV09]; see Section 3.6 for a review of their work.

Note that, since $\lambda_2^{\triangle}(G)$ (when setting $\pi(u)$ to be the total weighted degree of u) is a tightening of the $\vec{\lambda}_2^{e*}(G)$ program in Definition 5.1.7 and that $\vec{\lambda}_2^{e*}(G)$ can be used to certify expanders (in the sense that $\vec{\phi}(G) = \Theta(1)$ if and only if $\vec{\lambda}_2^{e*}(G) = \Theta(1)$), $\lambda_2^{\triangle}(G)$ can also be used to certify expanders. Note also that, as we shall see later, our rounding algorithm for Theorem 8.1.3 provides an alternative rounding algorithm to recover the Cheeger inequality in Theorem 5.1.8 for directed edge conductance.

Agarwal, Charikar, Makarychev and Makarychev [ACMM05] also derived an SDP-based rounding algorithm for directed edge expansion and proved $O(\sqrt{\log n})$ integrality gap; see Section 3.6.3 for a detailed review. Compared to their program $\mathsf{sdp}^{\triangle}(G)$ in (3.20), the $\lambda_2^{\triangle}(G)$ program that we introduce in Definition 8.1.2 is less constrained. We can see this by taking the linear programming dual of the inner maximization problem with respect to the $P(u, v)$ variables, so that

$$\max_{P \in \mathcal{F}(G)} \sum_{u,v \in V} \frac{1}{2}\big(P(u,v) + P(v,u)\big) \cdot \|f(u) - f(v)\|^2$$

becomes

$$\min_{r: V \to \mathbb{R}} \sum_{uv \in E} w(uv) \cdot \max\left\{0, \|f(u) - f(v)\|^2 - r(u) + r(v)\right\}.$$

Thus, we see that every feasible solution to (3.20) corresponds to a feasible solution to the $\lambda_2^{\triangle}(G)$ program with the same objective value by taking $r(u) = \|f(u) - f(0)\|^2$. (The reason we define $\lambda_2^{\triangle}$ in the min-max form is for ease of presentation of our analyses.)

Finally, we note that the same approach of adding ℓ_2^2 triangle inequalities to reweighted eigenvalues provides considerably simpler formulations and proofs for undirected vertex

[3] As remarked before, an almost linear-time algorithm is obtained in [LTW24].

expansion and hypergraph edge expansion than that in [FHL08] and in [LM14b] (see Section 3.6 for a review of their work), while having the same integrality gap $O(\sqrt{\log n})$; these results are presented in Section 8.4. Just as how the reweighted eigenvalue formulations in Chapter 4 and Chapter 5 provide a unifying framework to obtain Cheeger-type inequalities for generalized expansion quantities, we show in this study that in all these cases, adding ℓ_2^2 triangle inequality constraints to the reweighted eigenvalue formulations gives $O(\sqrt{\log n})$-approximation algorithms for estimating these quantities.

Remark 8.1.4 (Concurrent Work). *Concurrent to the publication of [LTW24], Chen, Orecchia, and Tani [COT23] designed an $O(\sqrt{\log n})$-approximation algorithm for "polymatroidal cut functions", which is a new subclass of submodular transformations that is more general than directed hypergraph expansion (with arbitrary vertex measure). Their result was in turn generalized by Chekuri and Louis [CL24] to "directed polymatroidal networks". Both results used an SDP relaxation similar to that in [ACMM05]; see Section 3.6.3. We note that in both papers, unlike in [LTW24], there is no fast (e.g. almost linear-time) algorithms with $O(\sqrt{\log n})$ approximation guarantees.*

In light of the negative evidence in Chapter 6, it remains to be seen if the reweighted eigenvalue formulation can be applied to models beyond generalized graphs, to obtain ARV-like approximation guarantees matching the aforementioned results.

8.1.2 Higher Expansions and Reweighted Higher Eigenvalues with Triangle Inequalities

Motivated by the positive results in the previous subsection, we then consider adding ℓ_2^2 triangle inequalities to the reweighted k-th eigenvalue formulation to obtain tighter relaxations of k-way expansion quantities. In previous chapters, higher-order Cheeger inequalities have been established for multi-way edge conductance (Theorem 3.1.6) and multi-way vertex expansion of undirected graphs (Theorem 4.1.10), as well as multi-way expansion of undirected hypergraphs ((3.11) and Theorem 5.11.1), but not for directed expansion quantities as explained in Section 5.10.1. Below, we formulate the problem and state the result for k-way hypergraph expansion, which is the most general symmetric expansion problem considered in this thesis. We begin by formally defining k-way hypergraph expansion.

Definition 8.1.5 (k-Way Hypergraph Edge Expansion). *Given an undirected hypergraph $H = (V, E, w)$ and a vertex measure $\pi : V \to \mathbb{R}^+$. For $1 \leq k \leq n$, the k-way hypergraph edge expansion of disjoint subsets $S_1, \ldots, S_k$ and of H are defined as*

$$(\phi_\pi)_k(S_1, \ldots, S_k) := \max_{i\in[k]} \frac{w(\delta(S_i))}{\pi(S_i)} \qquad \text{and} \qquad (\phi_\pi)_k(G) := \min_{S_1\sqcup\cdots S_k\subseteq V} (\phi_\pi)_k(S_1, \ldots, S_k),$$

where $\delta(S) := \{e \in E : e \cap S \neq \emptyset \wedge e \cap S^c \neq \emptyset\}$ *is the set of hyperedges cut by* S*.*

Theorem 8.1.6 (Approximating k-Way Hypergraph Edge Expansion)**.** *There is a randomized polynomial-time algorithm that, given an undirected hypergraph* $H = (V, E, w)$ *with vertex measure* $\pi : V \to \mathbb{R}^+$*, and also* $2 \leq k \leq n$ *and* $\varepsilon \in (0, 1/2]$*, returns* $\ell := \lfloor (1-\varepsilon)k \rfloor$ *disjoint subsets* $S_1, S_2, \cdots, S_\ell \subseteq V$*, such that*

$$(\phi_\pi)_\ell(S_1, S_2, \ldots, S_\ell) \leq O_\varepsilon\left(k \log k \log\log k \cdot \sqrt{\log n}\right) \cdot (\phi_\pi)_k(H).$$

We compare Theorem 8.1.6 with past work on the approximation of multi-way hypergraph expansion. We note that in the case where $\pi = \deg_w$ is the degree measure, for fixed ε, the multiplicative factor in the approximation guarantee of $\phi_{\lfloor (1-\varepsilon)k \rfloor}(H)$ in Theorem 8.1.6 is better than that in (3.11) by Chan, Louis, Tang, and Zhang [CLTZ18] roughly by a factor of $k^{1.5}$ but worse than that in Theorem 5.11.1 roughly by a factor of $\sqrt{k}$. However, there is no square-root loss in our approximation guarantee, and so it is significantly better than Cheeger-type results when the hypergraph has small k-way conductance.

By restricting the size of each hyperedge to be two, the result can be directly applied to k-way edge expansion of ordinary graphs. However, it is possible to obtain a better approximation factor of $O_\varepsilon(\sqrt{\log n \log k})$ for k-way edge expansion, which matches that in Theorem 3.6.13 by Louis and Makarychev [LM14a] for the related but different problem of sparsest k-partitioning $(\Phi_\pi)_k(G)$.

Theorem 8.1.7 (Approximating k-Way Edge Expansion)**.** *There is a randomized polynomial-time algorithm that, given an undirected graph* $G = (V, E, w)$ *with vertex measure* $\pi : V \to \mathbb{R}^+$*, and also* $2 \leq k \leq n$ *and* $\varepsilon \in (0, 1/2]$*, returns* $\ell := \lfloor (1-\varepsilon)k \rfloor$ *disjoint subsets* $S_1, S_2, \cdots, S_\ell \subseteq V$*, such that*

$$(\phi_\pi)_\ell(S_1, S_2, \ldots, S_\ell) \leq O_\varepsilon\left(\sqrt{\log n \log k}\right) \cdot (\phi_\pi)_k(G).$$

As remarked in Section 3.6.4, $(\phi_\pi)_k(G) \leq (\Phi_\pi)_k(G)$, and so Theorem 8.1.7 is a strengthening of Theorem 3.6.13. For k-way conductance by taking $\pi(u) = \deg_w(u)$, this result is worse by a factor of $\sqrt{\log n}$ than the higher-order Cheeger inequality Theorem 3.1.6 by Lee, Oveis Gharan, and Trevisan [LOT12], but without the square-root loss, and is likewise almost always better than (3.11) and Theorem 5.11.1 when specialized to graphs.

One significance of Theorem 8.1.7 is that it provides a better approximation algorithm for k-way expansion of low-rank hypergraphs via a reduction to graphs.

Corollary 8.1.8 (Approximating k-Way Hypergraph Edge Expansion for Low-Rank Hypergraphs)**.** *There is a randomized polynomial-time algorithm that, given an undirected hypergraph* $H = (V, E, w)$ *of rank* $r = r(H) := \max_{e \in E} |e|$ *with vertex measure* $\pi : V \to \mathbb{R}^+$*, and also* $2 \leq k \leq n$ *and* $\varepsilon \in (0, 1/2]$*, returns* $\ell := \lfloor (1-\varepsilon)k \rfloor$ *disjoint subsets* $S_1, S_2, \ldots, S_\ell \subseteq V$*, such that*

$$(\phi_\pi)_\ell(S_1, S_2, \ldots, S_\ell) \leq O_\varepsilon\left(\min(r\sqrt{\log k}, k \log k \log\log k) \cdot \sqrt{\log n}\right) \cdot (\phi_\pi)_k(H).$$

The reduction from small-set vertex expansion to hypergraph small-set expansion in Remark 3.6.20 applies to k-way expansions as well, and we obtain the following result about approximating k-way vertex expansions in graphs.

Corollary 8.1.9 (Approximating k-Way Vertex Expansion)**.** *There is a randomized polynomial-time algorithm, that, given an undirected graph* $G = (V, E)$ *of maximum degree* Δ *with vertex measure* $\pi : V \to \mathbb{R}^+$*, and also* $2 \leq k \leq n$ *and* $\varepsilon \in (0, 1/2]$*, returns* $\ell := \lfloor (1-\varepsilon)k \rfloor$ *disjoint subsets* $S_1, S_2, \ldots, S_\ell \subseteq V$*, such that*

$$\psi_\ell(S_1, \ldots, S_\ell) \leq O_\varepsilon\left(\min(\Delta\sqrt{\log k}, k \log k \log\log k) \cdot \sqrt{\log n}\right) \cdot \psi_k(G).$$

8.2 Our Techniques

Conceptually, the main contribution is that the reweighted eigenvalue formulation in Chapter 4 and Chapter 5 can be easily extended to obtain state-of-the-art approximation guarantees for generalized expansion and higher expansion quantities. On top of Theorem 8.1.3, we shall see in Section 8.4 that the technique of adding ℓ_2^2 triangle inequalities to reweighted eigenvalue formulations applies readily to directed hypergraph expansion, providing a unifying method to extend the results for undirected graphs to generalized graphs.

Technically, the proof of Theorem 8.1.3 begins similarly as [ARV09, FHL08, ACMM05], with the ℓ_2^2 structure theorem of [ARV09] in Theorem 3.6.2. After finding two vertex subsets that are well-separated in the average ℓ_2^2 embedding distance, a traditional region growing argument [LR99, ARV09] is used to produce a sparse cut.

As for the proof of Theorem 8.1.6 for k-way hypergraph expansion, our SDP relaxation is the $\sigma_k^*(H)$ program introduced previously in Section 5.11.2 plus ℓ_2^2 triangle inequalities. As discussed before, $\sigma_k^*(H)$ is itself introduced as a convex reformulation of $\lambda_k^*(H)$. Our SDP relaxation is defined formally as follows:

Definition 8.2.1 (Reweighted Sum of Eigenvalues with Triangle Inequalities)**.** *Given an undirected hypergraph $H = (V, E, w)$, vertex measure $\pi : V \to \mathbb{R}^+$, and $2 \le k \le n$, the $\lambda_k^{\Delta}(H)$ program is defined as*

$$
\begin{aligned}
\min_{\substack{f:V\to\mathbb{R}^n \\ g:V\to\mathbb{R}_{\ge 0}}} \quad & \frac{1}{k}\sum_{e\in E} w(e)g(e) \\
\text{subject to} \quad & g(e) \ge \|f(u) - f(v)\|^2 \quad \forall\{u,v\} \subseteq e, \forall e \in E \\
& \sum_{u\in V} \pi(u)\,\|f(u)\|^2 = k \\
& \sum_{u\in V} \pi(u) f(u) f(u)^T \preceq I_n \\
& \|f(u) - f(v)\|^2 + \|f(v) - f(u')\|^2 \ge \|f(u) - f(u')\|^2 \quad \forall u, v, u' \in V \cup \{0\}
\end{aligned}
$$

where we define $f(0) := \vec{0} \in \mathbb{R}^n$. Note that the scaling factor $\frac{1}{k}$ in the objective was absent from $\sigma_k^*(H)$. Note also that this definition makes sense for undirected graphs as well.

We apply the hypergraph orthogonal separators introduced by Louis and Makarychev [LM14b] (see Definition 3.6.15), generating many of them and postprocessing to obtain $\Theta(k)$ disjoint small expansion cuts. The proof flow follows that of Section 3.6.4.

As for Theorem 8.1.7 for k-way edge expansion, the only difference between $\mathsf{sdp}_k^{\Delta}(G)$ and $\lambda_k^{\Delta}(G)$ is that the spreading constraint

$$\sum_{v\in V} \pi(v)\langle f(u), f(v)\rangle = 1 \qquad \forall u \in V$$

is replaced by

$$\sum_{u\in V} \pi(u) f(u) f(u)^T \preceq I_n.$$

We show that the same key properties in Lemma 3.6.12 of the algorithm can also be inferred from the new constraints of the $\lambda_k^{\Delta}(G)$ program.

The reduction to obtain Corollary 8.1.8 from Theorem 8.1.7 is simply to construct an auxiliary graph by replacing each hyperedge by its star graph. It is possible to show that hypergraph expansion and ordinary edge expansion in the auxiliary graph are within a factor r from one another.

8.3 Rounding Reweighted Eigenvalues with Triangle Inequalities

In this section, we prove Theorem 8.1.3 that reweighted eigenvalue with triangle inequalities provides $O(\sqrt{\log n})$ approximation for directed edge expansion with general vertex weights. The proof is by applying a traditional threshold rounding on the two sets provided by the structure theorem of Arora, Rao, and Vazirani (see Theorem 3.6.2). The proof that $\lambda_2^{\triangle}(G)$ is indeed an SDP relaxation of directed edge expansion can be found in Appendix D.

Proposition 8.3.1 (Easy Direction)**.** *For any edge-capacitated directed graph $G = (V, E, w)$ with vertex weights $\pi : V \to \mathbb{R}^+$, it holds that $\lambda_2^{\triangle}(G) \le 2\vec{\phi}_\pi(G)$.*

We will use the structure theorem in [ARV09] for the proof of $\vec{\phi}_\pi(G) \lesssim \sqrt{\log n} \cdot \lambda_2^{\triangle}(G)$. Since we consider π-weighted directed edge expansion, we need the following weighted version of the structure theorem. The proof of the weighted version is a straightforward reduction to the unweighted version in Theorem 3.6.2 and is deferred to Appendix D (see [ACMM05, Algorithm 1] for a similar weighted structure theorem and reduction).

Lemma 8.3.2 (π-Weighted Structure Theorem)**.** *Let $G = (V, E, w)$ be a directed graph with vertex measure $\pi : V \to \mathbb{R}^+$ and $\pi(V) = 1$. Let $\{f(u)\}_{u \in V}$ be a set of embedding vectors satisfying ℓ_2^2 triangle inequalities and $\sum_{u,v \in V} \pi(u) \cdot \pi(v) \cdot \|f(u) - f(v)\|^2 = 1$. The embedding $\{f(u)\}_{u \in v}$ is said to be well-spread if $\pi(B(u, 1/\sqrt{10})) \le 1/10$ for all $u \in V$, where $B(u,t)$ denotes the set of points $v \in V$ such that $\|f(u) - f(v)\| < t$. If $\{f(u)\}_{u \in v}$ is well-spread, then there exists two subsets $L, R \subseteq V$ with $\pi(L), \pi(R) \ge \Omega(1)$ and*

$$d(L, R) := \min_{u \in L, v \in R} \|f(u) - f(v)\|^2 \gtrsim 1/\sqrt{\log n}.$$

Moreover, there is a randomized polynomial-time algorithm that finds such sets with high probability.

The proof of the hard direction below then follows a case analysis as in Theorem 3.6.3.

