

Evolution of Blockchain Technology: Dynamics of Knowledge Flow and Thematic Trends

by

Farzaneh Mardukh Rouhani

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Management Sciences

Waterloo, Ontario, Canada, 2024

© Farzaneh Mardukh Rouhani 2024

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

Blockchain technology has emerged as a transformative force across various industries, prompting significant attention and research efforts. This study examines the evolution and landscape of blockchain technology by analyzing the U.S. patent database. In particular, it explores knowledge flow in blockchain technology by incorporating various methodologies such as LDA topic modeling, citation network analysis and main path analysis. The categorization of blockchain technology into topics was conducted using the LDA machine learning model, which can provide more objective and robust topic classification than traditional subjective methods. Using this classification, this thesis explores the relationships between topics and analyzes the evolution process of each topic via network analysis. Based on the topic modeling, this paper mainly focuses on three predominant themes within blockchain research: data management, network security and privacy, and smart contracts.

This thesis first shows that specific themes are concentrated in certain geographic regions, with US inventors prominent in smart contracts, Chinese inventors in data management, and European inventors in network security and privacy. This thesis also observes distinct dynamics between established companies and new entrants across different thematic domains. New entrants initiate knowledge flows, and established companies leverage their expertise to consolidate and advance innovations, especially in the smart contract domain. In terms of assignees' role and position in the knowledge flow system, it shows that knowledge exchange hubs predominantly consist of established companies which leverage their experience and resources to shape industry standards, while new entrants introduce fresh perspectives and innovative approaches by challenging existing paradigms. These results offer valuable insights into the blockchain ecosystem, enabling both established firms and new entrants to strategically navigate the competitive market environments. The insights gained from this thesis can help companies enhance their R&D strategies, foster innovation, and maintain or attain technological leadership in the rapidly evolving blockchain market.

Acknowledgments

I am deeply thankful to my husband for his boundless love, constant support, and understanding throughout this journey. His patience and belief in me have been instrumental in the completion of this thesis.

I would also like to express my sincere gratitude to my supervisor Dr. Bon Koo, for the invaluable guidance and support throughout my thesis process. His expertise and insightful feedback have greatly enriched my work.

Dedication

In memory of Mahsa Zhina Amini, whose departure from this life has left an indelible mark and sparked a powerful movement for justice and freedom. To the brave Iranian girls who stand strong and resilient in the pursuit of their rights and dignity. Their strength and determination are a beacon of hope for all. This thesis is dedicated to their unwavering spirit and the enduring fight for equality and human rights.

Table of Contents

Author's Declaration.....	ii
Abstract.....	iii
Acknowledgments.....	iv
Dedication.....	v
List of Figures.....	vii
List of Abbreviations.....	viii
Chapter 1 Introduction.....	1
1.1 Background and Context.....	1
1.2 Research Objectives.....	2
1.3 Thesis Structure.....	4
Chapter 2 Literature Review.....	5
2.1 Patent Analysis and Technology Trajectories.....	5
2.1.1 Patent Co-citation Network.....	6
2.1.2 Main path analysis.....	8
2.2 Topic Modeling by Machine Learning.....	9
Chapter 3 Data and Method.....	13
3.1 Research Design.....	13
3.2 Phase 1: Data Collection.....	14
3.3 Phase 2: Text analysis (Topic modeling).....	15
3.4 Phase 3: Network Analysis.....	17
Chapter 4 Result.....	22
4.1 Dominant themes and their evolutionary trajectory.....	22
4.1.1 Topic 1- Decentralized Data Management and Storage.....	22
4.1.2 Topic 2- Smart Contract Execution.....	23
4.1.3 Topic 3- Network Security and User Authentication.....	24
4.2 Knowledge flow among dominant themes.....	26
4.2.1 Knowledge flow from data management.....	26
4.2.2 Knowledge flow from smart contract.....	28
4.2.3 Knowledge flow from network security.....	30
4.3 Assignees' positions and capabilities.....	33
Chapter 5 Discussion.....	36
5.1 Interpretation of Results.....	36
5.2 Theoretical implications.....	41
5.3 Practical implication.....	42
Chapter 6 Conclusion.....	44
6.1 Summary of Key Findings.....	44
6.2 Limitations of the Study.....	46
6.3 Future Recommendations.....	47
References.....	48
Appendix A Topic Modeling's Result.....	54

List of Figures

Figure 3.1 Research model	13
Figure 4.1 Main path analysis for topic 1	23
Figure 4.2 Main path analysis for topic 2	24
Figure 4.3 Main path analysis for topic 3	25
Figure 4.4 knowledge flow model analysis	26
Figure 4.5 knowledge flow from data management to smart contract	27
Figure 4.6 knowledge flow from data management to network security and privacy	28
Figure 4.7 knowledge flow from smart contract to network security and privacy	29
Figure 4.8 knowledge flow from smart contract to data management	30
Figure 4.9 knowledge flow from network security and privacy to smart contract	31
Figure 4.10 knowledge flow from network security and privacy to data management.....	32
Figure 4.11 Assignees' positions based on network citation attributes	34

List of Abbreviations

BC	Betweenness Centrality
LDA	Latent Dirichlet Allocation
MPA	Main Path Analysis
O-I Index	Outward-Inward Index

Chapter 1 Introduction

1.1 Background and Context

In recent years, blockchain technology has emerged as a disruptive innovation that is reshaping industries across the globe. Initially recognized as the foundational technology powering cryptocurrencies such as Bitcoin (Nakamoto 2009), blockchain has transcended its origins to become a multifaceted platform with far-reaching implications (Yu & Sheng, 2020). Its core promise lies in enabling decentralized and secure systems that fundamentally alter traditional approaches to transactions and data management. Beyond financial applications, blockchain technology has found extensive utility in diverse sectors including supply chain management, healthcare, and even voting systems (Bamakan et al., 2021; Perdana et al., 2021; Yu & Sheng, 2020). This versatility stems from its inherent characteristics of transparency, immutability, and distributed consensus, which not only enhance operational efficiency but also foster greater trust and reliability in digital interactions. Sim & Jang (2023) demonstrate that there is a positive correlation between a firm's blockchain patents and its financial performance and overall value. As blockchain continues to evolve, its potential to revolutionize various facets of modern society remains a compelling force in driving innovation and transforming multiple industries.

It is crucial to understand blockchain technology, its current status, and its potential impact on companies and markets. Several studies examine how blockchain technology is evolving, and other studies tackle the challenges blockchain is faced in its development. Scalability, security, privacy, regulatory compliance, energy consumption, user experience, and consensus are listed among the most important challenges for blockchain development (Bamakan et al., 2021; Yli-Huomo et al., 2016).

Several scholars analyze the current status of blockchain using publication and patent databases. Perdana et al. (2021) utilize journal article text sentiment analysis to understand societal resistance or support for blockchain as a disruptive technology. Hughes et al. (2019) explore barriers to blockchain adoption, such as trust issues, within the context of information systems (IS) and information management (IM) literature. Allen (2020) discusses blockchain as an institutional technology and examines how public policies and regulations may either hinder or facilitate blockchain adoption at the institutional level. Bamakan et al. (2021) conclude that blockchain has reached the growth phase in the United States, while Yu and Pan (2021a) recognize China and the United States as key players in the blockchain technology.

Given the path dependency nature of the technology, some studies focused on uncovering the knowledge flow path to identify the phases and development trajectories of blockchain technology. These studies assess technological trends and trajectories from various perspectives. For instance, Cavalheiro & Cavalheiro (2022) focus on the analysis of Bitcoin patent bibliographies. Other research has been explored by Bhatt et al. (2023) and Kim & Hyun (2023) to adopt a more detailed and comprehensive perspective by dividing blockchain into cryptocurrencies and distributed technologies, and discuss their respective trajectories, knowledge network attributes, and trajectory compatibility with the overall blockchain development path. In another study, the evolution path of blockchain applications across various industries, including the Internet of Things (IoT), healthcare, energy, voting, insurance, and supply chain management, has been investigated to gain a better understanding of blockchain's progress in each sector (Chang & Chen, 2020; S. Kim et al., 2021; Yu & Sheng, 2020). Furthermore, the knowledge spillover effect of emerging technologies like blockchain on other sectors has been studied to inform decision-making regarding technology investment and policy development (Someda et al., 2022).

1.2 Research Objectives

There still exist gap in the existing literature in that they do not thoroughly investigate the intricate connections among crucial themes and topics in blockchain research. They typically present generalized views on various facets of blockchain, including its technological features,

diverse applications, and technology forecasting. However, they lack a detailed examination of how specific blockchain topics develop over time, the dynamics of knowledge exchange within and between these topics, and their interactions. They do not examine how the types of companies, established firms, or new entrants, contribute to, and shape the flow of knowledge within the blockchain ecosystem. This gap in existing studies leads to an incomplete understanding of the complex and dynamic nature of blockchain research.

This thesis aims to address these gaps by exploring the following key questions. First, it determines how many distinct dominant topics blockchain technology can be classified into, and what are the insights of their evolutionary trajectory. Second, it investigates how the flow of knowledge among these topics influences and stimulates each other's evolution. Lastly, it evaluates patent assignees' relative positions and capabilities in blockchain domains. This thesis utilizes the U.S. patent data which are filed between 2017 and the end of 2023 and selects the patents that are related to blockchain technology using keywords.

This thesis employs three steps to address research objectives. To address the evolution of blockchain technology, it applies a machine learning approach to cluster text and conducts topic modeling using the LDA model. This method allows topics to be objectively categorized based on textual similarities, thereby mitigating potential biases that arise from relying solely on the researcher's knowledge. This categorization provides a foundation for subsequent analysis of each identified topic.

The next step focuses on constructing co-citation networks and performing main path analysis within and across these identified topics. This approach reveals how each thematic area within blockchain technology has evolved over time and how ideas and innovations within these topics have influenced each another. Moreover, it distinguishes between established companies and new entrants in blockchain innovation and allows for understanding of the dynamics of innovation within the blockchain ecosystem and assessing whether the development trajectory aligns with theories of disruptive innovation.

The final step encompasses the quantitative assessment of two key metrics within the assignees' co-citation networks: betweenness centrality and O-I index. These metrics provide insights into the strategic positioning and roles of each entity in facilitating knowledge dissemination and driving innovation within the blockchain domain.

This study demonstrates that blockchain technology can be classified into three primary domains or topics: decentralized data management, smart contracts, and network security and privacy. It identifies distinct geographic hubs of each topic, highlighting US dominance in smart contracts, Chinese proficiency in data management, and European advancements in network security and privacy. This study also identifies the dynamics between established companies and new entrants across these domains, discerning notable differences between their contributions. Established firms draw on their experience to influence industry standards, whereas newcomers introduce novel perspectives and innovative strategies, thereby enhancing competitiveness in the dynamic blockchain landscape.

The methodological steps in this study address the core research questions and provide a better understanding of how blockchain technology has developed and evolved. By shedding light on the intricate dynamics of knowledge flow, innovation networks, and the roles of different stakeholders, this study contributes valuable insights to both academics and practitioners within the rapidly evolving field of blockchain technology.

1.3 Thesis Structure

The remainder of the thesis is organized as follows: Chapter 2 explains the concept of knowledge flow, patent analysis, and LDA topic modeling. In Chapter 3, a detailed explanation of the research design, data selection criteria, and methodology is presented, outlining the steps taken to conduct the analysis. Chapter 4 provides empirical analysis process and presents the results. Chapter 5 explains the interpretations of the results, and their practical implications. Chapter 6 summarizes the key findings and comments the limitation and recommendations for future studies.

Chapter 2 Literature Review

2.1 Patent Analysis and Technology Trajectories

Technological evolution is intricately linked to the concept of technology trajectories and knowledge flow, which are fundamental to understanding how innovations develop and diffuse across different sectors. The concept of technology trajectory was first developed by Dosi (1982) who emphasizes that technological change is not random but rather follows specific paths shaped by a combination of scientific principles, accumulated knowledge, and socio-economic factors. This concept helps in understanding how technologies evolve, compete, and sometimes converge. Given the path-dependent nature of technology, proactive monitoring of past and present technological trends is essential for managers and decision-makers to anticipate future technological trajectories and mitigate the risk of organizational failure (Arthur, 2007).