Theorem 8.3.3 (Hard Direction)**.** *Let $G = (V, E, w)$ be a directed graph with vertex measure $\pi : V \to \mathbb{R}^+$. There is a polynomial-time algorithm which, with high probability, finds a set $S \subseteq V$ with $\vec{\phi}_\pi(S) \lesssim \lambda_2^{\triangle}(G) \cdot \sqrt{\log n}$.*

Proof. Let $\{f(u)\}_{u \in V}$ be an optimal solution to the $\lambda_2^{\triangle}(G)$ program. By Fact 2.10.4, the two normalization constraints in $\lambda_2^{\triangle}(G)$ in Definition 8.1.2 imply

$$\sum_{u,v \in V} \pi(u)\pi(v) \cdot \|f(u) - f(v)\|^2 = 2. \tag{8.2}$$

There are two cases to consider: the "well-spread" case and the "large core" case. In either case, we assume without loss of generality that $\pi(V) = 1$. In the context of the Cheeger-type proof for Theorem 5.1.8, the ARV-type proof here produces directly a one-dimensional ℓ_1 solution $h : V \to \mathbb{R}$, so that the loss is only $O(\sqrt{\log n})$:

$$\begin{aligned} & \frac{\max_{P \in \mathcal{F}(G)} \frac{1}{2}(P(u,v) + P(v,u)) \cdot |h(u) - h(v)|}{\sum_{u,v \in V} \pi(u)\pi(v) \cdot |h(u) - h(v)|} \\ \lesssim \quad & \sqrt{\log n} \cdot \frac{\max_{P \in \mathcal{F}(G)} \frac{1}{2}(P(u,v) + P(v,u)) \cdot \|f(u) - f(v)\|^2}{\sum_{u,v \in V} \pi(u)\pi(v) \cdot \|f(u) - f(v)\|^2}. \end{aligned} \tag{8.3}$$

- Suppose that f is well-spread. Then, since

  $$\sum_{u,v \in V} \pi(u) \cdot \pi(v) \cdot \left\| \frac{1}{\sqrt{2}} f(u) - \frac{1}{\sqrt{2}} f(v) \right\|^2 = 1,$$

  we can apply Lemma 8.3.2 to obtain, with high probability, two subsets $L, R \subseteq V$ with $\pi(L), \pi(R) \geq \Omega(1)$ and

  $$d(L,R) := \min_{u \in L, v \in R} \|f(u) - f(v)\|^2 \gtrsim 1/\sqrt{\log n}$$

  in randomized polynomial time. We then define $h(u) := d(u, L)$. Then, since

  $$|d(u,L) - d(v,L)| \leq d(u,v) = \|f(u) - f(v)\|^2$$

  for any $u, v \in V$, in the numerator we have

  $$\max_{P \in \mathcal{F}(G)} \frac{1}{2}(P(u,v)+P(v,u)) \cdot |h(u)-h(v)| \lesssim \max_{P \in \mathcal{F}(G)} \frac{1}{2}(P(u,v)+P(v,u)) \cdot \|f(u) - f(v)\|^2.$$

  As for the denominator, we have

  $$\begin{aligned} \sum_{u,v \in V} \pi(u)\pi(v)|h(u) - h(v)| \quad & \geq \quad \sum_{u \in L, v \in R} \pi(u)\pi(v)|d(u,L) - d(v,L)| \\ & \gtrsim \quad \pi(L)\pi(R) \cdot \frac{1}{\sqrt{\log n}} \\ & \gtrsim \quad \frac{1}{\sqrt{\log n}} \cdot \sum_{u,v \in V} \pi(u)\pi(v) \|f(u) - f(v)\|^2. \end{aligned}$$

  This proves (8.3).

- Suppose that f is not well-spread, so that it has a large core $C := B(u_0, \frac{1}{\sqrt{10}})$ with $\pi(C) > 1/10$ for some $u_0 \in V$. That means $d(u, u_0) \leq 1/10$ for all $u \in C$. Define $h(u) := d(u, C)$. Using the same reason as in the first case, in the numerator we have

  $$\max_{P \in \mathcal{F}(G)} \frac{1}{2}(P(u,v)+P(v,u)) \cdot |h(u)-h(v)| \lesssim \max_{P \in \mathcal{F}(G)} \frac{1}{2}(P(u,v)+P(v,u)) \cdot \|f(u)-f(v)\|^2.$$

  As for the denominator, we use that, for any $u \in C$ and $v \in V$,

  $$|h(u) - h(v)| = |d(u,C) - d(v,C)| = d(v,C) \geq d(v,u_0) - \frac{1}{10},$$

  so that

  $$\begin{aligned} \sum_{u,v \in V} \pi(u)\pi(v)|h(u)-h(v)| &\geq \sum_{u \in C, v \in V} \pi(u)\pi(v)\left(d(v,u_0) - \frac{1}{10}\right) \\ &\geq \pi(C) \cdot \left(\sum_{v \in V} \pi(v) d(v,u_0) - \frac{1}{10}\right). \end{aligned}$$

  This can be further lower bounded using

  $$\begin{aligned} \sum_{v \in V} \pi(v) d(v,u_0) &= \frac{1}{2} \sum_{u,v \in V} \pi(u)\pi(v)(d(u,u_0) + d(v,u_0)) \\ &\geq \frac{1}{2} \sum_{u,v \in V} \pi(u)\pi(v) \|f(u)-f(v)\|^2 = 1, \end{aligned}$$

  where the last equality is (8.2). Therefore, in fact we have

  $$\sum_{u,v \in V} \pi(u)\pi(v)|h(u)-h(v)| \geq \pi(C) \cdot \left(1 - \frac{1}{10}\right) \gtrsim \sum_{u,v \in V} \pi(u)\pi(v) \|f(u)-f(v)\|^2.$$

Once (8.3) is established, the rest of the proof is the same as the threshold rounding step in Section 5.7.3. By scaling we can assume that $\sum_{u,v \in V} \pi(u)\pi(v) \cdot |h(u)-h(v)| = 2$, and Fact 2.10.4 implies that

$$\sum_{v \in V} \pi(v) h(v) = 0 \qquad \text{and} \qquad \sum_{v \in V} \pi(v)|h(v)| = 1,$$

by appropriately shifting h by a constant. Then, h is a feasible solution to the π-weighted ℓ_1 dual program (similar to Lemma 5.7.6 but with vertex weight $\pi(u)$ instead of $\deg_w(u)$ in the constraints). It is straightforward to verify that the threshold rounding proof works for vertex weights other than just $\deg_w(u)$. □

Theorem 8.1.3 follows immediately from Theorem 8.3.3 and Proposition 8.3.1.

8.4 Generalization to Directed Hypergraphs

In this section, we show that adding ℓ_2^2 triangle inequality constraints to the reweighted eigenvalue SDP for directed hypergraph expansion, we obtain a tighter relaxation which has an integrality gap of $O(\sqrt{\log n})$. This encompasses all the expansion problems we have studied in this thesis, including undirected and directed vertex expansion and undirected hypergraph expansion. For definition of directed hypergraph and its expansion, refer to Section 2.2 and Section 2.3.3.

We again derive our SDP by adding ℓ_2^2 triangle inequalities to the reweighted eigenvalue program for directed hypergraphs, which we have introduced in Definition 6.3.11.

Definition 8.4.1 (Directed Hypergraph Reweighted Eigenvalue with Triangle Inequalities)**.** *Given a directed hypergraph $H = (V, E, w)$ over vertex measure $\pi : V \to \mathbb{R}^+$. Let*

$$\mathcal{F}(H) := \Bigg\{ P : V \times V \to \mathbb{R}_{\geq 0} \Bigg| \exists \{P_e : e^- \times e^+ \to \mathbb{R}_{\geq 0}\}_{e \in E} \ s.t. \quad P(u,v) = \sum_{e:(u,v)\in(e^-,e^+)} P_e(u,v), \quad \sum_{(u,v)\in(e^-,e^+)} P_e(u,v) \leq w(e) \quad \forall e \in E, \quad \sum_{v' \in V} P(v', u) = \sum_{v \in V} P(u,v) \quad \forall u \in V \Bigg\}$$

be the set of feasible reweightings on H. The $\lambda_2^{\Delta}(H)$ program for directed hypergraph expansion is defined as

$$\begin{aligned} \min_{f:V\to\mathbb{R}^n} \max_{P\in\mathcal{F}(H)} \quad & \sum_{u,v\in V} \frac{1}{2}(P(u,v) + P(v,u)) \left\| f(u) - f(v) \right\|^2 \\ \text{subject to} \quad & \sum_{u\in V} \pi(u) \cdot f(u) = \vec{0} \\ & \sum_{u\in V} \pi(u) \cdot \left\| f(u) \right\|^2 = 1 \\ & \left\| f(u) - f(v) \right\|^2 + \left\| f(v) - f(u') \right\|^2 \geq \left\| f(u) - f(u') \right\|^2 \quad \forall u, v, u' \in V. \end{aligned}$$

The intuition for defining feasible reweightings on directed hypergraphs this way is that they correspond to Eulerian reweightings of an underlying "clique graph" K_H of the directed hypergraph H, where for each directed hyperedge (e^-, e^+), we add an arc uv from every $u \in e^-$ to $v \in e^+$ with weight $w(e)$. The definition $\lambda_2^{\triangle}(H)$ is a natural one for various reasons. First, it can be shown that $\lambda_2^{\triangle}(H)$ is a relaxation of $\vec{\phi}_\pi(H)$. Second, when H is an undirected hypergraph and π is the total weighted degree, i.e. $e^- = e^+$, $\forall e \in E$, and $\pi(u) = \deg_w(u)$, then $\lambda_2^{\triangle}(H)$ is equivalent to the reweighted eigenvalue program for undirected hypergraphs as defined in Definition 5.1.9 but with ℓ_2^2 triangle inequalities. Third, just as our program for directed graphs is a relaxation of the SDP in [ACMM05], this program is a relaxation of the SDP in [CS18] (refer to Section 2 of their paper for details).

We show that our main result for directed edge expansion extends readily to the most general graph-like setting of directed hypergraphs.

Theorem 8.4.2 (Hypergraph Integrality Gap)**.** *Let $H = (V, E, w)$ be an edge-weighted directed hypergraph with vertex measure $\pi : V \to \mathbb{R}^+$. Then we have*

$$\lambda_2^{\triangle}(H) \lesssim \vec{\phi}_\pi(H) \lesssim \sqrt{\log n} \cdot \lambda_2^{\triangle}(H)$$

Proof outline. We first get the easy direction out of the way. The simplest proof is to apply Proposition 6.3.15 directly, noticing that the two-point embedding f in the proof satisfies ℓ_2^2 triangle inequalities. Now we focus on the hard direction.

One may attempt to prove the hard direction is by relating hypergraph expansion of H to the edge expansion of an ordinary derived graph G_H as in [CS18], which we have introduced in Definition 3.6.5. However, if we perform a black-box reduction from $\vec{\phi}_\pi(H)$ to $\vec{\phi}_{\pi'}(G_H)$, the approximation guarantee using this approach degrades to $O(\sqrt{\log(n+m)})$, which is worse when $m = \omega(\text{poly}(n))$.

To obtain the optimal approximation guarantee, we follow the proof of Theorem 8.3.3. The exact same argument using the ARV structure theorem (Theorem 3.6.2) produces an $h : V \to \mathbb{R}$ such that

$$\begin{aligned} & \frac{\max_{P \in \mathcal{F}(H)} \frac{1}{2}(P(u,v) + P(v,u)) \cdot |h(u) - h(v)|}{\sum_{u,v \in V} \pi(u)\pi(v) \cdot |h(u) - h(v)|} \\ \lesssim \quad & \sqrt{\log n} \cdot \frac{\max_{P \in \mathcal{F}(H)} \frac{1}{2}(P(u,v) + P(v,u)) \cdot \|f(u) - f(v)\|^2}{\sum_{u,v \in V} \pi(u)\pi(v) \cdot \|f(u) - f(v)\|^2}. \end{aligned}$$

Again, shifting gives a feasible solution to the π-weighted ℓ_1 dual program in Definition 6.3.3, and the threshold rounding proof in Proposition 6.3.8 works for arbitrary vertex weights. □

8.5 Many Hypergraph Sparse Cuts from Orthogonal Separators

In this section, we prove Theorem 8.1.6. Our approach is to use the SDP relaxation in Definition 8.2.1 and design a rounding algorithm using hypergraph orthogonal separators. We shall prove the following integrality gap about the $\lambda_k^{\Delta}(H)$ program.

Theorem 8.5.1 (Integrality Gap of $\lambda_k^{\Delta}(H)$ Program)**.** *Let $H = (V, E, w)$ be an undirected hypergraph with vertex measure $\pi : V \to \mathbb{R}^+$. Given also $2 \le k \le n$ and $\varepsilon \in (0, 1/2]$. Let $\ell := \lfloor (1-\varepsilon)k \rfloor$. Then,*

$$\phi_\ell(H) \lesssim_\varepsilon k \log k \log\log k \cdot \sqrt{\log n} \cdot \lambda_k^{\Delta}(H) \qquad \text{and} \qquad \lambda_k^{\Delta}(H) \lesssim \phi_k(H).$$

Moreover, there is a randomized polynomial-time algorithm that, given a feasible solution to the $\lambda_k^{\Delta}(H)$ program with objective value OBJ, returns ℓ disjoint vertex subsets $S_1, \ldots, S_\ell$ such that

$$\max_{i \in [\ell]} \phi(S_i) \lesssim_\varepsilon k \log k \log\log k \cdot \sqrt{\log n} \cdot OBJ.$$

Clearly, Theorem 8.1.6 follows from Theorem 8.5.1. We present the proof of the easy direction in Appendix D.

Proposition 8.5.2 (Easy Direction)**.** *For any undirected hypergraph $H = (V, E, \pi)$ with vertex measure $\pi : V \to \mathbb{R}^+$, and for any $2 \le k \le n$, we have*

$$\lambda_k^{\Delta}(H) \le \phi_k(H).$$

8.5.1 Hard Direction

Before giving the proof of the hard direction, we provide a technical overview here. Our rounding algorithm is similar to that of [LM14a, LM14b], with several differences in the analysis. The reader is strongly encouraged to read Section 3.6.4 before proceeding.

First, our vector solution to the $\lambda_k^{\Delta}(H)$ satisfies a different spreading constraint that

$$\sum_{u \in V} \pi(u) f(u) f(u)^T \preceq I_n,$$

and we need to show that the orthogonal separators generated are unlikely to have a large mass, where the mass of $S \subseteq V$ is again defined in relation to the vector solution f as

$$\mu(S) := \sum_{u \in S} \pi(u) \|f(u)\|^2 .$$

(Note that $\mu(V) = k$.) Second, we need to define a new ν potential for analyzing the expected total weight of hyperedges cut by the ℓ returned sets. Unlike in the k-way edge expansion and small-set hypergraph expansion settings, here each hyperedge can be cut by many sets, and a careful analysis is needed to limit the loss.

The following constraint, a direct consequence of applying ℓ_2^2 triangle inequality to vectors $f(u), f(v)$, and $f(0) = \vec{0}$, will be useful:

Fact 8.5.3. *Let f be a solution to the $\lambda_k^{\Delta}(H)$ program. Then, $0 \leq \langle f(u), f(v) \rangle \leq \|f(u)\|^2$ for any $u, v \in V$.*

We present our rounding algorithm in Algorithm 4, which is based on Algorithm 2. We prove its key properties in Lemma 8.5.4 and finally show how these properties imply Theorem 8.5.1.

Algorithm 4 Many Sparse Cuts using Hypergraph Orthogonal Separators

---

Input: Hypergraph $H = (V, E, w)$, solution $f : V \to \mathbb{R}^n$ to $\lambda_k^{\Delta}(H)$, parameters k, ε
Output: $\ell := \lfloor (1 - \varepsilon)k \rfloor$ disjoint subsets $S_1, \ldots, S_\ell \subseteq V$

1: Compute the normalization $\bar{f}$ from f per Proposition 3.6.11
2: Let $s := (12k/\varepsilon)$. Sample $T = \Theta(\frac{1}{\alpha} \log \frac{1}{\varepsilon})$ independent hypergraph s-orthogonal separators $S_1, \ldots, S_T$ for vectors $\bar{f}(u)$ and some choice of $\alpha \geq \max(1/s, 1/n) \geq \varepsilon/(12k)$, with separation threshold $\beta = 1 - \frac{\varepsilon}{12}$ and distortion $D = O_\beta(\sqrt{\log n} \cdot s \log s \log \log s)$
3: For each $i \in [T]$, define $S_i' := S_i$ if $\mu(S_i) \leq 1 + \varepsilon/2$ and $S_i' := \emptyset$ otherwise
4: For each $i \in [T]$, let $S_i'' := S_i' \setminus (\cup_{j<i} S_j')$
5: For each $i \in [T]$, let $P_i := \{u \in S_i'' : \|f(u)\|^2 \geq t_i\}$, where t_i is chosen to minimize $\phi(P_i)$
6: **return** the ℓ sets from P_i with the smallest edge conductance $\phi(P_i)$

Lemma 8.5.4 (Key Properties of Algorithm 4)**.** *Let $H = (V, E, w)$ be an undirected hypergraph with vertex measure $\pi : V \to \mathbb{R}^+$. Suppose that H is of rank r. Let $2 \leq k \leq n$ and $\varepsilon \in (0, 1/2]$. Then, the following guarantees about Algorithm 4 hold.*

(a) For every vertex $u \in V$ where $\|f(u)\| \neq 0$ and $i \in \{1, \ldots, T\}$, we have $\Pr[u \in S_i'] \geq \frac{\alpha}{2}$.

(b) All sets S_i'' are disjoint, and for a suitable choice of T,

$$\mathbb{E}\left[\mu(\cup S_i'')\right] \geq k\left(1 - \frac{\varepsilon}{8}\right)$$

(c) Let $f : V \to \mathbb{R}^n$ be an optimal solution to $\lambda_k^{\Delta}(H)$. For a set $S \subseteq V$, define

$$\nu(S) := \sum_{e \in E} w(e)\left[\mathbb{1}[e \in \delta(S)] \max_{u \in e \cap S} \|f(u)\|^2 + \mathbb{1}[e \subseteq S] \max_{u,v \in e} \left|\|f(u)\|^2 - \|f(v)\|^2\right|\right].$$

Then,

$$\mathbb{E}\left[\sum_{i \in [T]} \nu(S_i'')\right] \lesssim (1 + T \cdot \alpha D) \cdot k\lambda_k^{\Delta}(H).$$

Proof. To prove (a), note that we apply hypergraph orthogonal separators to the normalized vectors $\bar{f}(u)$. Then, if $\|f(u)\| \neq 0$ the normalization satisfies $\left\|\bar{f}(u)\right\| = 1$, and so $\Pr[u \in S_i] = \alpha \left\|\bar{f}(u)\right\|^2 = \alpha$ by Proposition 3.6.11. Thus, it suffices to prove that

$$\Pr[\mu(S_i) \leq 1 + \varepsilon/2 \,|\, u \in S_i] \geq 1/2$$

for all $i \in [T]$. Let $B := \{v \in V : \langle \bar{f}(u), \bar{f}(v) \rangle \leq \beta\}$. Whenever $v \in B$, we have

$$\Pr[v \in S_i \,|\, u \in S_i] = \frac{\Pr[u, v \in S_i]}{\Pr[u \in S_i]} \leq \frac{\varepsilon}{12k}$$

by the second property of orthogonal separators. Since $\mu(B) \leq \mu(V) = k$, by linearity of expectation and Markov's inequality it follows that

$$\Pr\left[\mu(S_i \cap B) \leq \frac{\varepsilon}{6} \,|\, u \in S_i\right] \geq \frac{1}{2}.$$

Next, use the spreading constraint to show that most of the vectors are in B. Indeed,

$$\begin{aligned}
\mu(V \setminus B) &= \sum_{v \notin B} \pi(v) \left\| f(v) \right\|^2 \\
&\le \sum_{v \notin B} \pi(v) \max(\left\| f(u) \right\|^2, \left\| f(v) \right\|^2) \\
&= \sum_{v \notin B} \pi(v) \frac{\langle f(u), f(v) \rangle}{\langle \bar{f}(u), \bar{f}(v) \rangle} \quad \text{(by Proposition 3.6.11)} \\
&\le \frac{1}{\beta} \sum_{v \notin B} \pi(v) \langle f(u), f(v) \rangle \quad (\text{since } \langle f(u), f(v) \rangle \ge 0 \text{ from Fact 8.5.3}) \\
&\le \frac{1}{\beta^2} \sum_{v \notin B} \pi(v) \frac{\langle f(u), f(v) \rangle^2}{\left\| f(u) \right\|^2} \quad \left(\text{since } \beta \le \langle \bar{f}(u), \bar{f}(v) \rangle = \frac{\langle f(u), f(v) \rangle}{\max(\left\| f(u) \right\|^2, \left\| f(v) \right\|^2)} \right) \\
&= \frac{1}{\beta^2} \cdot \frac{f(u)^T}{\left\| f(u) \right\|} \left(\sum_{v \notin B} \pi(v) f(v) f(v)^T \right) \frac{f(u)}{\left\| f(u) \right\|} \\
&\le \frac{1}{\beta^2} \cdot \frac{f(u)^T}{\left\| f(u) \right\|} \cdot I \cdot \frac{f(u)}{\left\| f(u) \right\|} \\
&= \frac{1}{\beta^2} \le 1 + \frac{\varepsilon}{3}.
\end{aligned}$$

which implies that $\mu(V \setminus B) \le 1 + \varepsilon/3$. To summarize, with $S_i = (S_i \cap B) \cup (S_i \setminus B)$,

$$\Pr\left[\mu(S_i) \le 1 + \frac{\varepsilon}{2} \mid u \in S_i\right] \ge \Pr\left[\mu(S_i \cap B) \le \frac{\varepsilon}{6} \mid u \in S_i\right] \ge \frac{1}{2}.$$

This completes the proof of (a). (b) immediately follows since for any vertex $u \in V$ with nonzero mass,

$$\Pr\left[u \in \cup_{i \in [T]} S_i'\right] \ge 1 - \left(1 - \frac{\alpha}{2}\right)^T \ge 1 - \exp\left(-\Theta\left(\log \frac{1}{\varepsilon}\right)\right) \ge 1 - \frac{\varepsilon}{8}$$

for suitably chosen T, so the expected total mass in the union of the S_i's is at least

$$\mu(V) \cdot \left(1 - \frac{\varepsilon}{8}\right) = k \cdot \left(1 - \frac{\varepsilon}{8}\right).$$

It remains to prove (c). To bound the expectation

$$\mathbb{E}\left[\sum_{i \in [T]} \sum_{e \in E} w(e) \left(\mathbb{1}[e \in \delta(S_i'')] \max_{u \in e \cap S_i''} \left\| f(u) \right\|^2 + \mathbb{1}[e \subseteq S_i''] \max_{u,v \in e} \left| \left\| f(u) \right\|^2 - \left\| f(v) \right\|^2 \right| \right)\right],$$

we break it into two parts. For any hyperedge $e \in E$, there is at most one $i \in [T]$ such that $e \subseteq S_i''$, and so

$$\begin{aligned} & \mathbb{E}\left[\sum_{i\in[T]}\sum_{e\in E} w(e)\mathbb{1}[e\subseteq S_i''] \max_{u,v\in e}\left|\|f(u)\|^2 - \|f(v)\|^2\right|\right] \\ \leq & \sum_{e\in E} w(e) \max_{u,v\in e}\left|\|f(u)\|^2 - \|f(v)\|^2\right|, \end{aligned}$$

which is at most $k\lambda_k^{\Delta}(H)$ by ℓ_2^2 triangle inequality. The remaining part to bound is

$$\mathbb{E}\left[\sum_{i\in[T]}\sum_{e\in E} w(e)\mathbb{1}[e\in\delta(S_i'')] \max_{u\in e\cap S_i''}\|f(u)\|^2\right].$$

For $e \in \delta(S_i'')$ to happen, it is necessary that either $e \in \delta(S_i)$, or $i \in [T]$ satisfies (1) i is the first index such that $\cup_{j\leq i} S_j'' \supseteq e$ and (2) $e \in \delta(S_j)$ for some $j < i$. Since the S_i's are i.i.d. generated, we can upper bound the above expectation by

$$\sum_{e\in E} w(e)\left[2T\cdot \Pr_S[e\in\delta(S)]\cdot \max_{u\in e}\|f(u)\|^2\right], \qquad (*)$$

where S is a randomly generated hypergraph orthogonal separator with the parameters specified in Algorithm 4. The following technical observation allows us to bound this.

Proposition 8.5.5. *Given $f : V \to \mathbb{R}^n$ from and its normalization $\bar{f} : V \to \mathbb{R}^n$ in accordance to Proposition 3.6.11, we have, for any $e \subseteq V$,*

$$\max_{u\in e}\|f(u)\|^2 \cdot \max_{u,v\in e}\left\|\bar{f}(u)-\bar{f}(v)\right\|^2 \leq 4\max_{u,v\in e}\|f(u)-f(v)\|^2.$$

Proof. We consider two cases.

- Case 1: $\max_{u\in e}\|f(u)\|^2 \geq 2\min_{u\in e}\|f(u)\|^2$. Then, $\max_{u,v\in e}\left\|\bar{f}(u)-\bar{f}(v)\right\|^2 \leq 2$ and

  $$\max_{u,v\in e}\|f(u)-f(v)\|^2 \geq \max_{u\in e}\|f(u)\|^2 - \min_{v\in e}\|f(v)\|^2 \geq \frac{1}{2}\max_{u\in e}\|f(u)\|^2.$$

  Rearranging, we get the desired inequality.

- Case 2: $\max_{u\in e}\|f(u)\|^2 < 2\min_{u\in e}\|f(u)\|^2$. Then, for any $u, v \in e$,

$$\left\|\bar{f}(u)-\bar{f}(v)\right\|^2 \le 2\frac{\|f(u)-f(v)\|^2}{\max(\|f(u)\|^2,\|f(v)\|^2)} \le 4\frac{\|f(u)-f(v)\|^2}{\max_{u'\in e}\|f(u')\|^2},$$

  where the first inequality is by Proposition 3.6.11. Taking maximum over all pairs (u, v) yields the desired inequality.

This completes the proof. □

By Proposition 8.5.5 and the third property of hypergraph orthogonal separators,

$$\Pr_S[e\in\delta(S)] \lesssim \alpha D\cdot\max_{u,v\in e}\left\|\bar{f}(u)-\bar{f}(v)\right\|^2 \lesssim \alpha D\cdot\frac{\max_{u,v\in e}\|f(u)-f(v)\|^2}{\max_{u\in e}\|f(u)\|^2},$$

and so $(*)$ may be further bounded above by

$$T\cdot\alpha D\cdot\sum_{e\in E} w(e)\max_{u,v\in e}\|f(u)-f(v)\|^2 \le T\cdot\alpha D\cdot k\lambda_k^{\Delta}(H).$$

The proof of (c) is complete. □

Proof of Theorem 8.5.1. We run Algorithm 4. Since Lemma 8.5.4(b) asserts that the expected total mass of S_i'' is at least $k(1-\varepsilon/8)$, so with probability at least $1/2$ the total mass $M := \sum_{i\in[T]}\mu(S_i'')$ is at least $k(1-\varepsilon/4)$. Assume that this is the case, and let

$$Z := \frac{\sum_{i\in[T]}\nu(S_i'')}{\sum_{i\in[T]}\mu(S_i'')} = \frac{1}{M}\sum_{i\in[T]}\nu(S_i'').$$

The correct way to view $\nu(S_i'')$ is the expected weight of the hyperedge boundary under a certain threshold rounding scheme, and so the following definition

$$\mathcal{I} := \left\{ i\in[T] : S_i''\neq\emptyset \text{ and } \frac{\nu(S_i'')}{\mu(S_i'')} \le \frac{4Z}{\varepsilon}\right\}$$

chooses precisely those sets that will round to a small hypergraph expansion set. Let $S_{\mathcal{I}}'' := \cup_{i\in\mathcal{I}} S_i''$. The total mass of sets S_i'' for i *outside* $\mathcal{I}$ is upper bounded by

$$\sum_{i\notin\mathcal{I}}\mu(S_i'') < \frac{\varepsilon}{4Z}\sum_{i\notin\mathcal{I}}\nu(S_i'') \le \frac{\varepsilon}{4Z}\cdot MZ,$$

and so

$$\mu(S''_{\mathcal{I}}) \geq M - \frac{M\varepsilon}{4} = M\left(1 - \frac{\varepsilon}{4}\right) \geq k\left(1 - \frac{\varepsilon}{4}\right)^2.$$

Since $\mu(S''_i) \leq (1+\varepsilon/2)$ for all $i \in [T]$, the set $\mathcal{I}$ has at least $\ell = \lfloor (1-\varepsilon)k \rfloor$ elements.

Now, for the threshold rounding, for any $i \in [T]$, let $M_i := \max\{\|f(u)\|^2 : u \in S''_i\}$ and define threshold sets

$$U_t := \{u \in S''_i : \|f(u)\|^2 > t\}$$

for $t \in [0, M_i]$. The "average" denominator is

$$\int_0^{M_i} \mathrm{vol}(U_t)\, dt = \sum_{u \in S''_i} \pi(u)\, \|f(u)\|^2 = \mu(S''_i).$$

For a hyperedge e to be cut by U_t, there are several cases to consider:

- if $e \cap S''_i = \emptyset$, then $e \notin \delta(U_t)$ for any $t \in [0, M_i]$.
- If e is cut by S''_i, then $e \in \delta(U_t)$ if and only if $t < \max_{u \in e \cap S''_i} \|f(u)\|^2$.
- If $e \subseteq S''_i$, then $e \in \delta(U_t)$ if and only if

$$\min_{u \in e} \|f(u)\|^2 \leq t < \max_{v \in e} \|f(v)\|^2.$$

Therefore, the "average" numerator is

$$\begin{aligned} & \int_0^{M_i} w(\delta(U_t))\, dt \\ = \quad & \sum_{e \in E} w(e) \int_0^{M_i} \mathbb{1}\left[e \in \delta(U_t)\right]\, dt \\ = \quad & \sum_{e \in E} w(e) \left[\mathbb{1}[e \in \delta(S''_i)] \max_{u \in e \cap S''_i} \|f(u)\|^2 + \mathbb{1}[e \subseteq S''_i] \max_{u,v \in e} \left| \|f(u)\|^2 - \|f(v)\|^2 \right| \right] = \nu(S''_i). \end{aligned}$$

Therefore, for $i \in \mathcal{I}$, P_i in step 5 of Algorithm 4 satisfies

$$\phi(P_i) \leq \frac{\nu(S''_i)}{\mu(S''_i)} \leq \frac{4Z}{\varepsilon}.$$

Since we have chosen

$$T = \Theta\left(\frac{1}{\alpha} \log \frac{1}{\varepsilon}\right) \qquad \text{and} \qquad D = O\left(\beta^{-1} \cdot s \log s \log\log s \cdot \sqrt{\log n}\right)$$

with $s = (12k/\varepsilon)$, we have

$$1 + T \cdot \alpha D = O_\varepsilon \left(k \log k \log\log k \cdot \sqrt{\log n} \right),$$

and so

$$\mathbb{E}[Z] \le O_\varepsilon \left(k \log k \log\log k \cdot \sqrt{\log n} \right) \cdot \lambda_k^\Delta(H)$$

by Lemma 8.5.4(c). Therefore, with probability at least 3/4, picking the ℓ sets $P_{i(1)}, \ldots, P_{i(\ell)}$ with the smallest conductance we have

$$\begin{aligned} \phi_\ell(P_{i(1)}, \ldots, P_{i(\ell)}) &\le O_\varepsilon(\varepsilon^{-1} k \log k \log\log k \cdot \sqrt{\log n} \cdot \lambda_k^\Delta(H)) \\ &= O_\varepsilon(k \log k \log\log k \cdot \sqrt{\log n} \cdot \lambda_k^\Delta(H)), \end{aligned}$$

as required. Finally, a careful but straightforward inspection of Algorithm 4 verifies that the algorithm indeed runs in randomized polynomial time. □

8.6 Multi-Way Edge Expansion and Vertex Expansion using Orthogonal Separators

8.6.1 Multi-Way Edge Expansion

We prove Theorem 8.1.7 in this subsection.

Proof of Theorem 8.1.7. Our algorithm is a straightforward modification of Algorithm 2, the only change being that the input embedding $f : V \to \mathbb{R}^n$ is a solution to the $\lambda_k^\Delta(G)$ program in Definition 8.2.1 instead of to the $\mathsf{sdp}_k^\Delta(G)$ program in [LM14a]. We would like to show that the properties of Theorem 3.6.10 hold for our algorithm as well. Indeed, only the proof of (a) is affected by the change, as we need to show that

$$\mu(V \setminus B) \le 1 + \varepsilon/3,$$

where $B := \{v \in V : \langle \bar{f}(u), \bar{f}(v) \rangle \le \beta\}$, using the new spreading constraint. The proof is exactly the same as in Lemma 8.5.4(a) for hypergraphs.

Following the proof of Theorem 3.6.13, we obtain the following hard direction that

$$\phi_{\lfloor (1-\varepsilon)k \rfloor}(G) \le O_\varepsilon(\sqrt{\log n \log k} \cdot \lambda_k^\Delta(G)).$$

The easy direction that $\lambda_k^\Delta(G) \le \phi_k(G)$ is a special case of Proposition 8.5.2. □

8.6.2 Implication about Low-Rank Hypergraphs

We prove Corollary 8.1.8 in this subsection.

Proof of Corollary 8.1.8. Given a hypergraph $H = (V, E, w)$ with vertex measure $\pi : V \to \mathbb{R}^+$, consider the following graph $G = (V', E', w')$ with $V' = V$ and equipped with the same vertex measure. For each hyperedge $e \in E$ with weight $w(e)$, add a star graph to G, i.e. fix a vertex $u_0 \in e$ and add weight $w(e)$ to edge (u_0, v) for all $v \in e \setminus \{u_0\}$. From Theorem 8.1.7, there is a randomized polynomial-time algorithm that, given G, $2 \le k \le n$ and $\varepsilon \in (0, 1/2]$, produces $\ell = \lfloor (1-\varepsilon)k \rfloor$ disjoint subsets $S_1, \ldots, S_\ell \subseteq V$, such that

$$\phi_\pi^G(S_1, \ldots, S_\ell) \le O_\varepsilon(\sqrt{\log n \log k}) \cdot (\phi_\pi)_k(G).$$

(We use ϕ_π^G to mean expansion in G.) Let $S \subseteq V$. By our construction, $\pi(S) = \pi'(S)$. If e is cut by S, then at least one and at most $|e| - 1$ edges in the star graph in G gets cut by S, whereas if e is not cut by S then none of the edges in the star graph in G gets cut by S. Therefore,

$$\frac{\phi_\pi^G(S)}{r(H)} \le \phi_\pi^H(S) \le \phi_\pi^G(S),$$

and this shows that the same subsets $S_1, \ldots, S_\ell$ generated above satisfies

$$\phi_\pi^H(S_1, \ldots, S_\ell) \le O_\varepsilon(r\sqrt{\log n \log k}) \cdot (\phi_\pi)_k(H).$$

By choosing the better solution between the above and that from Theorem 8.1.6, the result is proven. □

8.6.3 Multi-Way Vertex Expansion

We prove Corollary 8.1.9 in this subsection.

Proof of Corollary 8.1.9. Given $G = (V, E)$ and distribution $\pi : V \to \mathbb{R}^+$. Suppose the maximum degree of G is Δ, and let

$$C := O_\varepsilon\left(\min(\Delta\sqrt{\log k}, k \log k \log\log k) \cdot \sqrt{\log n}\right)$$

for an appropriate constant in the big-O. We require also that $C \ge 2$.