Various methods have been utilized in the literature to identify technological trajectories, including systematic literature reviews and other qualitative approaches, as well as quantitative techniques like text-based analysis and bibliometric analysis. The utilization of patent data in trajectory and landscape analysis has emerged as a key strategy for researchers and organizations aiming to stay up to date of emerging technologies and innovation trends (Abbas et al., 2014; Altuntas et al., 2015). Indeed, patents are considered the tangible outcomes of research and invention, encapsulating significant scientific, technical, and conceptual advancements within specific fields. (Adamuthe & Thampi, 2019; Altuntas et al., 2015). Trappey et al. (2011) utilize Chinese patent data to forecast trends in radio-frequency identification (RFID) technology, providing stakeholders with essential information for strategic decision-making. Similarly, Wang et al. (2015) explore the technology life cycles and market dynamics within the optical disc industry through patent analysis, offering insights into the interplay between market demand and technological advancements.

Building upon this literature, several studies have utilized forward citations as a measure of technological advancement and evolution across diverse domains. For instance, Schmid (2018) investigates the advancement of military technologies, while Brown and Campi (2015) explore technological evolution within the oil and gas industry. Furthermore, Lanjouw

and Schankerman (2004) devise an index of patent quality based on citation frequency and show that patents receiving more frequent citations contribute significantly to societal value by serving as building blocks for future innovations a positively influence the stock market value of firms. Hagedoorn and Cloudt (2003) use patent citation analysis to assess innovation performance in companies operating in high-tech sectors, and Kyebambe et al. (2017) developed an innovative algorithm using patent citation techniques to forecast emerging technologies.

Several studies adopt the text-based analysis. For instance, the application of co-word analysis to knowledge management (KM) research using the Web of Science articles has facilitated the identification of evolving trends and advancements in the field. This approach, augmented by the integration of local and global citation scores, enables researchers to gain deeper insights into the trajectory of KM research and its broader impact (Mehri et al., 2014). In addition, several studies use both co-word analysis and local and global citation scores to discover the interdisciplinary nature and dominant topics in surgical robots via a database of related articles (Shen et al., 2019). Building upon this literature, scholars emphasise the significance of understanding how knowledge flows propagate across diverse networks and contexts, shaping innovation trajectories and economic landscapes (Duguet & MacGarvie, 2005; Singh, 2005).

2.1.1 Patent Co-citation Network

The measurement of knowledge flow is generally divided into two categories: qualitative and quantitative. Qualitative measurements aim to explore the depth and richness of knowledge exchanged through methods such as interviews, case studies, and surveys (S. Kim et al., 2003; Siu & Bao, 2008; Williams, 2011). Scholars have extensively investigated the strength and weakness of network ties within qualitative research, examining how they facilitate the spread of knowledge throughout networks (Jack, 2005). Within the quantitative category, several studies employ regression analysis, and other studies utilize citation network analysis (Fernández et al., 2022; Noailly & Shestalova, 2017; Singh, 2005). In social sciences, network theories provide a powerful framework for understanding the complex interactions and interdependencies among actors within a network. By examining individual roles, relationship

structures, and systemic behaviors, perceptions, and attitudes, network analysis offers a comprehensive overview of the dynamics driving social systems (Acemoglu & Akcigit, 2012; Aghion & Howitt, 1992; Grossman & Helpman, 2018).

Network analysis serves as a valuable tool for studying the intricate relationships underlying knowledge spillovers between companies. By examining network properties such as nodes, ties, density, and cluster coefficients, scholars gain valuable insights into the mechanisms underpinning knowledge diffusion. Researchers can gain insights into how these connections within the knowledge spillover network are utilized and activated to drive various business activities (Jack, 2005; C. C. Wang et al., 2017). Furthermore, network analysis enables the examination of the broader structure of international technology diffusion, allowing for the measurement of key metrics such as the degree, closeness, and betweenness centralities of nodes within the network (Shih & Chang, 2009).

Xiang et al. (2013) investigate international knowledge flow networks within the power system domain for Chinese assignees, offering valuable insights into the evolving dynamics of knowledge exchange across geographical and organizational boundaries. Gilsing et al. (2016) explore the relationship between a firm's betweenness and closeness centrality positions and its alliance portfolio size and strategic choices using a firm-level citation network. Scholars also study firms' innovation potential through the lens of network analysis across diverse industries, including pharmaceuticals, chemicals, and automotive sectors. By employing metrics such as degree centrality and network density, they gauge the prominence of firms within networks and their capacity to leverage network connections for innovation. Central nodes within networks wield considerable influence, serving as conduits for knowledge flow and facilitating collaboration among disparate actors. On the other hand, less influential peripheral nodes may serve as bridges between distinct network clusters, enabling the dissemination of novel insights across boundaries (Gilsing et al., 2008). In another study, closeness and betweenness centralities are used to assess knowledge flow and identify key players and hubs within the knowledge transfer landscape (Shih & Chang, 2009). He et al. (2009) employs degree centrality as a critical indicator to assess the performance of innovation in the inventor network between two US regions. In summary, network analysis offers a robust framework for understanding knowledge flow within and across industries. By employing

metrics such as degree, closeness, and betweenness centralities, researchers can identify key players and influential nodes that drive innovation and collaboration.

2.1.2 Main path analysis

Regarding the path dependencies of technologies, a robust methodology is needed to understand their trajectory and evolution. Main path analysis (MPA) is a powerful tool for discovering technological advancements and mapping the flow of knowledge (Liu et al., 2019). This method was first introduced by researchers examining the citation networks of DNA-related articles, with the aim of simplifying the complex citation networks and identifying the most significant developmental paths in DNA theory (Hummon & Dereian, 1989).

The novelty of MPA has attracted considerable scholarly attention, leading to its application in various citation contexts, including patent citations. Over time, MPA has undergone significant evolution in terms of identifying optimal algorithms, refining citation weighting methods, and developing variations of main paths (Batagelj, 2003; Liu & Lu, 2012; Yu & Pan, 2021b, 2021a). For example, Liu and Lu (2012) propose several types of main path analyses, such as backward and forward local paths, global main paths, and key-route paths, each designed to serve specific functions and purposes. Batagelj (2003) introduces a novel algorithm for citation weighting that is characterized by its simplicity, speed, and efficiency, which has subsequently been widely adopted. These advancements illustrate the dynamic nature of MPA and its increasing sophistication in capturing the complexities of technological and knowledge development.