If $\psi_k(G) \ge 1/C$, then by the definition of multi-way vertex expansion we can return any ℓ disjoint subsets. Otherwise, use Proposition 2.3.6 followed by Proposition 2.3.7 to construct a hypergraph $H = (V'', E'', w'')$ with distribution $\pi'' : V'' \to \mathbb{R}^+$ on the vertices of H. Run the algorithm for Corollary 8.1.8 to obtain $\ell = \lfloor (1-\varepsilon)k \rfloor$ disjoint subsets $S_1'', \ldots, S_\ell'' \subseteq V''$, such that

$$(\phi_{\pi''})_\ell(S_1'', S_2'', \ldots, S_\ell'') \leq \frac{C}{6} \cdot (\phi_{\pi''})_k(H).$$

By the assumption that $\psi_k(G) \leq 1/C \leq 1/2$, Proposition 2.3.6 and Proposition 2.3.7 imply

$$\psi_k(G) \geq \frac{1-\frac{1}{2}}{1+\frac{1}{2}} \cdot (\phi_{\pi''})_k(H) = \frac{1}{3}(\phi_{\pi''})_k(H).$$

Finally, since

$$(\phi_{\pi''})(S_i'') \leq \frac{C}{6} \cdot 3\psi_k(G) \leq \frac{1}{2},$$

for all $i \in [\ell]$, we can again combine Proposition 2.3.6 and Proposition 2.3.7 to construct disjoint subsets $S_1, \ldots, S_\ell \subseteq V$ such that

$$\psi(S_i) \leq \frac{\phi_{\pi''}(S_i'')}{1-\frac{1}{2}} = 2\phi_{\pi''}(S_i'').$$

Output them at the end of the algorithm, so that

$$\psi_\ell(S_1, \ldots, S_\ell) \leq 2(\phi_{\pi''})_\ell(S_1'', \ldots, S_\ell'') \lesssim C \cdot (\phi_{\pi''})_k(H) \lesssim C \cdot \psi_k(G),$$

as desired. Since the algorithm for Corollary 8.1.8 and the reductions all run in randomized polynomial time, our algorithm runs in randomized polynomial time as well. □

8.7 Concluding Remarks

In this chapter, we have provided a unifying approach for obtaining ARV-type approximation results for directed hypergraph expansion using reweighted eigenvalues, which encompasses directed edge expansion and directed vertex expansion. The SDP relaxation of the expansion quantities is simply by adding ℓ_2^2 triangle inequality constraints to the corresponding reweighted second eigenvalue. Using the same idea, we obtain new approximation results of multi-way expansion quantities such as vertex expansion and hypergraph expansion, by adding ℓ_2^2 triangle inequality constraints to the corresponding reweighted sum of first k eigenvalues. The approximation ratios obtained are similar to those in [LM14a, LM14b].

On the practical side, it is worth noting that the ARV-type approximation results in this chapter can be implemented in almost-linear time; see [LTW24] for details. While we have theoretical guarantee on their runtimes and approximation ratios, we are curious about whether they may be implemented to find good sparse cuts in large graphs quickly. Such implementation would bring these algorithms into the practical realm; in particular, fast spectral algorithms for computing hypergraph sparse cuts would be useful in certain machine learning applications, and fast algorithms for finding reweightings could be useful in graphical neural networks for hypergraphs and directed graphs.

Since multi-way graph partitioning has found many applications in clustering and classification, one interesting open area of research is to design fast approximation algorithms for multi-way graph partitioning that at least match the approximation guarantees presented in Section 8.1.2, and to generalize it to the directed (hyper)graph setting. Another open problem would be to generalize these results to more general classes of submodular functions, which were partially obtained in [COT23, CL24] using a different approach.

Chapter 9

Conclusion

In this thesis, we have developed the reweighted eigenvalue framework and used it to build a spectral theory for a general class of graph expansion problems. The inception of the framework was inspired by the fastest mixing Markov chain in [BDX04, Roc05].

Our results demonstrated that reweighted eigenvalues is a simple and effective framework for reducing the study of expansion quantities as general as directed hypergraph conductance to the basic setting of edge conductance in undirected graphs. Our first new idea of Eulerian reweighting was the key to extending the framework to directed graphs and directed hypergraphs. Our second new idea of k-th reweighted eigenvalue allowed us to capture direct analogues of the work of [Tre09, LOT12, LRTV12, KLL+13] in vastly more general settings. Our results have applications in graph partitioning, expander characterization, and mixing time analysis.

The table below summarizes the results we have obtained using reweighted eigenvalues.

	EC	VE	Dir. EC	Dir. VE	H. EC	Dir. H. EC
Cheeger	Y	Y	Y	Y	Y	Y
"Bipartite"	Y	**Y**	N	N	N	N
"Higher-Order"	Y	Y	N	N	Y	N
"Improved"	Y	**Y**	**Y**	**Y**	**Y**	**Y**
$O(\sqrt{\log n})$	Y	Y	Y	Y	Y	Y
Ortho. Sep.[1]	**Y**	**Y**	N	N	**Y**	N

Table 9.1: Summary of results. (Y = known result, **Y** = new result, N = no result)
(EC = Edge Conductance, VE = Vertex Expansion, Dir. = Directed, H. = Hypergraph.)

Philosophically, reweighted eigenvalues marks a significant departure from past work for designing rounding-based approximation algorithms. While most past work [BHT00, LRV13, Yos16, Yos19, Lou15, CLTZ18] defined a non-computable spectral quantity before relaxing it to a computable spectral quantity, reweighted eigenvalues operates on the ℓ_1 fractional program, symmetrizing it so that it can be lifted to a polynomial-time computable spectral quantity in a straightforward manner. It is possible that this idea may be useful in the design of approximation algorithms for very different problems.

It is our hope that reweighted eigenvalues may prove to be useful in tackling other important problems in graph theory and beyond.

We wrap up this thesis with some future research directions and open problems.

- Investigate the tightness of the Cheeger-type inequalities in Theorem 5.1.4, Theorem 5.1.8, and Theorem 6.3.23. A related problem is whether the projection loss using the Large Optimal Property can be improved (e.g. using generic chaining).

- Is there a combinatorial interpretation of the asymmetry ratio $\alpha(G_H)$ for directed hypergraphs in Proposition 6.3.19?

- One major obstacle for deriving a directed analogue of higher-order Cheeger inequality [LOT12, KLL$^+$13] is that the k-th reweighted eigenvalue does not relate to k-way expansion in directed graphs (see Section 5.10.1). Find the "correct" multi-way combinatorial quantity to relate to higher reweighted eigenvalues. Similarly, find a robust combinatorial characterization of λ_n^*.

- Establish a Cheeger-type inequality for small-set expansion of directed hypergraphs similar to [ABS10] (see Section 3.1.5). This has potential consequences in subexponential algorithms for generalized small-set expansion problems.

- Currently, the reweighted eigenvalues framework is largely limited to directed hypergraphs. Explore the second approach for applying reweighted eigenvalues to submodular transformations in Section 6.4. Find an interesting subclass of problems where this approach yields Cheeger-type inequalities and ARV-type results, potentially matching or generalizing the results of [COT23, CL24].

- The Cheeger rounding algorithms presented in this thesis all involve the dual solution (e.g. using r to order the vertices in Section 5.7.3). Other than being directly related to fastest mixing Markov chain, are there any algorithmic applications of the computed optimal reweighting?

[1] This refers to approximations of k-way expansion problems using orthogonal separators; see Chapter 8.

- While our results come with polynomial-time algorithms, they might not be fast enough for modern, large graphs. For ARV-type approximation of directed hypergraphs, an almost linear time algorithm was established in [LTW24]. Can we establish a fast algorithm for finding k disjoint sparse cuts (e.g. in undirected hypergraphs) that matches the guarantees in Section 8.1.2?

- Local algorithms [ST13, ACL06, ACL07, AP09] are pivotal in applying spectral algorithms to finding local sparse cuts in large graphs. In this context, one limitation of reweighted eigenvalues is that finding a reweighting seems to be a "global" task. Can we bypass this obstacle and design local algorithms for generalized expansion problems, using reweighted eigenvalues?

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 563–572, 2010.

[ACL06] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using pagerank vectors. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 475–486. IEEE, 2006.

[ACL07] Reid Andersen, Fan Chung, and Kevin Lang. Local partitioning for directed graphs using pagerank. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 166–178. Springer, 2007.

[ACMM05] Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev. $O(\sqrt{\log n})$ approximation algorithms for min UnCut, min 2CNF deletion, and directed cut problems. In *Proceedings of the 37th Annual Symposium on Theory of computing (STOC)*, pages 573–581, 2005.

[AF02] David Aldous and James Fill. Reversible Markov chains and random walks on graphs. *Unfinished monograph*, 2002 (recompiled 2014), 2002.

[AG11] Sanjeev Arora and Rong Ge. New tools for graph coloring. In *Proceedings of the 14th International Workshop on Approximation Algorithms for Combinatorial Optimization (APPROX)*, pages 1–12, 2011.

[AHK05] Sanjeev Arora, Elad Hazan, and Satyen Kale. Fast algorithms for approximate semidefinite programming using the multiplicative weights update method. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 339–348. IEEE, 2005.

[AK07] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the 39th Annual Symposium on Theory of Computing (STOC)*, pages 227–236, 2007.

[Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st Annual Symposium on Theory of Computing (STOC)*, pages 1–12, 2019.

[AM85] Noga Alon and Vitali D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[And70a] E.M. Andreev. On convex polyhedra in Lobačevskiĭ spaces. *Mathematics of the USSR-Sbornik*, 81:445–478, 1970.

[And70b] E.M. Andreev. On convex polyhedra of finite volume in Lobačevskiĭ spaces. *Mathematics of the USSR-Sbornik*, 83:256–260, 1970.

[AP09] Reid Andersen and Yuval Peres. Finding sparse cuts locally using evolving sets. In *Proceedings of the 41st Annual Symposium on Theory of Computing (STOC)*, pages 235–244, 2009.

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM*, 56(2):1–37, 2009.

[Bal97] Keith Ball. An elementary introduction to modern convex geometry. *Flavors of Geometry*, 1997.

[Bar96] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 184–193. IEEE, 1996.

[BDPX09] Stephen Boyd, Persi Diaconis, Pablo Parrilo, and Lin Xiao. Fastest mixing Markov chain on graphs with symmetries. *SIAM Journal on Optimization*, 20(2):792–819, 2009.

[BDSX06] Stephen Boyd, Persi Diaconis, Jun Sun, and Lin Xiao. Fastest mixing Markov chain on a path. *The American Mathematical Monthly*, 113(1):70–74, 2006.

[BDX04] Stephen Boyd, Persi Diaconis, and Lin Xiao. Fastest mixing Markov chain on a graph. *SIAM review*, 46(4):667–689, 2004.

[BFK+14] Nikhil Bansal, Uriel Feige, Robert Krauthgamer, Konstantin Makarychev, Viswanath Nagarajan, Joseph Seffi, and Roy Schwartz. Min-max graph partitioning and small set expansion. *SIAM Journal on Computing*, 43(2):872–904, 2014.

[BH11] Andries E. Brouwer and Willem H. Haemers. *Spectra of graphs*. Springer Science & Business Media, 2011.

[BHT00] Sergey Bobkov, Christian Houdré, and Prasad Tetali. λ_∞, vertex isoperimetry and concentration. *Combinatorica*, 20(2):153–172, 2000.

[BLR10] Punyashloka Biswal, James R. Lee, and Satish Rao. Eigenvalue bounds, spectral partitioning, and metrical deformations via flows. *Journal of the ACM (JACM)*, 57(3):1–23, 2010.

[Blu94] Avrim Blum. New approximation algorithms for graph coloring. *Journal of the ACM*, 41(3):470–516, 1994.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[CA15] Onur Cihan and Mehmet Akar. Fastest mixing reversible Markov chains on graphs with degree proportional stationary distributions. *IEEE Transactions on Automatic Control*, 60(1):227–232, 2015.

[Che70] Jeff Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. *Problems in Analysis*, 625:195–199, 1970.

[Chu05] Fan Chung. Laplacians and the Cheeger inequality for directed graphs. *Annals of Combinatorics*, 9:1–19, 2005.

[CKL+22] Li Chen, Rasmus Kyng, Yang P. Liu, Richard Peng, Maximilian Probst Gutenberg, and Sushant Sachdeva. Maximum flow and minimum-cost flow in almost-linear time. In *Proceedings of the 63rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 612–623. IEEE, 2022.

[CKN20] Yu Chen, Sanjeev Khanna, and Ansh Nagda. Near-linear size hypergraph cut sparsifiers. In *Proceedings of the 61st Conference on Foundations of Computer Science (FOCS)*, pages 61–72. IEEE, 2020.

[CKP+16] Michael B. Cohen, Jonathan Kelner, John Peebles, Richard Peng, Aaron Sidford, and Adrian Vladu. Faster algorithms for computing the stationary distribution, simulating random walks, and more. In *Proceedings of the 57th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 583–592, 2016.

[CKP+17] Michael B. Cohen, Jonathan Kelner, John Peebles, Richard Peng, Anup B. Rao, Aaron Sidford, and Adrian Vladu. Almost-linear-time algorithms for Markov chains and new spectral primitives for directed graphs. In *Proceedings of the 49th Annual Symposium on Theory of Computing (STOC)*, pages 410–419, 2017.

[CL24] Chandra Chekuri and Anand Louis. On sparsest cut and conductance in directed polymatroidal networks. *arXiv:2410.20525*, 2024.

[CLSZ20] Mihai Cucuringu, Huan Li, He Sun, and Luca Zanetti. Hermitian matrices for clustering directed graphs: insights and applications. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 983–992, 2020.

[CLTZ18] T.-H. Hubert Chan, Anand Louis, Zhihao Gavin Tang, and Chenzi Zhang. Spectral properties of hypergraph Laplacian and approximation algorithms. *Journal of the ACM*, 65(3):1–48, 2018.

[CMM06] Eden Chlamtac, Konstantin Makarychev, and Yury Makarychev. How to play unique games using embeddings. In *Proceedings of the 47th Conference on Foundations of Computer Science (FOCS)*, pages 687–696. IEEE, 2006.

[Col91] Yves Colin de Verdière. Un principe variationnel pour les empilements de cercles. *Inventiones Mathematicae*, 104:655–669, 1991.

[COT23] Antares Chen, Lorenzo Orecchia, and Erasmo Tani. Submodular hypergraph partitioning: Metric relaxations and fast algorithms via an improved cut-matching game. *arXiv:2301.08920*, 2023.

[Cov65] Thomas M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, EC-14(3):326–334, 1965.

[CS13] Chandra Chekuri and Anastasios Sidiropoulos. Approximation algorithms for Euler genus and related problems. In *Proceedings of the 54th Annual IEEE*

Symposium on Foundations of Computer Science (FOCS), pages 167–176. IEEE, 2013.

[CS18] T.-H. Hubert Chan and Bintao Sun. SDP primal-dual approximation algorithms for directed hypergraph expansion and sparsest cut with product demands. In *Proceedings of the 24th Annual International Computing and Combinatorics Conference (COCOON)*, pages 688–700. Springer, 2018.

[CTWZ19] T.-H. Hubert Chan, Zhihao Gavin Tang, Xiaowei Wu, and Chenzi Zhang. Diffusion operator and spectral analysis for directed hypergraph Laplacian. *Theoretical Computer Science*, 784:46–64, 2019.

[CTZ15] T.-H. Hubert Chan, Zhihao Gavin Tang, and Chenzi Zhang. Cheeger inequalities for general edge-weighted directed graphs. In *Proceedings of the 21st Annual International Computing and Combinatorics Conference (COCOON)*, pages 30–41. Springer, 2015.

[DLQ20] Sally Dong, Yin Tat Lee, and Kent Quanrud. Computing circle packing representations of planar graphs. In *Proceedings of the 31st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2860–2875. SIAM, 2020.

[EMPS16] Alina Ene, Gary L. Miller, Jakub Pachocki, and Aaron Sidford. Routing under balance. In *Proceedings of the 48th Annual Symposium on Theory of Computing (STOC)*, pages 598–611, 2016.

[FHL08] Uriel Feige, MohammadTaghi Hajiaghayi, and James R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM Journal on Computing*, 38(2):629–657, 2008.

[Fil91] James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *The Annals of Applied Probability*, 1(1):62–87, 1991.

[FK13] James Allen Fill and Jonas Kahn. Comparison inequalities and fastest mixing Markov chains. *Annals of Applied Probability*, 23(5):1778–1816, 2013.

[FLT20] Majid Farhadi, Anand Louis, and Prasad Tetali. On the complexity of λ_∞, vertex expansion, and spread constant of trees. *arXiv:2003.05582v1*, 2020.

[FM92] Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the 24th Annual Symposium on Theory of Computing (STOC)*, pages 26–38, 1992.

[FT03] Jittat Fakcharoenphol and Kunal Talwar. An improved decomposition theorem for graphs excluding a fixed minor. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 36–46. Springer, 2003.

[Fuj05] Satoru Fujishige. *Submodular functions and optimization*. Elsevier, 2005.

[GHT84] John R. Gilbert, Joan P. Hutchinson, and Robert Endre Tarjan. A separator theorem for graphs of bounded genus. *Journal of Algorithms*, 5(3):391–407, 1984.

[Gil07] Rafael Gillman. 0/1-*polytopes: typical and extremal properties*. PhD thesis, Technische Universität Berlin, 2007.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *Proceedings of the 44th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 534–543, 2003.

[GM12] Bernd Gärtner and Jiri Matousek. *Approximation algorithms and semidefinite programming*. Springer, 2012.

[GM17] Krystal Guo and Bojan Mohar. Hermitian adjacency matrix of digraphs and mixed graphs. *Journal of Graph Theory*, 85(1):217–248, 2017.

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, 2013.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(4):439–561, 2006.

[JLS86] William B. Johnson, Joram Lindenstrauss, and Gideon Schechtman. Extensions of Lipschitz maps into Banach spaces. *Israel Journal of Mathematics*, 54(2):129–138, 1986.

[JPV22] Vishesh Jain, Huy Tuan Pham, and Thuy-Duong Vuong. Dimension reduction for maximum matchings and the fastest mixing Markov chain. *arXiv:2203.03858*, 2022.

[Kai04] Volker Kaibel. On the expansion of graphs of 0/1-polytopes. In *The Sharpest Cut: The Impact of Manfred Padberg and His Work. MOS-SIAM Series on Optimization*, pages 199–216. SIAM, 2004.

[Kel06] Jonathan Kelner. Spectral partitioning, eigenvalue bounds, and circle packings for graphs of bounded genus. *SIAM Journal on Computing*, 35(4):882–902, 2006.

[KKTY22] Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Spectral hypergraph sparsifiers of nearly linear size. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1159–1170. IEEE, 2022.

[KLL+13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved Cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *Proceedings of the 45th Annual Symposium on Theory of Computing (STOC)*, pages 11–20, 2013.

[KLL16] Tsz Chiu Kwok, Lap Chi Lau, and Yin Tat Lee. Improved Cheeger's inequality and analysis of local graph partitioning using vertex expansion and expansion profile. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1848–1861. SIAM, 2016.

[KLPT11] Jonathan Kelner, James R. Lee, Gregory N. Price, and Shang-Hua Teng. Metric uniformization and spectral bounds for graphs. *Geometric and Functional Analysis*, 21(5):1117–1143, 2011.