Main Path Analysis (MPA) has been used to study technological evolution and trajectories in diverse industries. For instance, Liang et al. (2016) use MPT to identify the major research paths and developmental stages of IT outsourcing over time, and Ho et al. (2014) apply MPA and regression analysis to examine the impact of companies' networks on the evolution of fuel cells. In another investigation, the core competencies of companies and the flow of knowledge were assessed through MPA and network citation analysis to uncover opportunities and suggest future investments in thin-film solar technology (Lai et al., 2021).

The development of autonomous automobiles has also been analyzed using MPA and cross-citation analysis to determine strategic positions for businesses to adapt to the rapid advancements in this field (Cho et al., 2021).

Similarly, blockchain researchers have employed MPA for various purposes, including mapping trajectories and enhancing the methodology itself. In their comprehensive study, Yu & Sheng (2020) analyze all blockchain-related articles to construct global main paths, backward and forward paths, and key-route paths, thereby identifying trends, prominent authors, and active journals in the blockchain domain and its application in different industries like supply chain and health. Also, key-route main path analysis has been used to discover probable blockchain categories and their development phases in the subfields (Tseng & Ting, 2013). Bhatt et al. (2023) assessed the similarity between concepts in technological innovation within a disruptive innovation theory framework using MPA. Furthermore, another study explored the main path analysis of blockchain patents from the United States and China to identify structural similarities and differences in their processes, providing insights for future research directions (Yu & Pan, 2021b).

However, while there have been several studies on blockchain development and trajectories aimed at identifying the backbone structure, none have delved deeply into the blockchain thematic domain. These studies have not thoroughly explored the intricate relationships between various topics within blockchain technology. Consequently, there remains a gap in understanding the detailed evolutionary path of blockchain. A more comprehensive analysis that maps out the main paths and interconnections between these subdomains is necessary to provide a nuanced understanding of blockchain's development and future directions.

2.2 Topic Modeling by Machine Learning

Within the realm of patent analysis, the extraction of knowledge and recognizing patterns from patent database is facilitated by text mining methodologies which offer valuable insights for technology investment decision-making (Abbas et al., 2014; Choi & Song, 2018). Clustering emerges as a potent text mining technique which functions within an unsupervised framework

to organize similar objects into coherent groups characterized by intra-group cohesion and inter-group differentiation (Chen & Hu, 2006). The clustering process unfolds through a series of iterative steps, including data aggregation, establishment of similarity measures, method selection, performance assessment, and interpretation of clustering results, as elucidated by Dang and Ahmad (Dang & Ahmad, 2014).

Compared to existing text-based clustering methods, topic modeling approach offers a higher degree of automation and captures the R&D status of the target field through comprehensive data analysis. Furthermore, topic modeling emerges as a cutting-edge approach in semantic-based mining methods for elucidating constructs within large datasets, such as patent databases or journal papers. Topic modeling has emerged as a favored method among researchers due to its efficacy in analyzing large, unstructured document collections. This approach offers significant performance and benefits in uncovering patterns and trends within vast datasets (Griffiths & Steyvers, 2004; Welbers, 2015). Topic modeling analyzes distributions of words and topics, where researchers can gain insights into the underlying themes and patterns present in large datasets. Studies by Nikolenko et al. (2017) and Welbers (2015) demonstrate the effectiveness of topic modeling in elucidating complex textual data.

In the fields of machine learning and statistics, advanced topic modeling techniques play a crucial role in analyzing large document collections. Among these techniques, Latent Dirichlet Allocation (LDA) stands out as a highly trusted and adopted method. Introduced by Blei et al. (2003), LDA is a generative probabilistic model capable of automatically extracting latent topics from word distributions in documents without the need for predefined labels. This unsupervised approach assumes that documents contain multiple topics distributed across a fixed vocabulary. By leveraging distributions of topics and topic words following a Dirichlet distribution, LDA generates lists of weighted topics and topic words, facilitating interpretation and labeling by researchers (Campbell et al., 2015; Dang & Ahmad, 2014). One of the main features of this method is that it bypasses the need for predefined labels or human-classified training data (C. Lee et al., 2017).

Researchers have integrated LDA with various machine learning techniques in several studies. For example, Mo et al. (2015) employed LDA topic modeling alongside Support

Vector Machine (SVM) models for automatic term recognition (ATR), and Venugopalan and Rai (2015) applied LDA topic modeling to analyze solar photovoltaics. Additionally, in research concerning non-petroleum-based fuel vehicles, Hong Kim et al. (2019) utilized the k-means method in conjunction with LDA. LDA is also combined with various research concepts, making it a versatile technique with applications across multiple domains. For instance, Yang et al. (2014) applied an LDA method with amino acid sequence features to predict the type III secretion system (T3SS) in medicine, and Qi et al. (2018) employed mLDA and mHLDA methods to identify driving styles from real-world driving behavior data. In the realm of behavioral sciences, LDA topic modeling has been instrumental in uncovering patterns related to eating disorders (Hwang et al., 2020), gender differences in online technology communities (Sun et al., 2020), and abstract concepts like research design (Nie & Sun, 2017). In addition, LDA has been utilized to analyze social media data, focusing on topics such as Covid-19 and earthquakes (Eligüz el al., 2020; Kwok et al., 2021).

Furthermore, LDA has been pivotal in trend recognition and forecasting across various sectors. Its applications range from identifying trends in country-specific policy research to predicting technology convergence within industries like medicine and transportation (Ghaffari et al., 2023; H. Kim et al., 2018; S. Lee et al., 2015). Blockchain research has also benefitted from LDA, with studies employing it alone or in combination with the technology life cycle to uncover emerging trends (Bamakan et al., 2021; Zhang et al., 2021). Dynamic LDA Topic Modeling (DTM) has been particularly effective in studying the evolution of topics over different time periods (J. Lee et al., 2023). This functionality underscores LDA's versatility and its significant role in advancing knowledge discovery and trend analysis across diverse fields of research.

While previous research use patent data to explore various aspects of blockchain technology, such as its technological evolution, common challenges, geographical distribution, and innovation trends, this thesis uses different methodology and addresses unanswered questions regarding blockchain knowledge flow networks. Specifically, we undertake a thorough examination of the knowledge flow within blockchain technology using a

combination of text-based machine learning clustering, co-citation network analysis, and main path analysis. This multifaceted analytical approach enables a better understanding of how knowledge in blockchain technology evolves over time. It provides deeper insights into the technological knowledge encapsulated within patent data, as well as the dynamics among companies and key stakeholders in terms of knowledge sharing and innovation production within the field.