[KLT22] Tsz Chiu Kwok, Lap Chi Lau, and Kam Chuen Tung. Cheeger inequalities for vertex expansion and reweighted eigenvalues. In *Proceedings of the 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 366–377. IEEE, 2022.

[Koe36] Paul Koebe. Kontaktprobleme der konformen Abbildung. *Ber. Verh. Sächs. Akad. Leipzig*, 88:141–164, 1936.

[KPR93] Philip Klein, Serge A. Plotkin, and Satish Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th Annual Symposium on Theory of Computing (STOC)*, pages 682–690, 1993.

[KR03] Volker Kaibel and Anja Remshagen. On the graph-density of random 0/1-polytopes. In *Proceedings of the 6th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 318–328, 2003.

[KRV09] Rohit Khandekar, Satish Rao, and Umesh Vazirani. Graph partitioning using single commodity flows. *Journal of the ACM (JACM)*, 56(4):1–15, 2009.

[KS15] Ken-ichi Kawarabayashi and Anastasios Sidiropoulos. Beyond the Euler characteristic: Approximating the genus of general graphs. In *Proceedings of the 47th Annual Symposium on Theory of Computing (STOC)*, pages 675–682, 2015.

[KS19] Ken-ichi Kawarabayashi and Anastasios Sidiropoulos. Polylogarithmic approximation for Euler genus on bounded degree graphs. In *Proceedings of the 51st Annual Symposium on Theory of Computing (STOC)*, pages 164–175, 2019.

[Kur30] Casimir Kuratowski. Sur le probleme des courbes gauches en topologie. *Fundamenta mathematicae*, 15(1):271–283, 1930.

[Li11] Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics and Statistics*, 4(1):66–70, 2011.

[LL15] Jianxi Liu and Xueliang Li. Hermitian-adjacency matrices and Hermitian energies of mixed graphs. *Linear Algebra and its Applications*, 466(1):182–207, 2015.

[LM00] Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, pages 1302–1338, 2000.

[LM14a] Anand Louis and Konstantin Makarychev. Approximation algorithm for sparsest k-partitioning. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1244–1255. SIAM, 2014.

[LM14b] Anand Louis and Yury Makarychev. Approximation algorithms for hypergraph small set expansion and small set vertex expansion. In *Proceedings of Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 339–355. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2014.

[LM18] Pan Li and Olgica Milenkovic. Submodular hypergraphs: p-Laplacians, Cheeger inequalities and spectral clustering. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 3014–3023. PMLR, 2018.

[LOT12] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order Cheeger inequalities. In *Proceedings of the 44th Annual Symposium on Theory of Computing (STOC)*, pages 1117–1130, 2012.

[Lou15] Anand Louis. Hypergraph Markov operators, eigenvalues and approximation algorithms. In *Proceedings of the 47th Annual Symposium on Theory of Computing (STOC)*, pages 713–722, 2015.

[Lov83] László Lovász. Submodular functions and convexity. *Mathematical Programming — The State of the Art*, pages 235–257, 1983.

[LP17] David Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Society, 2017.

[LR99] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, 1999.

[LR05] Monique Laurent and Franz Rendl. Semidefinite programming and integer programming. *Handbooks in Operations Research and Management Science*, 12:393–514, 2005.

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1131–1140, 2012.

[LRV13] Anand Louis, Prasad Raghavendra, and Santosh Vempala. The complexity of approximating vertex expansion. In *Proceedings of the 54th IEEE Annual Symposium on Foundations of Computer Secience (FOCS)*, pages 360–369, 2013.

[LS10] James R. Lee and Anastasios Sidiropoulos. Genus and the geometry of the cut graph. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 193–201. SIAM, 2010.

[LS20] Steinar Laenen and He Sun. Higher-order spectral clustering of directed graphs. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020.

[LSZ19] Huan Li, He Sun, and Luca Zanetti. Hermitian Laplacians and a Cheeger inequality for the Max-2-Lin Problem. In *Proceedings of the 27th Annual European Symposium on Algorithms (ESA)*, pages 71:1–71:14. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[LT79] Richard J. Lipton and Robert Endre Tarjan. A separator theorem for planar graphs. *SIAM Journal on Applied Mathematics*, 36(2):177–189, 1979.

[LTW23] Lap Chi Lau, Kam Chuen Tung, and Robert Wang. Cheeger inequalities for directed graphs and hypergraphs using reweighted eigenvalues. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1834–1847, 2023.

[LTW24] Lap Chi Lau, Kam Chuen Tung, and Robert Wang. Fast algorithms for directed graph partitioning using flows and reweighted eigenvalues. In *Proceedings of the 35th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 591–624. SIAM, 2024.

[Lux07] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[Mih89] Milena Mihail. Conductance and convergence of Markov chains — a combinatorial treatment of expanders. In *Proceedings of the 30th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 526–531, 1989.

[MMR19] Konstantin Makarychev, Yury Makarychev, and Ilya Razenshteyn. Performance of Johnson-Lindenstrauss transform for k-means and k-medians clustering. In *Proceedings of the 51st Annual Symposium on Theory of Computing (STOC)*, pages 1027–1038, 2019.

[Moh93] Bojan Mohar. A polynomial time circle packing algorithm. *Discrete Mathematics*, 117(1-3):257–263, 1993.

[OSVV08] Lorenzo Orecchia, Leonard Schulman, Umesh Vazirani, and Nisheeth Vishnoi. On partitioning graphs via single commodity flows. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 461–470, 2008.

[OZ22] Sam Olesker-Taylor and Luca Zanetti. Geometric bounds on the fastest mixing markov chain. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2022.

[Roc05] Sébastien Roch. Bounding fastest mixing. *Electronic Communications in Probability*, 10:282–296, 2005.

[Rot16] Thomas Rothvoss. Lecture notes on the ARV algorithm for sparsest cut. *arXiv:1607.00854*, 2016.

[RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 755–764, 2010.

[RY68] Gerhard Ringel and John W. T. Youngs. Solution of the Heawood map-coloring problem. *Proceedings of the National Academy of Sciences*, 60(2):438–445, 1968.

[RY22] Akbar Rafiey and Yuichi Yoshida. Sparsification of decomposable submodular functions. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence*, pages 10336–10344, 2022.

[She09] Jonah Sherman. Breaking the multicommodity flow barrier for $O(\sqrt{\log n})$-approximations to sparsest cut. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 363–372. IEEE, 2009.

[Sim95] Stephen Simons. Minimax theorems and their proofs. In *Minimax and applications*, pages 1–23. Springer, 1995.

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, 1989.

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

[Spi19] Daniel Spielman. Spectral and algebraic graph theory. *Yale lecture notes (draft)*, 4:47, 2019.

[ŠS06] Jirı Šıma and Satu Elisa Schaeffer. On the NP-completeness of some graph cluster measures. *SOFSEM 2006: Theory and Practice of Computer Science*, page 530, 2006.

[ST96] Daniel A Spielman and Shang-Hua Teng. Spectral partitioning works: Planar graphs and finite element meshes. In *Proceedings of the 37th Conference on Foundations of Computer Science (FOCS)*, pages 96–105. IEEE, 1996.

[ST13] Daniel A Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM Journal on computing*, 42(1):1–26, 2013.

[ST14] Daniel A Spielman and Shang-Hua Teng. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM Journal on Matrix Analysis and Applications*, 35(3):835–885, 2014.

[Ste10] David Steurer. *On the complexity of unique games and graph expansion*. PhD thesis, Princeton University, 2010.

[Str16] Gilbert Strang. *Introduction to linear algebra*. SIAM, 2016.

[Tal05] Michel Talagrand. *The generic chaining: upper and lower bounds of stochastic processes*. Springer Science & Business Media, 2005.

[Tan84] R. M. Tanner. Explicit concentrators from generalized N-gons. *SIAM Journal on Algebraic Discrete Methods*, 5(3):287–294, 1984.

[Tho89] Carsten Thomassen. The graph genus problem is NP-complete. *Journal of Algorithms*, 10(4):568–576, 1989.

[Thu78] William Thurston. The geometry and topology of 3-manifolds. *Princeton University Lecture Notes*, 1978.

[Tko12] Tomasz Tkocz. An upper bound for spherical caps. *The American Mathematical Monthly*, 119(7):606–607, 2012.

[Tre09] Luca Trevisan. Max cut and the smallest eigenvalue. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 263–272, 2009.

[Tre16] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. *University of California, Berkeley*, 2016.

[WSV12] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Science & Business Media, 2012.

[Yos16] Yuichi Yoshida. Nonlinear Laplacian for digraphs and its applications to network analysis. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 483–492, 2016.

[Yos19] Yuichi Yoshida. Cheeger inequalities for submodular transformations. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2582–2601. SIAM, 2019.

Appendix A

Deferred Proofs for Chapter 4

A.1 Deferred Proofs for Weighted Vertex Expansion

Proof of Lemma 4.3.8. The case where there is an optimizer $S \subseteq V$ to $\psi(G)$, with $0 < \pi(S) \leq 1/2$ and $\psi(S) = \psi(G)$, has been proven in Proposition 3.2.8. The other case is when $\psi(G) = 1$. We will show that $\gamma^{(1)}(G) \leq 2$ and this would imply that $\gamma^{(1)}(G) \leq 2\psi(G)$. Let v be a vertex with $0 < \pi(v) \leq 1/2$, which must exist as long as the graph has at least two vertices. We define a solution f, g to $\gamma^{(1)}(G)$ as in Proposition 3.2.8 with $S = V \setminus \{v\}$. Following the same arguments, f, g is a feasible solution to $\gamma^{(1)}(G)$ with objective value

$$\sum_{v \in V} \pi(v) g(v) = \pi(\partial S) \left(\frac{1}{\pi(v)} + \frac{1}{1 - \pi(v)} \right) \leq \pi(v) \cdot \frac{2}{\pi(v)} = 2.$$

This proves $\gamma^{(1)}(G) \leq 2\psi(G)$ in the other case when $\psi(G) = 1$. □

Proof of Theorem 4.3.7. The easy direction $\gamma^{(1)}(G) \lesssim \psi(G)$ is proved in Lemma 4.3.8. For the hard direction, given a solution $(f, \overrightarrow{E})$ to $\vec{\gamma}^{(1)}(G)$, we apply the ℓ_2^2 to ℓ_1 step to obtain $h : V \to \mathbb{R}$ with π-weighted median 0, then the threshold rounding step in Proposition 4.3.10 on h, to obtain a set S with $0 < \pi(S) \leq 1/2$ and $\vec{\psi}_\tau(S) \lesssim \sqrt{\vec{\gamma}^{(1)}(G)}$. If $\vec{\psi}_\tau(S) \geq 1/2$, then it implies that $\vec{\gamma}^{(1)}(G) = \Omega(1)$, and so the inequality $\psi(G)^2 \lesssim \vec{\gamma}^{(1)}(G)$ holds trivially as $\psi(G) \leq 1$ by definition. Otherwise, if $\vec{\psi}_\tau(S) < 1/2$, we apply the postprocessing step in Lemma 4.3.11 on S to obtain $S' \subseteq S$ with $\psi(S') \leq 2\vec{\psi}_\tau(S)$. Therefore, S' is a set with $\pi(S') \leq \pi(S) \leq \pi(\text{supp}(x)) \leq 1/2$ and $\psi(S') \leq 2\vec{\psi}_\tau(S) \lesssim \sqrt{\vec{\gamma}^{(1)}(G)}$. Thus we conclude the hard direction that $\psi(G) \leq \psi(S') \lesssim \sqrt{\vec{\gamma}^{(1)}(G)} \lesssim \sqrt{\gamma^{(1)}(G)}$. □

A.2 Deferred Proofs for Bipartite Vertex Expansion

Proof of Lemma 4.4.8. First we consider the case that there is an optimizer $S \subseteq V$ to $\psi_B(G)$, with bipartition $S = S_1 \sqcup S_2$ and $\pi(\partial S)/\pi(S) = \psi_B(G)$. Define $f : V \to \mathbb{R}$ and $g : V \to \mathbb{R}$ as follows:

- $f(u) = 1$ if $u \in S_1$, and $f(u) = -1$ if $u \in S_2$, and $f(u) = 0$ if $u \notin S$;
- $g(u) = 1$ if $u \in \partial S$, and $g(u) = 0$ otherwise.

For each edge $uv \in E$, we claim that $g(u)+g(v) \geq (f(u)+f(v))^2$. Note that u and v cannot both belong to S_1 or both belong to S_2. One can check that the constraint is satisfied in all the remaining cases: (1) $u \in S_1$ and $v \in S_2$ (or vice versa), (2) $u \in S$ and $v \notin S$ (or vice versa), and (3) $u, v \notin S$. So, (f, g) is a feasible solution to the $\nu^{(1)}(G)$ program, and the objective value is

$$\frac{\sum_{v \in V} \pi(v) g(v)}{\sum_{v \in V} \pi(v) f(v)^2} = \frac{\sum_{v \in \partial S} \pi(v)}{\sum_{v \in S_1 \cup S_2} \pi(v)} = \frac{\pi(\partial S)}{\pi(S)} = \psi(S).$$

This implies that $\nu^{(1)}(G) \leq 2\psi(S)$ in this case. The other case is when $\psi_B(G) = 1$. Choosing the feasible solution $f \equiv 1$ and $g \equiv 2$ to the $\nu^{(1)}(G)$ program shows that $\nu^{(1)}(G) \leq 2$. This implies that $\nu^{(1)}(G) \leq 2\psi_B(G)$ in the other case. □

Proof of Theorem 4.4.4. The easy direction $\nu^{(1)}(G) \lesssim \psi_B(G)$ is proved in Lemma 4.4.8. For the hard direction, given a solution $(f, \overrightarrow{E})$ to $\vec{\nu}^{(1)}(G)$, we apply the ℓ_2^2 to ℓ_1 step and then the threshold rounding step in Proposition 4.4.10 to obtain two disjoint sets $S_1, S_2 \subseteq V$ with $\vec{\psi}_\tau(S_1, S_2) \lesssim \sqrt{\vec{\nu}^{(1)}(G)}$. If $\vec{\psi}_\tau(S_1, S_2) \geq 1/2$, then it implies that $\vec{\nu}^{(1)}(G) = \Omega(1)$, and so the inequality $\psi_B(G)^2 \lesssim \vec{\nu}^{(1)}(G)$ holds trivially as $\psi_B(G) \leq 1$ by definition. Otherwise, if $\vec{\psi}_\tau(S_1, S_2) < 1/2$, we apply the postprocessing step in Lemma 4.4.11 on (S_1, S_2) to obtain (S_1', S_2') so that $S_1' \cup S_2'$ is an induced bipartite graph in G and $\psi(S_1' \cup S_2') \leq 2\vec{\psi}_\tau(S_1, S_2)$. Thus we conclude the hard direction that $\psi_B(G) \leq \psi(S_1' \cup S_2') \leq 2\vec{\psi}_\tau(S_1, S_2) \lesssim \sqrt{\vec{\nu}^{(1)}(G)}$. □

A.3 Deferred Proofs for Multiway Vertex Expansion

Proof of Lemma 4.5.11. If $\psi_k(G) \geq 1$, then the lemma holds trivially as $\lambda_k^*(G) \leq 2$. Henceforth, we assume $\psi_k(G) < 1$, and there are nonempty disjoint subsets $S_1, \ldots, S_k \subseteq V$ with $\max_{1 \leq i \leq k} \psi(S_i) = \psi_k(G)$.

Using the notation in Proposition 4.5.2, the λ_k^* program in Definition 4.1.9 can be written as

$$\begin{aligned}
\lambda_k^*(G) \;:=\; \max_{Q \geq 0} \min_{f_1,\ldots,f_k: V \to \mathbb{R}} \max_{1 \leq i \leq k} \quad & f_i^T(\Pi - Q) f_i \\
\text{subject to} \quad & \sum_v \pi(v) f_i(v)^2 = 1 && \forall 1 \leq i \leq k \\
& \sum_v \pi(v) f_i(v) f_j(v) = 0 && \forall 1 \leq i \neq j \leq k \\
& Q(u,v) = 0 && \forall uv \notin E \\
& \sum_{v \in V} Q(u,v) = \pi(u) && \forall u \in V \\
& Q(u,v) = Q(v,u) && \forall uv \in E.
\end{aligned}$$

Each $f_i^T(\Pi - Q) f_i$ can be written as

$$\sum_{uv \in E} Q(u,v)\big(f_i(u) - f_i(v)\big)^2 = \sum_{uv \in E} \pi(u) P(u,v) \big(f_i(u) - f_i(v)\big)^2.$$

We set

$$f_i(u) := \begin{cases} \frac{1}{\sqrt{\pi(S_i)}}, & \text{if } u \in S_i \\ 0, & \text{otherwise.} \end{cases}$$

We can check that the constraints on f_i are satisfied. Moreover, for any P satisfying the constraints,

$$\begin{aligned}
\sum_{uv \in E} \pi(u) P(u,v) \big(f_i(u) - f_i(v)\big)^2 \;&=\; \sum_{v \in S_i, u \in \partial(S_i)} \pi(u) P(u,v) \frac{1}{\pi(S_i)} \\
&\leq\; \frac{1}{\pi(S_i)} \sum_{v \in V, u \in \partial(S_i)} \pi(u) P(u,v) \;=\; \frac{\pi(\partial(S_i))}{\pi(S_i)} \;=\; \psi(S_i).
\end{aligned}$$

So, $\max_{1 \leq i \leq k} f_i^T(\Pi - \Pi P) f_i \leq \max_{1 \leq i \leq k} \psi(S_i)$. Taking maximum over P gives $\lambda_k^*(G) \leq \psi_k(G)$. $\square$

A.4 Deferred Proofs for 0/1-Polytopes with Poor Vertex Expansion

Proof of Lemma 4.7.4. The plan is to prove that the stated conditions imply $\frac{1}{2}(x+y) \in \operatorname{conv}(M)$, which would then immediately imply that there is no edge connecting x and y in Q. We will prove the contrapositive: if $\frac{1}{2}(x+y) \notin \operatorname{conv}(M)$, then there is a p-consistent affine function l such that $l(z) < 0$ for all $z \in M$ matching the pattern p.