Chapter 3 Data and Method

3.1 Research Design

The research methodology is structured into three phases as shown in Figure 3.1, Phase 1 focuses on data collection and preparation which entails the initial gathering of patent data related to blockchain technology from the USPTO database. It first identifies relevant patents using CPC classification and keywords (Clarke et al., 2020), and then analyzes the patent abstracts and titles to extract relevant keywords. Subsequently, it undertakes data cleaning to ensure the dataset's quality and readiness for subsequent analysis.

Phase 2 involves the LDA topic modeling. It begins with the identification of customized stopwords relevant to blockchain and then executes the machine learning model under various configurations to ascertain the optimal number of topics for clustering the blockchain patents. After determining the most appropriate number of topics, this phase applies the LDA topic modeling to the patent titles and abstracts to reveal underlying themes and topics.

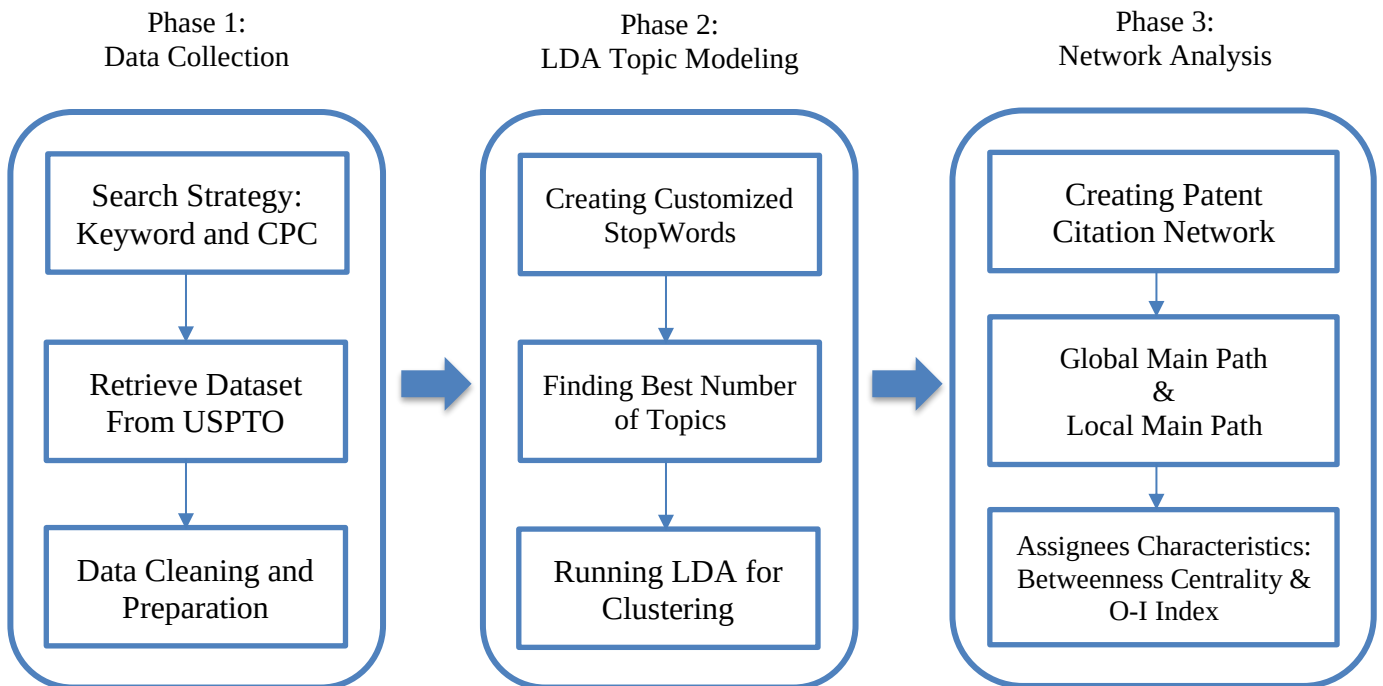


Figure 3.1 Research model

The knowledge flow analysis in phase 3 involves examining the knowledge flow pattern within each topic and the interrelations among them. It constructs a network-based patent citations to evaluate the knowledge flow, and then creates both global and local paths to identify significant pathways, patents, and assignees. In addition, it calculates assignee characteristics such as the betweenness centrality and the I-O index to analyze the role and position of assignees within each topic. These structured methodological steps help provide a systematic and rigorous approach to analyze the research questions.

3.2 Phase 1: Data Collection

The data collection process involves several sequential steps. The patent data are collected from the Patentsview website¹ which store all patents granted by the United States Patent and Trademark Office (USPTO) from 2017 to the end of 2023. To collect the patents which are related to blockchain technology, this study utilizes Clarke et al. (2020) method which using both search queries and CPC codes. This method proposes a search query which incorporates collaborative input from blockchain experts and patent examiners at the European Patent Office. The set of keywords were meticulously evaluated against the patent database to discern the primary, secondary, and excluded terms. Evaluation criteria focuses on identifying false positives resulting from the alignment of these keywords with the classification codes recommended by the examiners. This study proposes a set of keywords and CPC codes (G06 and H04) that are considered highly relevant to blockchain patents. The following list of keywords is utilized for retrieving blockchain patents based on titles and abstracts.

List of selected key words:

'blockchain', 'bitcoin', 'bit-coin', 'block-chain', 'blocksign', 'codius', 'colored coin', 'colored-coin', 'crypto currency', 'crypto-currency', 'cryptocurrency', 'distributed ledger', 'distributed-ledger', 'dogecoin', 'doge-coin', 'ethereum', 'factom', 'litecoin', 'lite-coin', 'pay-to-script-hash', 'p2sh', 'proof of stake', 'proof-of-stake', 'sidechain', 'smart contract', 'smart-contract'

Based on the above keywords, a total of 5,760 patents were detected, of which 3,810 were cited and used in this research.

¹ patentsview.org

3.3 Phase 2: Text analysis (Topic modeling)

Latent Dirichlet Allocation (LDA) is an unsupervised learning algorithm employed for analyzing textual data. The term "Latent" refers to the hidden or underlying structures within the data and the term "Dirichlet" refers to the Dirichlet distribution, a family of probability distributions that can be used to describe the distribution of topics across documents as well as words within topics. The algorithm in this model, iteratively assigns words in the documents to topics based on their co-occurrence patterns and adjusts these assignments to maximize the likelihood of the observed data. This process reveals the underlying topics and their distribution across documents, enabling the extraction of abstract themes from large text corpora (Blei et al., 2003).

In the data preparation stage, we use several natural language processing (NLP) techniques to transform abstract texts into meaningful units representing different technological topics. The process begins with tokenization, where each sentence is divided into individual words or "tokens." This is followed by the removal of stop-words, which are common words that do not add significant meaning, such as "and," "the," and "in." We utilize the NLTK² package for this task. Additionally, we refine our stop-words list by identifying topic-specific stop-words. For instance, in our blockchain topic model, words like "first" and "second" appear frequently but lack meaningful context, so we add them to our stop-words list.

Next, we identify and eliminate bigrams and trigrams—pairs or triplets of adjacent words—to simplify text representations and facilitate a deeper understanding of the content. Finally, we filter out high-frequency words that appear in over 90% of the documents and low-frequency words that appear in less than 0.1% of the documents. This ensures that our corpus is clean and uncluttered, making it suitable for efficient analysis and interpretation. By employing systematic operations, we extract relevant data from abstract and title texts, ultimately improving the precision and accuracy of our preprocessing steps. This careful preparation lays the foundation for meaningful analysis and interpretation of the technological topics under study.

² Python's Natural Language Toolkit

List of extended stop-words:

'from', 'first', 'second', 'one', 'method', 'methods', 'blockchain', 'may', 'based', 'include'

To find the optimal number of topics, we ran the LDA model with different configurations and examine two criteria, coherency, and perplexity, to choose the best number of topics. Several studies have employed both coherence and perplexity metrics to determine the optimal number of topics (Ghaffari et al., 2023; J. Lee et al., 2023; Montobbio et al., 2022; Nie & Sun, 2017). Coherence score indicates how related these words are to one another, and generally speaking, topics with higher coherence scores are easier to interpret. In this research coherency was calculated with the *c_v* measure (Mimno et al., n.d.), which evaluates the semantic interpretability of the topics based on the co-occurrence of the top words.

In order to determine whether a topic is meaningful and coherent, we can look for a number of topics where coherence peaks or plateaus. However, adding more subjects may not always result in higher-quality topics because of overfitting errors, which might provide unnaturally high coherence values. In order to solve this issue, perplexity is also used to identify the optimal number of topics. Perplexity is a metric for assessing how well a probabilistic topic model performs. Perplexity in the LDA model is calculated based on the log-likelihood of the words in the corpus under the model's learned topic distributions. It shows the degree to which the model can forecast a set of documents that are either held out or unseen. A model that fits the data better is indicated by lower perplexity numbers. However, perplexity decreases with the increasing number of topics, and it's crucial to assess whether the improvement in perplexity is worth the increased complexity of new topics.

To ensure high confidence in this determination, a systematic analysis was conducted. The LDA model was tested under twelve different conditions, varying the values of seed, passes, and alpha, which are key variables in the Genism LDA model. Seed represents the initial random number used at the start of the process, passes indicate the number of iterations through the text, and alpha controls the sparsity of document-topic distributions. The specific values chosen for each variable were selected based on common numbers found in relevant literature, ensuring consistency and comparability with prior research (Grün et al., 2011; Kumari et al., 2021). This rigorous approach enhances the reliability of the analysis and

strengthens confidence in the resulting topic modeling outcomes. The results of the 12-plot analysis, driven by the specified conditions, indicate that having three topics yields both the highest local coherence and the lowest local perplexity. Thus, we determine the optimal number for clustering blockchain-related content based on patent texts as three topics.

3.4 Phase 3: Network Analysis

We use patent citations to build the network and eliminate the citations that did not relate to the chosen topics. The remaining database contains only citations from patents that are within the chosen topic which was 8973 citations. The co-citation networks show how patents are related to one another in terms of their shared underlying knowledge. Network analysis consists of two steps: a patent citation network is created for the main path analysis, and a citation network for assignees is constructed.

3.4.1 Main path Analysis

Citation information includes many interlink relationships; some of these relationships are critical paths, while others are not. It is necessary to perform the main path analysis to determine the critical developing paths. The main path analysis involves two steps. In the first step, the traversal score, which indicates the importance of each citation link, is calculated. In the second step, the most important path is identified based on the nodes with the highest traversal scores (Hummon & Dereian, 1989). In a citation network, a ‘source’ is a node that is cited but cites no other nodes; a ‘sink’ is a node that cites other nodes but is not cited. In other words, sources are the origins of knowledge, whereas sinks are the end points of knowledge dissemination. Search path count (SPC), search path link count, search path node pair, and node pair projection count are algorithms available to calculate traversal weight (Batagelj, 2003; Hummon & Dereian, 1989). This study applied the SPC algorithm because of its efficiency compared to other algorithms (Liu & Lu, 2012). This algorithm searches all possible paths between sources and sinks, and then counts the number of times each arc lays on one of these paths.

To identify the main path, we employed the "priority first search" algorithm proposed by Hummon and Doreian (1989). This algorithm operates by selecting the next link with the highest traversal count at each node, continuing this process until a terminal node is reached, thereby constructing what is referred to as a main path. The main path represents the most critical sequence of links through which knowledge disseminates from an initial node. However, this approach primarily identifies a "local" main path, which focuses on the most direct routes within a specific segment and may not always reflect the path with the highest overall traversal counts across the network. Consequently, we also consider a "global" main path, which is the path with the highest total traversal counts, highlighting the most significant routes in the overall knowledge flow. The local main path is typically utilized when the objective is to understand the most immediate and direct routes of knowledge dissemination within a confined part of the network which is usually useful for predicting the future impact of a node. In contrast, the global main path is used when the goal is to identify the most important and overarching routes of knowledge flow throughout the entire network, providing a broader perspective on the dissemination of information.