Denote $w := \frac{1}{2}(x+y)$. As $w \notin \operatorname{conv}(M)$, by Proposition 2.7.2, there is an affine function $l' : (u_1, \ldots, u_n) \mapsto \beta' + \sum_i \alpha'_i u_i$ such that $l'(w) = 0$ and $l'(z) < 0$ for all $z \in M$. We would like to modify l' to obtain an affine function $l : (u_1, \ldots, u_n) \mapsto \beta + \sum_i \alpha_i u_i$ such that (i) $l(w) = 0$, (ii) $\alpha_i = 0$ for $i \in \operatorname{supp}(p)$, and (iii) $l(z) < 0$ for all $z \in M$ matching the pattern p.

Note that, by the definition of common pattern, for $i \in \operatorname{supp}(p)$, either $x_i = y_i = 1$ or $x_i = y_i = 0$, and so $w_i \in \{0, 1\}$. Also, for any $z \in \{0,1\}^n$ that matches the pattern p, we must have $z_i = w_i$ for $i \in \operatorname{supp}(p)$. So, for any such z,

$$\begin{aligned} l'(z) = \beta' + \sum_{i=1}^{n} \alpha'_i z_i &= \left(\beta' + \sum_{i \in \operatorname{supp}(p)} \alpha'_i z_i\right) + \sum_{i \notin \operatorname{supp}(p)} \alpha'_i z_i \\ &= \left(\beta' + \sum_{i \in \operatorname{supp}(p)} \alpha'_i w_i\right) + \sum_{i \notin \operatorname{supp}(p)} \alpha'_i z_i. \end{aligned}$$

Hence, if we set $\beta := \beta' + \sum_{i \in \operatorname{supp}(p)} \alpha'_i w_i$, and $\alpha_i = \alpha'_i$ for $i \notin \operatorname{supp}(p)$ and $\alpha_i = 0$ for $i \in \operatorname{supp}(p)$, then the affine function $l : (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n \mapsto \beta + \sum_i \alpha_i u_i$ satisfies $l(z) = l'(z)$ for any z that matches the pattern p. Therefore, l is an affine function that satisfies the three properties that (i) $l(w) = l'(w) = 0$, (ii) $\alpha_i = 0$ for $i \in \operatorname{supp}(p)$, and (iii) $l(z) = l'(z) < 0$ for $z \in M$ matching the pattern p. □

Proof of Lemma 4.7.5. Note that for $x \in L$ and $y \in R$ in our construction in Definition 4.7.1, their common pattern p satisfies $|\operatorname{supp}_0(p)| = |\operatorname{supp}_1(p)| \leq k$. The lemma follows by applying Lemma 4.7.4 on all possible such patterns. □

Proof of Lemma 4.7.6. Let $z \in \{0,1\}^n$ be such that $|z| = \frac{n}{2}$ and z matching the pattern p. Let $z^\tau \in \{0,1\}^n$ be its "opposite point" formed by toggling the coordinates of z_i for $i \notin \operatorname{supp}(p)$ and leaving other coordinates unchanged. Note that the lemma follows from the following two facts: (i) $|z^\tau| = \frac{n}{2}$ and z^τ matches the pattern p, and (ii) $l(z) + l(z^\tau) = 0$.

For the first fact, z^τ matches the pattern p because $z_i^\tau = z_i$ for $i \in \mathrm{supp}(p)$. And $|z^\tau| = \frac{n}{2}$ because $|z| = \frac{n}{2}$ and there are the same number of zeroes and ones in p, the latter being a consequence of $|x| + |y| = k + (n-k) = n$.

For the second fact, as l is p-consistent,

$$l(z) + l(z^\tau) = \Big(\beta + \sum_{i \notin \mathrm{supp}(p)} \alpha_i z_i\Big) + \Big(\beta + \sum_{i \notin \mathrm{supp}(p)} \alpha_i z_i^\tau\Big) = 2\Big(\beta + \frac{1}{2} \sum_{i \notin \mathrm{supp}(p)} \alpha_i\Big) = 0,$$

where the last equality is from the second condition in Definition 4.7.3. □

Proof of Lemma 4.7.7. The number of points $z \in \{0,1\}^n$ with $|z| = \frac{n}{2}$ is $\binom{n}{n/2}$, whereas the number of such points that matches pattern p is $\binom{n-2s}{n/2-s}$. Therefore,

$$\begin{aligned} \Pr_{z \sim Z}[z \text{ matches pattern } p] \quad &= \quad \binom{n-2s}{n/2-s} \Big/ \binom{n}{n/2} \\ &\gtrsim \quad \left(\sqrt{\frac{2}{\pi(n-2s)}} \cdot 2^{n-2s}\right) \Big/ \left(\sqrt{\frac{2}{\pi n}} \cdot 2^n\right) \ \geq \ 4^{-s}, \end{aligned}$$

where we used Stirling's approximation $n! \approx_n \sqrt{2\pi n}(n/e)^n$. □

Proof of Theorem 4.1.12. Let Q be a 0/1-polytope from Definition 4.7.1. We would like to apply Lemma 4.7.5 to prove the theorem.

Let Z be the uniform distribution on $\{z \in \{0,1\}^n : |z| = n/2\}$. Combining Lemma 4.7.6 and Lemma 4.7.7, it follows that for any pattern p with $|\mathrm{supp}_0(p)| = |\mathrm{supp}_1(p)| = s \leq k$ and any p-consistent affine function l,

$$\Pr_{z \sim Z}\left[\, l(z) \geq 0 \text{ and } z \text{ matches pattern } p \,\right] \geq c \cdot 4^{-s}$$

for some universal constant $c > 0$. Therefore, if we take independent samples $z_1, \ldots, z_m \sim Z$ and set $M := \{z_1, \ldots, z_m\}$ where m is a value to be determined later, then for any pattern p with $|\mathrm{supp}_0(p)| = |\mathrm{supp}_1(p)| = s \leq k$ and any p-consistent affine function l,

$$\Pr[\nexists z \in M \text{ with } l(z) \geq 0 \text{ and } z \text{ matching pattern } p] \ \leq \ (1 - c \cdot 4^{-s})^m \ \leq \ (1 - c \cdot 4^{-k})^m.$$

To apply a union bound, we upper bound the numbers of different such p and l. The number of patterns p with $|\mathrm{supp}_0(p)| = |\mathrm{supp}_1(p)| = s \leq k$ is

$$\sum_{s=0}^{k} \binom{n}{2s} \cdot \binom{2s}{s} \leq (k+1) \cdot n^{2k}.$$

The number of p-consistent affine functions l with different sign patterns on the boolean hypercube $\{0,1\}^n$ is upper bounded by the number of affine threshold functions on $\{0,1\}^n$, which is at most 2^{n^2} by Proposition 4.7.8. Combining the two estimates and the above probability bound, the failure probability is

$$\Pr[\exists p\, \exists l \text{ s.t. } \nexists z \in M \text{ with } l(z) \geq 0 \text{ and } z \text{ matching pattern } p] \leq (1-c\cdot 4^{-k})^m \cdot (k+1) \cdot n^{2k} \cdot 2^{n^2}.$$

Setting

$$m \;=\; \frac{4^k}{c} \cdot (1 + \log(k+1) + 2k \log n + n^2 \log 2) \;\lesssim\; 4^k n^2,$$

the failure probability is at most

$$\begin{aligned} & (1 - c \cdot 4^{-k})^m \cdot (k+1) \cdot n^{2k} \cdot 2^{n^2} \\ \leq\;& \exp(-c \cdot 4^{-k} \cdot m) \cdot \exp(\log(k+1) + 2k \log n + n^2 \log 2) \\ \leq\;& \exp(-(1 + \log(k+1) + 2k \log n + n^2 \log 2)) \cdot \exp(\log(k+1) + 2k \log n + n^2 \log 2) \\ =\;& e^{-1}. \end{aligned}$$

Therefore, by Lemma 4.7.5, we conclude that there exists $M \subseteq \{z \in \{0,1\}^n : |z| = n/2\}$ with $|M| \lesssim 4^k n^2$ such that there are no edges between L and R in the graph of Q.

Finally, as $|L| = \binom{n}{k} \geq (n/k)^k$ and $\partial L \subseteq M$, it follows that

$$\psi(L) \leq \frac{|M|}{|L|} \lesssim \frac{4^k n^2}{(n/k)^k} = \frac{(4k)^k}{n^{k-2}}.$$

□

A.5 Deferred Proofs for Spherical Proximity Graph

Proof of Proposition 4.8.3. The construction largely follows that of [GM12, Lemma 8.3.22]. Given n and $\gamma > 0$, we iteratively choose points $y_1, y_2, \ldots, y_m \in \mathbb{S}^{k-1}$, such that each new point y_{i+1} has distance at least $\gamma/2$ from $y_1, \ldots, y_i$. We stop when it is no longer possible to choose a point that is $(\gamma/2)$-far from all existing points.

We now let $S'_1, S'_2, \ldots, S'_m$ be the cells of the Voronoi diagram of $y_1, y_2, \ldots, y_m$. That is, for any $x \in \mathbb{S}^{k-1}$, $x \in S'_i$ iff $d(x, y_i) = \min_{j \in [m]} d(x, y_j)$. Note that cell S'_i contains $B(y_i, \gamma/4)$ and is contained in $B(y_i, \gamma/2)$. Therefore, the measure of each S'_i is at least $\varepsilon = \mu(\mathrm{Cap}(\gamma/4))$ and the diamater of each S'_i is at most γ. By further subdividing the cells (evenly) until each cell has measure $\leq 2\varepsilon$, we obtain $S_1, S_2, \ldots, S_n$, such that the measure of each S_i is between ε and 2ε, and $\mathrm{diam}(S_i) \leq \gamma$ for each $i \in [n]$. We can choose the points $x_i \in S_i$ arbitrarily. □

Proof of Lemma 4.8.4. First, we prove the bound on the maximum degree. For any vertex $u \in V$, its degree is equal to the number of points x_v that are δ-close to x_u. We can count the number of such points using volume estimation. If a point x_v is within distance δ from x_u, then the entire cell S_v is within distance $\delta + \gamma$ from x_u. This contributes at least $\varepsilon = \mu(\mathrm{Cap}(\gamma/4))$ total measure to $B(x_u, \delta + \gamma)$. The total measure of these cells is at most $\mu(B(x_u, \delta + \gamma))$, so

$$|\partial(u)| \cdot \mu(\mathrm{Cap}(\gamma/4)) \leq \mu(B(x_u, \delta + \gamma)).$$

Rearranging gives the desired upper bound on $|\partial(u)|$, and thus on the maximum degree.

Next, we prove the bound on the vertex expansion. Given any $T \subseteq V$, we wish to lower bound $|\partial(T)|$. For $j \in V$, if the cell S_j is completely contained in $S_T + \mathrm{Cap}(\delta)$, then $x_j \in T \cup \partial(T)$. If we take the union of all such cells, then the set will contain $S_T + \mathrm{Cap}(\delta - \gamma)$:

$$\cup\{S_j : S_j \subseteq S_T + \mathrm{Cap}(\delta)\} \supseteq S_T + \mathrm{Cap}(\delta - \gamma).$$

This is because, for any point $u \in S_T + \mathrm{Cap}(\delta - \gamma)$, the cell containing u will be inside

$$\{u\} + \mathrm{Cap}(\gamma) \subseteq S_T + \mathrm{Cap}(\delta - \gamma) + \mathrm{Cap}(\gamma) = S_T + \mathrm{Cap}(\delta).$$

It follows that

$$\cup\{S_j : j \notin T \text{ and } S_j \subseteq S_T + \mathrm{Cap}(\delta)\} \supseteq (S_T + \mathrm{Cap}(\delta - \gamma)) \setminus S_T.$$

Combining previous observations, and since each cell has measure at most 2ε,

$$\begin{aligned} |\partial(T)| = |(T \cup \partial(T)) \setminus T| &\geq |\{S_j : j \notin T \text{ and } S_j \subseteq S_T + \mathrm{Cap}(\delta)\}| \\ &\geq \frac{\mu(S_T + \mathrm{Cap}(\delta - \gamma)) - \mu(S_T)}{2\varepsilon}. \end{aligned}$$

We are done after substituting this and $|T| \leq \mu(S_T)/\varepsilon$ into $\psi(T) = |\partial(T)|/|T|$. □

Proof of Proposition 4.8.5. We will use the following formula for spherical cap volume:

$$\mu(\mathrm{Cap}(x)) = \frac{\int_0^x \sin^{k-2}\theta \, d\theta}{\int_0^\pi \sin^{k-2}\theta \, d\theta}.$$

This comes from the formula for unnormalized spherical cap volume [Li11]:

$$\mu_0(\mathrm{Cap}(x)) = \frac{2\pi^{(k-1)/2}}{\Gamma((k-1)/2)} \cdot \int_0^x \sin^{k-2}\theta \, d\theta.$$

We shall use the approximation

$$\sin^k \theta \approx_k \left(\theta - \frac{\theta^3}{3!}\right)^k$$

for $0 \leq \theta \leq O(1/\sqrt{k})$. The third-degree approximation is sufficient because for θ in this range, $(\theta - \theta^3/3! + O(\theta^5))^k \approx_k (\theta - \theta^3/3!)^k$. Then, for $x = O(1/\sqrt{k})$,

$$\begin{aligned}
\int_0^x \sin^k \theta \, d\theta &\approx_k \int_0^x \left(\theta - \frac{\theta^3}{3!}\right)^k d\theta \\
&\approx_k \int_0^x \left(\theta - \frac{\theta^3}{3!}\right)^k \cdot (1 - \theta^2/2) \, d\theta \quad (\because (1 - \theta^2/2) \text{ is close to } 1) \\
&\overset{y:=\theta-\theta^3/6}{=} \int_0^{x-x^3/6} y^k \, dy \quad (\because y(\theta) := \theta - \theta^3/6 \text{ is increasing for } \theta \in [0, x]) \\
&= \frac{(x - x^3/6)^{k+1}}{k+1}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{\mu(\mathrm{Cap}(\delta + \gamma))}{\mu(\mathrm{Cap}(\gamma/4))} &= \frac{\int_0^{\delta+\gamma} \sin^{k-2} \theta \, d\theta}{\int_0^{\gamma/4} \sin^{k-2} \theta \, d\theta} \\
&\approx_k \left(\frac{\delta + \gamma}{\gamma/4}\right)^{k-1} \cdot \left(\frac{1 - (\delta + \gamma)^2/6}{1 - (\gamma/4)^2/6}\right)^{k-1} \\
&= \left(\frac{4(c_1 + c_2)}{c_1}\right)^{k-1} \cdot \left(\frac{1 - (c_1 + c_2)/6k}{1 - c_1/96k}\right)^{k-1} \\
&\lesssim \left(\frac{4(c_1 + c_2)}{c_1}\right)^k \\
&\leq 2^{O(k)}.
\end{aligned}$$

□

Proof of Proposition 4.8.6. By isoperimetric inequality on the sphere (see e.g. [Bal97] for reference), LHS is minimized when the set S_T is a spherical cap. Therefore, it suffices to prove that, for any $\tau \in (0, \pi/2]$,

$$\mu(\mathrm{Cap}(\tau + (\delta - \gamma))) \geq (1 + \Omega(1)) \cdot \mu(\mathrm{Cap}(\tau)).$$

Let $c := c_2 - c_1 > 0$. This is equivalent to

$$\int_0^{\tau + c/\sqrt{k}} \sin^{k-2}\theta \, d\theta \geq (1 + \Omega(1)) \int_0^{\tau} \sin^{k-2}\theta \, d\theta.$$

For technical reasons, we first deal with the case where τ is close to $\pi/2$. Indeed, when $\tau \geq \pi/2 - c/\sqrt{k}$, the result follows from well-known upper bounds on spherical cap volume. For example, we may use the upper bound in [Tko12]:

$$\mu(\mathrm{Cap}(\pi/2 - \theta)) \leq e^{-k \sin^2 \theta / 2}, \qquad \theta \in [0, \pi/2)$$

and the fact that $\sin\theta \approx_k \theta$ for $\theta = O(1/\sqrt{k})$. Then, if c is such that

$$\exp\left[-k \sin^2\left(\frac{c}{2\sqrt{k}}\right)/2\right] \leq 1/3,$$

one has

$$\frac{\mu(\mathrm{Cap}(\tau + c/\sqrt{k}))}{\mu(\mathrm{Cap}(\tau))} \geq \min\left(\frac{\mu(\mathrm{Cap}(\pi/2))}{\mu(\mathrm{Cap}(\pi/2 - c/2\sqrt{k}))}, \frac{\mu(\mathrm{Cap}(\pi/2 + c/2\sqrt{k}))}{\mu(\mathrm{Cap}(\pi/2))}\right) \geq \frac{4}{3}.$$

Therefore, we may assume that $\tau < \pi/2 - c/\sqrt{k}$.

We will actually prove the following relation: for all $x \in [0, \pi/2 - c/\sqrt{k}]$,

$$\int_x^{x + c/\sqrt{k}} \sin^{k-2}\theta \, d\theta \geq (1 + \Omega(1)) \int_{x - c/\sqrt{k}}^{x} \sin^{k-2}\theta \, d\theta. \tag{A.1}$$

If this relation is proven to be true, then letting $t := \left\lfloor \tau\sqrt{k}/c \right\rfloor$, writing

$$\int_0^{\tau + c/\sqrt{k}} \sin^{k-2}\theta \, d\theta = \left(\int_\tau^{\tau + c/\sqrt{k}} + \int_{\tau - c/\sqrt{k}}^{\tau} + \cdots + \int_0^{\tau - tc/\sqrt{k}}\right) \sin^{k-2}\theta \, d\theta,$$

and applying (A.1) to each term on RHS, we obtain the desired result.