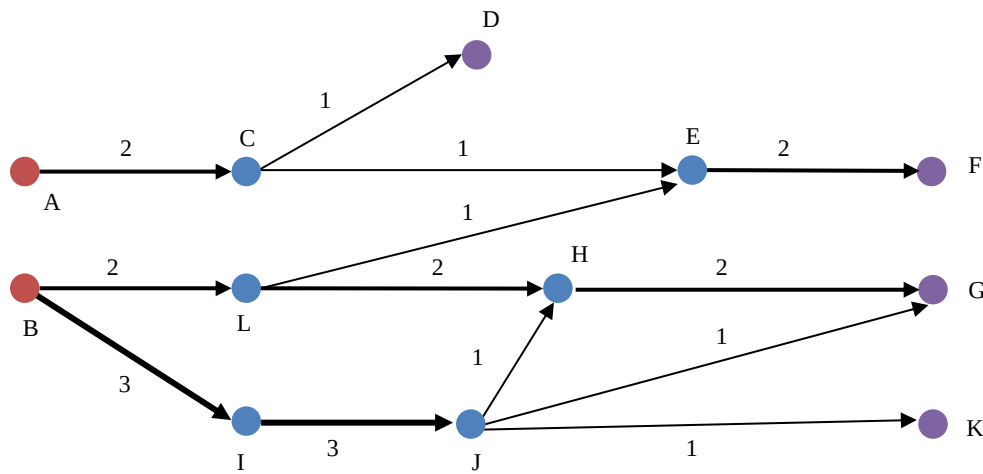


Figure 3.2 A simple network citation with SPC values

To better understand how a main path is determined, consider the example depicted in Figure 2. In this network, sources A and B are marked in red, while sinks D, F, G, and K are

marked in purple. The link H-G has an SPC value of 2 because it is traversed by the paths B-L-H-G and B-I-J-H-G. Similarly, the link B-I has a score of 3 because it is traversed by the paths B-I-J-K, B-I-J-G, and B-I-J-H-G. Higher SPC values indicate a link's significant role in transmitting knowledge. These SPC values help in identifying the main path by highlighting the routes most frequently used for knowledge dissemination. By focusing on these high-traffic links, the main path provides insights into how information flows through the network, emphasizing the most critical connections.

In summary, MPA highlights the critical paths of knowledge dissemination within citation networks by calculating traversal counts and identifying significant paths both locally and globally, providing insights into the evolution and trajectory of technologies. We use Pajek software to implement the specified algorithms and derive various main paths in our research. Pajek's advanced network analysis capabilities enable us to accurately trace and visualize the technological trajectories within each identified blockchain topic, ensuring a comprehensive understanding of their evolution and development.

3.4.2 Network Characteristics

In order to find the role and position of assignees, we constructed a citation network for assignees. This study employs social network analysis (SNA) methods to analyze betweenness centrality and O-I index measures, which serve to identify important nodes in the flow of knowledge within the network.

Betweenness centrality (BC): It quantifies how frequently a node serves as a pair of vertices' shortest path. High BC indicates strong potential to influence and regulate the information flow inside the network. In other words, a node with high betweenness centrality facilitates the flow of information between different parts of the network, serving as a key facilitator. This advantage contrasts with closeness centrality, which focuses on how quickly a node can reach other nodes but does not account for its role in connecting different network segments (Yuan & Li, 2021).

$$\text{Betweenness Centrality}(i) = \sum_{x \neq i \neq y} \frac{\sigma_{xy}(i)}{\sigma_{xy}} \quad (1)$$

where, σ_{xy} is the number of shortest path distance between nodes x and y and $\sigma_{xy}(i)$ is the number of those path that pass-through node i .

O-I index: The outward-inward (O-I) index measures how much one group of nodes refers to other nodes. This index can help identify the knowledge producers or absorbers within a network. The O-I index ranges from -1 to 1. A value closer to 1 suggests higher invention quality, while a value closer to -1 indicates greater accumulation of technical knowledge by the node.

$$O - I \text{ index} = \frac{\text{outdegree} - \text{indegree}}{\text{outdegree} + \text{indegree}} \quad (2)$$

While betweenness centrality indicates the extent to which a node facilitates communication and acts as a broker within the network, the O-I index reflects the degree of innovation associated with a node. Based on these metrics, nodes can be classified into four clusters (Choe et al., 2016):

- High betweenness centrality and a high O-I index, identified as "brokers based on producing knowledge," which indicates strong facilitation and high innovation.
- High betweenness centrality and a low O-I index, labeled as "brokers based on intermediary absorbing knowledge," signifying strong facilitation but lower innovation.
- Low betweenness centrality and a high O-I index, termed "knowledge producers," indicating high innovation but limited facilitation.
- Low betweenness centrality and a low O-I index, known as "knowledge absorbers," reflecting both low facilitation and innovation.

We use Gephi software to construct the network and calculate the betweenness centrality, indegree and outdegree for each node. To enhance our comparative analysis, we employ

normalized betweenness centrality. Gephi's tools allow us to accurately map and analyze the roles and influence of different nodes within the blockchain ecosystem.

Chapter 4 Result

4.1 Dominant themes and their evolutionary trajectory

Based on the LDA analysis results, we determined that the optimal number of topics is three. This configuration appeared most frequently among the 12 LDA runs, exhibiting the highest coherence and lowest perplexity. The most relevant terms for each topic and the number of patents associated with each topic are detailed in Appendix A. These topics were identified as the most significant and representative categories within the data. The three topics are as follows:

topic 1- Decentralized Data Management and Storage

topic 2- Smart Contract Execution

topic 3- Network Security and User Authentication

These are the three dominant topics in blockchain, identified from patent data using the LDA machine learning model. In the following sections, we will explore the global main path for each topic using the specified algorithms. This approach allows us to analyze the technological trajectory and evolution within each topic over time, providing a detailed understanding of the progress and development specific to each area, rather than a general overview of blockchain evolution.

4.1.1 Topic 1- Decentralized Data Management and Storage

The main path analysis of patents related to decentralized data management and storage as shown in Fig 4.1 reveals a structured progression of innovations contributing to the field's advancement. Foundational patents from Nasdaq, Inc. and an individual inventor address critical aspects of data integrity and transaction validation, laying the groundwork for subsequent developments. Innovations by Alibaba Group introduce efficient blockchain address mapping, enhancing data routing within decentralized networks. Patents from AliPay focus on authentication and identifier management, addressing key challenges in identity and access control within decentralized systems.