In order to prove the relation, we show that, for all $x \in [0, \pi/2 - c/\sqrt{k}]$,

$$\sin^{k-2}(x + c/\sqrt{k}) \geq (1 + \Omega(1)) \cdot \sin^{k-2} x.$$

It is equivalent to showing that the function

$$f(x) := \frac{\sin^{k-2}(x + c/\sqrt{k})}{\sin^{k-2}(x)}$$

is at least $1 + \Omega(1)$ for $x \in [0, \pi/2 - c/\sqrt{k}]$. Check that, for $k \geq 3$, $f(x)$ is decreasing by differentiating $\sin(x + c/\sqrt{k})/\sin(x)$. It remains to compute

$$\begin{aligned} f(\pi/2 - c/\sqrt{k}) &= \frac{\sin^{k-2}(\pi/2)}{\sin^{k-2}(\pi/2 - c/\sqrt{k})} \\ &= [\cos(c/\sqrt{k})]^{-(k-2)} \\ &= \left(1 - \frac{c^2}{2k}\right)^{-(k-2)} + o(1) \qquad \text{(by Taylor expansion)} \\ &= \exp(c^2/2) + o_k(1). \end{aligned}$$

Therefore, $f(x) \geq 1 + \Omega(1)$ for sufficiently large k, as required. □

Proof of Proposition 4.8.7. Let $G_k = (V, E)$ be the (γ, δ)-spherical proximity graph defined in Definition 4.8.1 with embedding $u \mapsto x_u \in \mathbb{S}^{k-1}$, where $\gamma = c_1/\sqrt{k}$ and $\delta = c_2/\sqrt{k}$. We wish to construct test vectors $f : V \to \mathbb{R}^k$, such that $\sum_{u \in V} f(u) = \vec{0}$ and

$$\frac{\sum_{uv \in E} P(u,v) \left\|f(u) - f(v)\right\|^2}{\sum_{u \in V} \left\|f(u)\right\|^2}$$

is small for any doubly stochastic reweighting P of the graph G_k. Referring to the proof of Proposition 4.8.3 in the beginning of this subsection, We claim that setting $f(u) := x_u - \bar{x}$ works, where $\bar{x} := \frac{1}{n} \sum_{u \in V} x_u$. By construction, $\sum_{u \in V} f(u) = \vec{0}$. We next show that

$$\sum_{u \in V} \left\|f(u)\right\|^2 \geq \Omega(n).$$

Since $\sum_{u \in V} f(u) = \vec{0}$, by Fact 2.10.4 we have

$$\sum_{u \in V} \left\|f(u)\right\|^2 = \frac{1}{2n} \sum_{u,v \in V} \left\|f(u) - f(v)\right\|^2 = \frac{1}{2n} \sum_{u,v \in V} \left\|x_u - x_v\right\|^2.$$

It then suffices to show that, for all $u \in V$,

$$\sum_{v \in V} \left\|x_u - x_v\right\|^2 \geq \Omega(n). \tag{A.2}$$

Divide the sphere $\mathbb{S}^{k-1}$ into two halves: the half H^+ closer to x_u and the half H^- closer to the antipodal point $-x_u$. Referring to the construction in Proposition 4.8.3, since the cells S_u have volume between ε and 2ε, at most $2n/3$ cells will be completely contained in H^+. Therefore, at least $n/3$ cells S_v will have a nontrivial intersection with H^-, which implies that the corresponding chosen points x_v satisfy $d(x_u, x_v \geq \frac{\pi}{2} - \gamma \geq \Omega(1)$, and so are of Euclidean distance $\Omega(1)$ from x_u. This proves (A.2).

Since $\|f(u) - f(v)\|^2 = \|x_u - x_v\|^2 \leq d(x_u, x_v)^2 \leq \delta^2$ for any $uv \in E$, we conclude that, for any doubly stochastic reweighting P,

$$\lambda_2(I - P) \leq \frac{\sum_{uv \in E} P(u,v) \, \|f(u) - f(v)\|^2}{\sum_{u \in V} \|f(u)\|^2} \leq \frac{\sum_{u \in V} \sum_{v: uv \in E} P(u,v) \cdot \delta^2}{\Omega(n)} = O(\delta^2),$$

and hence $\lambda_2^*(G_k) \lesssim \delta^2 \lesssim 1/k$. Combining with the degree bound in Proposition 4.8.5,

$$\lambda_2^*(G_k) \lesssim \frac{1}{k} \lesssim \frac{1}{\log \Delta},$$

and the proof is complete. $\square$

Appendix B

Deferred Proofs for Chapter 6

Proof of Proposition 6.3.26. The proof largely follows Proposition 3.1.18. Given a one-dimensional solution $X(u,v) = \langle x(u), x(v) \rangle$ to the reweighted eigenvalue program $(\sigma_\mu^*)_k(F_H)$ for the directed hypergraph H, the inner maximization objective $M(X)$ is

$$\begin{aligned}
\max_{\lambda_{e,(u,v)}} \quad & \sum_{e \in E} \sum_{\substack{u \in e^- \\ v \in e^+}} \lambda_{e,(u,v)} (x(u) - x(v))^2 \\
\text{subject to} \quad & \sum_{\substack{u \in e^- \\ v \in e^+}} \lambda_{e,(u,v)} \leq 1 \qquad \forall e \in E \\
& \sum_{v \in V} \sum_{e \in E} \lambda_{e,(u,v)} = \sum_{v \in V} \sum_{e \in E} \lambda_{e,(v,u)} \qquad \forall u \in V \\
& \lambda_{e,(u,v)} \geq 0.
\end{aligned}$$

Let $y_x : V \to \mathbb{R}$ be a k-step function, taking values $t_1 < t_2 < \cdots < t_k$. Let $c \in \mathbb{R}$ be a parameter, and define $h : V \to \mathbb{R}$ so that

$$h(u) := \int_c^{f(u)} \nu(t)\, dt,$$

where $\nu(t) := \min_{i \in [k]} |t - t_i|$. We then choose c so that $\sum_{u \in V} \mu(u) h(u) = 0$. By the same argument as in Proposition 3.1.18, we have the denominator lower bound that

$$\sum_{v \in V} \mu(v) |h(v)| \gtrsim \frac{1}{k} \sum_{v \in V} \mu(v)(x(v) - c)^2 \geq \frac{1}{k} \sum_{v \in V} \mu(v) x(v)^2 \geq \frac{1}{k}.$$

As for the numerator, we have from Proposition 3.1.18 that for any $u, v \in V$,

$$|h(u) - h(v)| \leq \frac{1}{2}|x(u) - x(v)| \left(|x(u) - y_x(u)| + |x(v) - y_x(v)| + |x(u) - x(v)|\right),$$

so that for any feasible reweighting $\{\lambda_{e,(u,v)}\}$,

$$\begin{aligned}
& \sum_{e \in E} \sum_{\substack{u \in e^- \\ v \in e^+}} \lambda_{e,(u,v)} |h(u) - h(v)| \\
\leq \quad & \frac{1}{2} \sum_{e \in E} \sum_{\substack{u \in e^- \\ v \in e^+}} \lambda_{e,(u,v)} |x(u) - x(v)| \left(|x(u) - y_x(u)| + |x(v) - y_x(v)| + |x(u) - x(v)|\right) \\
\overset{(*)}{\leq} \quad & \frac{1}{2} \left[M(x) + \sqrt{2M(x) \cdot \sum_{e \in E} \sum_{u \in e^-, v \in e^+} \lambda_{e,(u,v)} \left((x(u) - y_x(u))^2 + (x(v) - y_x(v))^2\right)} \right] \\
\overset{(**)}{\lesssim} \quad & M(x) + \|x - y_x\|_\mu \cdot \sqrt{M(x)},
\end{aligned}$$

where we used Cauchy-Schwarz inequality in $(*)$ and the fact that μ is the total degree measure in $(**)$. Combining the numerator and denominator bounds yields the result. □

Proof of Proposition 6.3.27. The proof follows closely that of Proposition 3.1.19. First, the dual of the inner maximization program in the proof of Proposition 6.3.26 is

$$\begin{aligned}
\min_{\substack{g: V \to \mathbb{R}_{\geq 0} \\ r: V \to \mathbb{R}}} \quad & \sum_{e \in E} g(e) \\
\text{subject to} \quad & (x(u) - x(v))^2 \leq g(e) - r(u) + r(v) \qquad \forall e \in E, (u, v) \in (e^-, e^+),
\end{aligned}$$

and the dual of the inner maximization program in $(\sigma_\mu^*)_k(F_H)$ is the same, except that $(x(u) - x(v))^2$ is replaced by $\|f(u) - f(v)\|^2$ if $X(u, v) = \langle f(u), f(v) \rangle$.

Suppose that a feasible dual solution (x, g, r) is given and $X(u, v) = \langle x(u), x(v) \rangle$. Let $M > 0$ be a parameter to be determined later. Let $t_0 = -\infty$ and successively choose $t_1, t_2, \ldots$ so that $t_i > t_{i-1}$ is the smallest real number such that the following function

$$\bar{f}_i(u) := \begin{cases} \min\left(x(u) - t_{i-1}, t_i - x(u)\right), & \text{if } t_{i-1} < x(u) \leq t_i \\ 0, & \text{otherwise} \end{cases}$$

satisfies $\left\|\bar{f}_i\right\|_\mu^2 \geq M$. If such a t_i does not exist, we set $t_i = \infty$ and terminate the process. The process always terminates within n steps, and if it terminates with $t_{k+1} = \infty$ then the following function (which is determined once f and the t_i's are fixed)

$$y_x(u) := \arg \min_{t_i : i \in [k]} |x(u) - t_i|$$

is a k-step function. Again, the h_i's have disjoint support, and in fact

$$\sum_{i=1}^{k+1} \left\|\bar{f}_i\right\|_\mu^2 = \|x - y_x\|_\mu^2 .$$

Consider the scenario that the process does *not* terminate after k steps. That means $\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_k$ are all well-defined and each having mass $\left\|\bar{f}_i\right\|_\pi^2$ *exactly* M. We will construct from $\bar{f}_1, \ldots, \bar{f}_k$ a solution $(\bar{f}, \bar{g}, \bar{r})$ to the dual program of $(\sigma_\mu^*)_k(F)$ with small objective value. Define $\bar{f} : V \to \mathbb{R}^n$, $\bar{g} : V \to \mathbb{R}_{\geq 0}$ and $\bar{r} : V \to \mathbb{R}$ as follows:

$$\bar{f}(v) := \Big(\frac{\bar{f}_1(v)}{\sqrt{M}}, \ldots, \frac{\bar{f}_k(v)}{\sqrt{M}}, 0, \ldots, 0\Big)^T, \qquad \bar{g}(v) := \frac{1}{M} g(v), \qquad \bar{r}(v) := \frac{1}{M} r(v).$$

We will check that $(\bar{f}, \bar{g}, \bar{r})$ is a feasible solution to the dual program of $(\sigma_\mu^*)_k(F_H)$. Define $S_i := \operatorname{supp} \bar{f}_i \subseteq V$. For the sub-isotropy condition, note that each $\bar{f}(v)$ has at most one nonzero entry, and the Gram matrix $\bar{X}(u, v) = \langle \bar{f}(u), \bar{f}(v) \rangle$ satisfies

$$\begin{aligned}
& \operatorname{diag}(\mu)^{\frac{1}{2}} \bar{X} \operatorname{diag}(\mu)^{\frac{1}{2}} \\
= \; & \operatorname{diag}\Big(\frac{1}{M} \sum_{v \in S_1} \mu(u) \bar{f}_1(u)^2, \frac{1}{M} \sum_{v \in S_2} \mu(v) \bar{f}_2(u)^2, \ldots, \frac{1}{M} \sum_{v \in S_k} \mu(v) \bar{f}_k(v)^2, 0, \ldots, 0\Big) \\
= \; & \frac{1}{M} \operatorname{diag}\Big(\left\|\bar{f}_1\right\|_\mu^2, \left\|\bar{f}_2\right\|_\mu^2, \ldots, \left\|\bar{f}_k\right\|_\mu^2, 0, \ldots, 0\Big) \\
= \; & \operatorname{diag}(1, 1, \ldots, 1, 0, \ldots, 0) \preceq I_n.
\end{aligned}$$

The mass constraint is satisfied as

$$\operatorname{tr}\Big(\operatorname{diag}(\mu)^{\frac{1}{2}} \bar{X} \operatorname{diag}(\mu)^{\frac{1}{2}}\Big) = \operatorname{tr}\Big(\operatorname{diag}\big(1, 1, \ldots, 1, 0, \ldots, 0\big)\Big) = k.$$

The dual constraints are satisfied as

$$\begin{aligned}
\left\|\bar{f}(u) - \bar{f}(v)\right\|^2 &= \frac{1}{M} \sum_{i=1}^{k} (\bar{f}_i(u) - \bar{f}_i(v))^2 \leq \frac{1}{M} (x(u) - x(v))^2 \\
&\overset{(*)}{\leq} \frac{1}{M} \left(g(e) - r(u) + r(v)\right) = \bar{g}(e) - \bar{r}(u) + \bar{r}(v),
\end{aligned}$$

where $(*)$ follows the same reasoning as in the proof of Proposition 3.1.19. Therefore, $(\bar{f}, \bar{g}, \bar{r})$ is a feasible solution to the dual program of $(\sigma_\mu^*)_k(F_H)$, and its objective value is

$$\sum_{e \in E} \bar{g}(e) = \frac{1}{M} \sum_{e \in E} g(e) = \frac{M(X)}{M} \geq (\sigma_\mu^*)_k(F).$$

Choose $M = 2M(X)/(\sigma_\mu^*)_k(F_H)$ so that the above inequality *fails*. This means that the process terminates after at most k steps, and with $t_k = \infty$, which gives

$$\|x - y_x\|_\mu^2 = \sum_{i=1}^{k} \left\|\bar{f}_i\right\|_\mu^2 \leq kM \lesssim \frac{k \cdot M(X)}{(\sigma_\mu^*)_k(F_H)}.$$

This completes the proof. $\square$

Appendix C

Deferred Proofs for Chapter 7

Proof of Lemma 7.3.3. The proof basically follows that of the unweighted version in Lemma 3.5.7, and bears similarities with the ARV-type proofs such as Theorem 8.3.3. Recall that π is a distribution on V, so that $\pi(V) = 1$.

For convenience, normalize the vertex weights so that $\sum_{u,v \in V} \pi(u)\pi(v)d_s(u,v)^2 = 1$. We would like to construct a 1-Lipschitz mapping $u \mapsto f(u) \in \mathbb{R}$ such that $d_f(u,v) \leq d_s(u,v)$ for all $u, v \in V$, and

$$\sum_{u,v \in V} \pi(u)\pi(v)d_f(u,v) \gtrsim \alpha^{-2}. \tag{C.1}$$

There are two cases to consider: the "well-spread" case and the "large core" case. Define "well-spread" similarly to Lemma 8.3.2: that the π-weight of all d_s-balls

$$B_{d_s}(u, 1/\sqrt{20}) := \left\{ v \in V : d_s(u,v) \leq 1/\sqrt{20} \right\}$$

is at most $1/10$. In this case, using the definition of $\alpha = \alpha(G)$ (which involves padded decomposition in Definition 3.1.12), fix a partitioning $\mathcal{P}$ of V such that each partition in $\mathcal{P}$ has d_s-diameter at most $1/\sqrt{20}$, and that $\pi(U) \geq 1/2$ for

$$U := \left\{ u \in V : B_{d_s}\left(u, \frac{1}{\sqrt{10} \cdot \alpha}\right) \subseteq P(u) \right\},$$

where $P(u)$ denotes the (unique) partition in $\mathcal{P}$ that contains $u \in V$. Such partitioning $\mathcal{P}$ exists by Markov's inequality. Let $\tau : \mathcal{P} \to \{0, 1\}$ be a random variable that assigns i.i.d. 0/1 values to each partition in $\mathcal{P}$, and let

$$L := \{u \in V : \tau(P(u)) = 1\}$$

be a random subset. Define

$$f(u) := d_s(u, L) = \min_{v \in L} d_s(u, v).^{1}$$

Clearly this is 1-Lipschitz that $d_f(u, v) \le d_s(u, v)$. We show that

$$\mathbb{E}_f \left[\sum_{u,v \in V} \pi(u)\pi(v) d_f(u, v)^2 \right] \gtrsim \alpha^{-2},$$

which implies the result by averaging argument. For any $u \in U$, either $\tau(P(u)) = 1$ so that $f(u) = 0$, or $\tau(P(u)) = 0$ so that $f(u) \ge 1/(\sqrt{20} \cdot \alpha)$. Therefore, for any $u, v \in U$ in different partitions, there is at least $1/2$ probability that

$$d_f(u, v)^2 = |f(u) - f(v)|^2 \ge \frac{1}{20\alpha^2}.$$

By the well-spread property, for any $u \in U$ the total π-weight of vertices $v \in U$ that is not in $P(u)$ is at least $1/2 - 1/10 = 3/10$. Therefore,

$$\mathbb{E}_f \left[\sum_{u,v \in V} \pi(u)\pi(v) d_f(u, v)^2 \right] \ge \frac{1}{2} \cdot \frac{3}{10} \cdot \frac{1}{20\alpha^2} \gtrsim \alpha^{-2},$$

so such f exists that satisfies (C.1).

In the "large core" case, there exists a $u_0 \in V$ such that

$$\pi\left(B_{d_s}(u_0, 1/\sqrt{20})\right) > \frac{1}{10}.$$

Take $L := B_{d_s}(u_0, 1/\sqrt{20})$ and define $f(u) := d_s(u, L)$. Again, $d_f(u, v) \le d_s(u, v)$ for all $u, v \in V$. We prove a stronger average distortion bound that

$$\sum_{u,v \in V} \pi(u)\pi(v) d_f(u, v)^2 \ge \Omega(1).$$

Note that $\text{diam}(L)^2 \le (2/\sqrt{20})^2 = 1/5$, and so

$$\begin{aligned} 2 &= \sum_{u,v \in V} \pi(u)\pi(v) d_s(u, v)^2 \\ &\le 4 \sum_{u,v \in V} \pi(u)\pi(v) \left[d_s(u, L)^2 + \text{diam}(L)^2 + d_s(v, L)^2 \right] \\ &\le 4 \left(\frac{1}{5} + 2 \sum_{u \in V} \pi(u) d_s(u, L)^2 \right), \end{aligned}$$

[1] When $L = \emptyset$, this is not well-defined, so we just set $f \equiv 0$. This does not affect the rest of the proof.

where the first equality is by Fact 2.10.4 and the first inequality is by applying the relaxed triangle inequality $2d_s(u,v)^2 + 2d_s(v,u')^2 \geq d_s(u,u')^2$ twice. Thus,

$$\sum_{v \in V} \pi(v) f(v)^2 = \sum_{v \in V} \pi(v) d_s(v, L)^2 \geq \frac{1}{4} - \frac{1}{10} = \frac{3}{20}.$$

It follows that

$$\begin{aligned} \sum_{u,v \in V} \pi(u)\pi(v) d_f(u,v)^2 &\geq \sum_{u \in L} \sum_{v \notin L} \pi(u)\pi(v) d_f(u,v)^2 \\ &= \pi(L) \cdot \sum_{v \in V} \pi(v) f(v)^2 \\ &\geq \frac{1}{10} \cdot \frac{3}{20} \geq \Omega(1), \end{aligned}$$

which concludes the proof for the large core case. □

Proof of Lemma 7.3.4. The proof is by writing out the Lagrangian of the minimum congestion program, simplifying it to obtain $\Lambda_s(G)$ as a dual program, and lastly finding a Slater point to establish strong duality. It is modeled heavily after the proof of Lemma 3.5.9.

Introduce primal variables $c(v)$ for the congestion at vertex v and $w(p)$ for the amount of flow sent along a path p. Use $\mathcal{P}$ to denote the set of all paths on G and $\mathcal{P}(u,v)$ to denote the set of all u-v paths on G. The minimum congestion program can then be written as

$$\begin{aligned} \min_{c,w} \quad & \left(\sum_{v \in V} \frac{c(v)^2}{\pi(v)} \right)^{1/2} \\ \text{subject to} \quad & c(v) = \sum_{p \ni v} w(p) \qquad \forall v \in V \\ & w(p) \geq 0 \qquad \forall p \in \mathcal{P} \\ & \sum_{p \in \mathcal{P}(u,v)} w(p) = \pi(u)\pi(v) \qquad \forall u, v \in V, u \neq v. \end{aligned}$$

Note that the objective function is convex. Using dual variables $s(v), \mu(p), \alpha(u,v)$ for the three constraints, we obtain the Lagrangian dual program as

$$\begin{aligned} \max_{s,\mu,\alpha} \min_{c,w} \quad & \left(\sum_{v\in V} \frac{c(v)^2}{\pi(v)}\right)^{1/2} - \sum_{v\in V} s(v)\left[c(v) - \sum_{p\ni v} w(p)\right] \\ & - \sum_{p\in\mathcal{P}} \mu(p)w(p) - \sum_{u\neq v} \alpha(u,v)\left[\sum_{p\in\mathcal{P}(u,v)} w(p) - \pi(u)\pi(v)\right] \\ \text{subject to} \quad & \mu(p) \geq 0 \qquad \forall p \in \mathcal{P}. \end{aligned}$$

For fixed s, μ, α, we solve the inner minimization problem. We isolate the part relevant to c from the part relevant to w and minimize them separately. The part relevant to c is

$$\left(\sum_{v\in V} \frac{c(v)^2}{\pi(v)}\right)^{1/2} - \sum_{v\in V} s(v)c(v).$$

First derivative test yields the local minimizer condition

$$c(v) = \pi(v)s(v) \cdot \left(\sum_{v\in V} \frac{c(v)^2}{\pi(v)}\right)^{1/2} \qquad \forall v \in V,$$

and the objective becomes

$$\left(\sum_{v\in V} \frac{c(v)^2}{\pi(v)}\right)^{1/2} \left(1 - \sum_{v\in V} \pi(v)s(v)^2\right).$$

We see that the minimum value is 0 if $\sum_{v\in V} \pi(v)s(v)^2 \leq 1$ and $-\infty$ otherwise. Therefore, we may remove this from the objective of the Lagrangian dual and instead add the constraints that $\sum_{v\in V} \pi(v)s(v)^2 \leq 1$.

The part relevant to w is

$$\sum_{u,v\in V} \sum_{p\in\mathcal{P}(u,v)} w(p)\left[\sum_{v\in p} s(v) - \mu(p) - \mathbb{1}[u \neq v] \cdot \alpha(u,v)\right].$$

For the minimum value to not be $-\infty$, we need $\sum_{v\in p} s(v) - \mu(p) - \mathbb{1}[u \neq v] \cdot \alpha(u,v) = 0$ for all $p \in \mathcal{P}(u,v)$, which is equivalent to $s(v) \geq 0$ for all $v \in V$ and

$$\alpha(u,v) \leq \sum_{v\in p} s(v) \qquad \forall u \neq v, p \in \mathcal{P}(u,v).$$

Again, we add these as constraints and remove the w part from the objective of the Lagrangian. After these steps, the primal variables w and h are eliminated, μ becomes redundant, and we end up with the following maximization problem:

$$\begin{aligned}
\max_{s\geq 0,\alpha} \quad & \sum_{u\neq v} \pi(u)\pi(v)\alpha(u,v) \\
\text{subject to} \quad & \alpha(u,v) \leq \sum_{v\in p} s(v) \qquad \forall u\neq v, p\in\mathcal{P}(u,v) \\
& \sum_{v\in V} \pi(v)s(v)^2 \leq 1.
\end{aligned}$$

Clearly, the best choice of $\alpha(u,v)$ is $\alpha(u,v) = d_s(u,v)$ where $d_s(u,v)$ is the s-weighted shortest path length from u to v. Since $\Lambda_s(G)$ is homogeneous in s, we see that $\max_{s:V\to\mathbb{R}_{\geq 0}} \Lambda_s(G)$ is equivalent to the above program.

It remains to establish strong duality. It follows from the convexity of the primal objective and the existence of Slater point by taking $w(p) = \pi(u)\pi(v)/|\mathcal{P}(u,v)|$ for any $p\in\mathcal{P}(u,v)$ and $c(v) = \sum_{p\ni v} w(p)$. □

Proof of Lemma 7.4.9. The proof is again by standard Lagrangian duality and is modeled heavily after the proof of Lemma 3.5.17. Rewrite the metric spreading maximization problem as follows:

$$\begin{aligned}
\max_{\varepsilon,\delta,s} \quad & \varepsilon \\
\text{subject to} \quad & \frac{1}{\pi(A)^2}\sum_{u,v\in A} \pi(u)\pi(v)\delta(u,v) \geq \varepsilon \qquad \forall A\in\Psi \\
& \delta(u,v) \leq \sum_{u'\in p} s(u') \qquad \forall u,v\in V\ \forall p\in\mathcal{P}(u,v) \\
& s(u)\geq 0 \qquad \forall u\in V \\
& \sum_{u\in V} \pi(u)s(u)^2 = 1.
\end{aligned}$$

Here $\mathcal{P}(u,v)$ denotes the set of all paths on G from u to v and $\delta(u,v)$ represents the length of the shortest path from u to v. Introduce dual variables $h(A)$ for the first constraint, $\alpha(p)$ for the second constraint, and μ for the final constraint. We obtain the Lagrangian dual program as

$$\begin{aligned}
\min_{h,\alpha,\mu}\max_{\varepsilon,\delta,s} \quad & \varepsilon + \sum_{A\in\Psi} h(A)\left(\frac{1}{\pi(A)^2}\sum_{u,v\in A}\pi(u)\pi(v)\delta(u,v) - \varepsilon\right) \\
& + \sum_{u,v}\sum_{p\in\mathcal{P}(u,v)} \alpha(p)\left(\sum_{u'\in p} s(u') - \delta(u,v)\right) + \mu\left(1 - \sum_{u\in V}\pi(u)s(u)^2\right) \\
\text{subject to} \quad & h(A) \geq 0 \qquad \forall A\in\Psi \\
& \alpha(p) \geq 0 \qquad \forall p \in \mathcal{P} \\
& s(u) \geq 0 \qquad \forall u \in V.
\end{aligned}$$

We first solve the inner maximization problem to eliminate the primal variables, and then interpret the dual variables as subset flow parameters. For the inner maximization problem, the part involving ε is

$$\varepsilon\left(1 - \sum_{A\in\Psi} h(A)\right),$$

as ε is unconstrained, for the maximum value to not be ∞, we need $\sum_{A\in\Psi} h(A) = 1$, in which case the maximum is 0. The part involving $\delta(u,v)$ is

$$\delta(u,v)\left[\pi(u)\pi(v)\sum_{A\in\Psi:\{u,v\}\subseteq A}\frac{h(A)}{\pi(A)^2} - \sum_{p\in\mathcal{P}(u,v)}\alpha(p)\right].$$

Again, as $\delta(u,v)$ is unconstrained, for the maximum value to not be ∞, we need

$$\sum_{p\in\mathcal{P}(u,v)}\alpha(p) = \pi(u)\pi(v)\sum_{A\in\Psi:\{u,v\}\subseteq A}\frac{h(A)}{\pi(A)^2} =: D_h(u,v),$$

in which case the maximum is 0. Finally, the part involving $s(u)$ is

$$\left(\sum_{p\in\mathcal{P}:u\in p}\alpha(p)\right)\cdot s(u) - \mu\cdot\pi(u)s(u)^2.$$

Write

$$C(u) := \sum_{p\in P:u\in p}\alpha(p).$$

When $\mu < 0$ this is unbounded, and otherwise first derivative test gives the optimizer

$$s(u) = \frac{C(u)}{2\mu \cdot \pi(u)}.$$ [2]

Simplifying, the dual program becomes

$$\begin{aligned}
\min_{h,\alpha,\mu} \quad & \sum_{u\in V} C(u)\left(\frac{C(u)}{2\mu\cdot\pi(u)}\right) + \mu\left(1 - \sum_{u\in V}\pi(u)\left(\frac{C(u)}{2\mu\cdot\pi(u)}\right)^2\right) \\
\text{subject to} \quad & h(A) \geq 0 \qquad \forall A \in \Psi \\
& \sum_{A\in\Psi} h(A) = 1 \\
& \alpha(p) \geq 0 \qquad \forall p \in \mathcal{P} \\
& \sum_{p\in\mathcal{P}(u,v)} \alpha(p) = D_h(u,v) \qquad \forall u,v \in V.
\end{aligned}$$

The objective is

$$\mu + \frac{1}{4\mu}\sum_{u\in V}\frac{C(u)^2}{\pi(u)}$$

and by first derivative test the minimizer is $\mu = \sqrt{\sum_{u\in V} C(u)^2/\pi(u)}/2$, attaining a minimum value of 2μ.

By now the subset flow interpretation should be clear: $h : \Psi \to \mathbb{R}_{\geq 0}$ is a distribution on Ψ, $D_h(u,v)$ is the corresponding demand between u and v in the subset flow problem, $\alpha(p)$ is the amount of flow along path p in the flow solution, which we denote by F. Then, the objective is exactly $con_\pi(F)$, and the above constraints can be condensed into $F \in \mathcal{F}_\pi^\Psi(G)$. Therefore, the Lagrangian dual of the metric spreading maximization problem is

$$\min_{F\in\mathcal{F}_\pi^\Psi(G)} con_\pi(F).$$

It remains to establish strong duality. Clearly, the primal maximization problem is concave. We can find a Slater point as follows: take $s(u) = 1$ for all $u \in V$, $\delta(u,v) = 1$ for all $u, v \in V$, and $\varepsilon = 1/10$. This is feasible and all inequality constraints are strict. This completes the proof. □

[2] We treat $0/0 = 0$ here.

Appendix D

Deferred Proofs for Chapter 8

Proof of Proposition 8.3.1. Let $\emptyset \neq S \subset V$. We construct an SDP solution to show that $\xi(G) \leq 2\vec{\phi}_\pi(S)$. Consider the vector solution

$$v_i := \begin{cases} (a, 0, \ldots, 0), & \text{if } i \in S, \\ (b, 0, \ldots, 0), & \text{otherwise }. \end{cases}$$

where $a, b \in \mathbb{R}$ satisfies $a\pi(S) + b\pi(S^c) = 0$ and $a^2\pi(S) + b^2\pi(S^c) = 1$. Note that such (a, b) must exist. A routine check reveals that all the constraints on v_i are satisfied. It remains to show that

$$\frac{1}{2}\sum_{uv \in E} P(u,v)\, \|f(u) - f(v)\|^2 \leq 2\vec{\phi}_\pi(S) \qquad \forall F \in \mathcal{P}(G).$$

Solving for a and b, we see that $(a-b)^2 = \pi(V)/\pi(S)\pi(S^c)$. Then,

$$\begin{aligned} \frac{1}{2}\sum_{uv \in E} P(u,v)\, \|f(u) - f(v)\|^2 &= \frac{1}{2}\left[\sum_{u \in S, v \in S^c} + \sum_{u \in S^c, v \in S}\right] P(u,v)(a-b)^2 \\ &= (P(S,S^c) + P(S^c,S)) \cdot \frac{\pi(V)}{2\pi(S)\pi(S^c)} \\ &\leq \frac{P(S,S^c) + P(S^c,S)}{\min(\pi(S),\pi(S^c))} \\ &\leq \frac{2\min(\delta^+(S),\delta^+(S^c))}{\min(\pi(S),\pi(S^c))} = 2\vec{\phi}_\pi(S), \end{aligned}$$

where the last inequality uses the fact that $P \in \mathcal{F}(G)$ is an Eulerian reweighting, so that $P(S, S^c) = P(S^c, S) \leq \min(w(\delta^+(S)), w(\delta^+(S^c)))$. This finishes the proof that $\lambda_\pi^\Delta(G) \leq 2\vec{\phi}_\pi(G)$. □

Proof of Lemma 8.3.2. The algorithm for the unweighted version in Theorem 3.6.2 proceeds as follows. Let $\sigma > 0$ be a suitable absolute constant. First, choose a random direction $w \sim \mathbb{S}^{n-1} \subseteq \mathbb{R}^n$, and order the vertices u by $\langle f(u), w\rangle$. Second, if the median value is M, set L to be the set of vertices u such that $\langle f(u), w\rangle \geq M + \sigma/\sqrt{n}$, and set R to be the set of vertices u such that $\langle f(u), w\rangle < M$. Third, while there are pairs $(u, v) \in L \times R$ such that $\|f(u) - f(v)\|^2 < \Delta = \Theta(1/\sqrt{\log n})$, remove u from L and v from R. If $|L| \geq \Omega(n)$ and $|R| \geq \Omega(n)$ at the end, the procedure successfully finds two large subsets that are at least $\Delta \geq \Omega(1/\sqrt{\log n})$ apart in ℓ_2^2 distance. Refer to [ARV09] for complete details.

To prove the π-weighted version in Lemma 8.3.2, we do a reduction to the unweighted case. Recall the assumption that $\pi(V) = 1$. Let $K \in \mathbb{N}$ such that $K \cdot \min_{u\in V} \pi(u) \geq 1/2$, and let $\pi'(u) := \lceil K\pi(u)\rceil$ for $u \in V$. We may further assume that $\min_{u\in V} \pi(u) \geq \Omega(1/\operatorname{poly}(n))$ (vertices with smaller measure may be ignored), so that $K \leq O(\operatorname{poly}(n))$. Create $\pi'(u)$ copies of $f(u)$ and feed the embedding to the unweighted algorithm. Note that the embedding consists of $\Theta(K)$ vectors. In the end of the unweighted algorithm, w.h.p. the output sets L and R will have size $\Theta(K)$ each, and they will be at least $\Omega(1/\sqrt{\log n})$ apart in ℓ_2^2 distance.

Note that if one copy of $f(u)$ is in either of the output set, we may include all copies of $f(u)$ in that output set, without affecting the distance between L and R. Then, the π-measure of vertices in L will be at least

$$\sum_{u\in L} \pi(u) \geq \sum_{u\in L} \frac{\pi'(u)}{2K} = |L|/2K \geq \Omega(1);$$

same for R. We have proved that w.h.p. $\pi(L), \pi(R) \geq \Omega(1)$. The runtime is polynomial in the number of vectors which is $\Theta(K)$, and hence polynomial in n.

To get rid of the K-dependence in the runtime, we may modify the unweighted algorithm as follows: In the second step, compute the weighted median. In the third step, instead of removing both vertices u and v, subtract $\min(\pi(u), \pi(v))$ from both $\pi(u)$ and $\pi(v)$, and remove the vertex whose π-measure drops to zero. □

Proof of Proposition 8.5.2. The proof is standard. Given k disjoint subsets $S_1, \ldots, S_k$, define the following solution to the $\lambda_k^{\Delta}(H)$ program:

$$f(u)_i := \begin{cases} \frac{1}{\sqrt{\pi(S_i)}}, & \text{if } i \leq k \text{ and } u \in S_i; \\ 0, & \text{otherwise,} \end{cases} \qquad \text{and} \qquad g(e) := \max_{u,v \in e} \|f(u) - f(v)\|^2 .$$

Here $f(u)_i$ denotes the i-th coordinate of $f(u)$. It is routine to check that (f, g) is a feasible solution. Its objective value is

$$\begin{aligned} \frac{1}{k} \sum_{e \in E} w(e) g(e) &= \frac{1}{k} \sum_{e \in E} w(e) \max_{u,v \in e} \|f(u) - f(v)\|^2 \\ &\leq \frac{1}{k} \sum_{e \in E} w(e) \cdot \left[\sum_{i \in [k]} \max_{u,v \in e} (f(u)_i - f(v)_i)^2 \right] \\ &= \frac{1}{k} \sum_{i \in [k]} \sum_{e \in \delta(S_i)} \frac{w(e)}{\pi(S_i)} \\ &= \frac{1}{k} \sum_{i \in [k]} \phi(S_i) \\ &\leq \max_{i \in [k]} \phi(S_i). \end{aligned}$$

This completes the proof. □