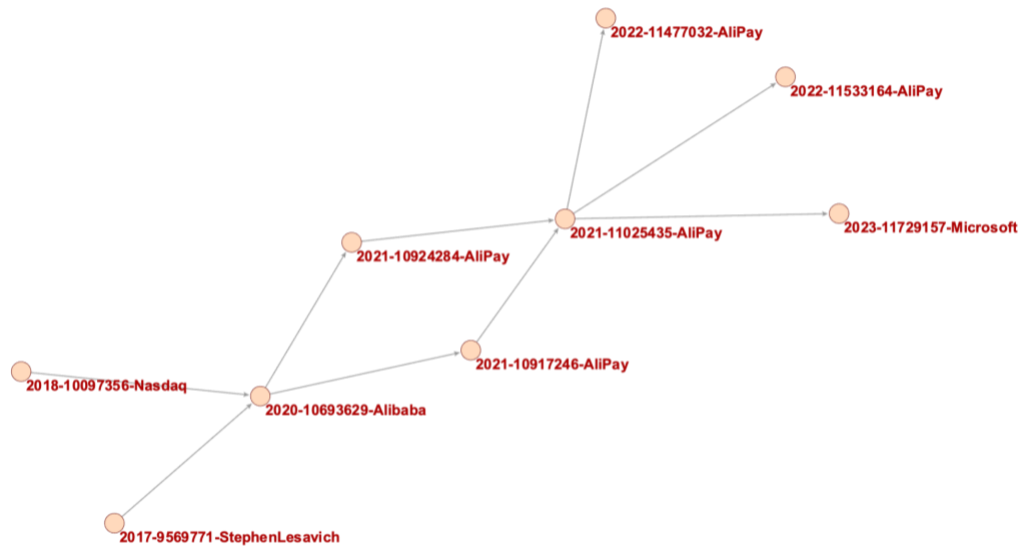


Figure 4.1 Main path analysis for topic 1

The latest technologies in this field originated from 11025435 are from AliPay and Microsoft. An interesting observation in the main path is the dominance of Chinese companies. While U.S. companies start the foundation of the decentralized data management and transactions, a financial Chinese company, AliPay, becomes a dominant player in developing blockchain and introducing novel approaches in this topic. In addition, it's noteworthy that despite IBM's large number of patents, it doesn't feature in the main path. This suggests that while IBM has a significant number of patents, they may not be as influential in driving the core advancements within the network.

4.1.2 Topic 2- Smart Contract Execution

The main path analysis of patents in smart contract as shown in Fig 4.2 highlights a progression of innovations focused on enhancing the functionality, efficiency, and stability of blockchain-based systems. At the forefront, Mastercard's 2018 patent introduces a method for linking blockchain-based assets to fiat currency accounts, establishing a crucial bridge between traditional financial systems and blockchain technology. Following this, Winklevoss IP's 2019 patent presents a system for generating stable value digital assets, addressing the volatility issues in digital currencies. Afterwards, Inveniam Capital plays a significant role with multiple

patents from 2022 and 2023, focusing on digital or "smart" contracts. These patents simplify the blockchain by using contract identifiers, which reduces the blockchain's size and processing requirements by offloading the execution code. This streamlines the deployment and execution of smart contracts, making blockchain systems more efficient. Additionally, their innovations include mechanisms for maintaining stable cryptocurrency coinage without requiring reserves, further enhancing the stability and usability of blockchain currencies. A notable point is that the starting point is from a well-known company, Mastercard, while the following innovations derives with new startups. This observation proves that blockchain as a disruptive technology create an environment to new entrants to access to huge resources and accumulated knowledge, contribute significantly to the development of blockchain.

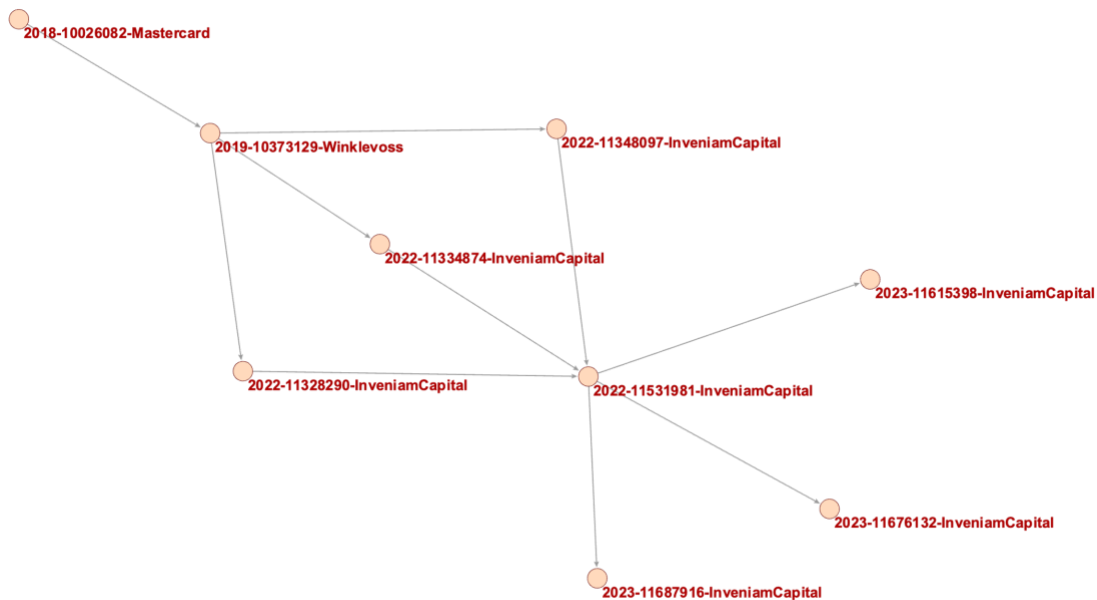


Figure 4.2 Main path analysis for topic 2

4.1.3 Topic 3- Network Security and User Authentication

The main path analysis of patents in the context of blockchain network security and privacy as shown in fig 4.3 reveals a strategic evolution of technologies aimed at enhancing the integrity, privacy, and efficiency of distributed ledger systems. Accenture along with GSC Secrypt leads the initial innovation with patents from 2017 and 2018 focusing on distributed key secrets and

hybrid blockchains. Subsequently, EYGS LLP's patents from 2020 onwards introduce sophisticated methods utilizing zero-knowledge proofs (ZKPs) to enhance privacy and efficiency in distributed ledger networks. These ZKP-enabled systems allow participants to verify transaction information without revealing private data, addressing critical privacy concerns while maintaining network consensus.

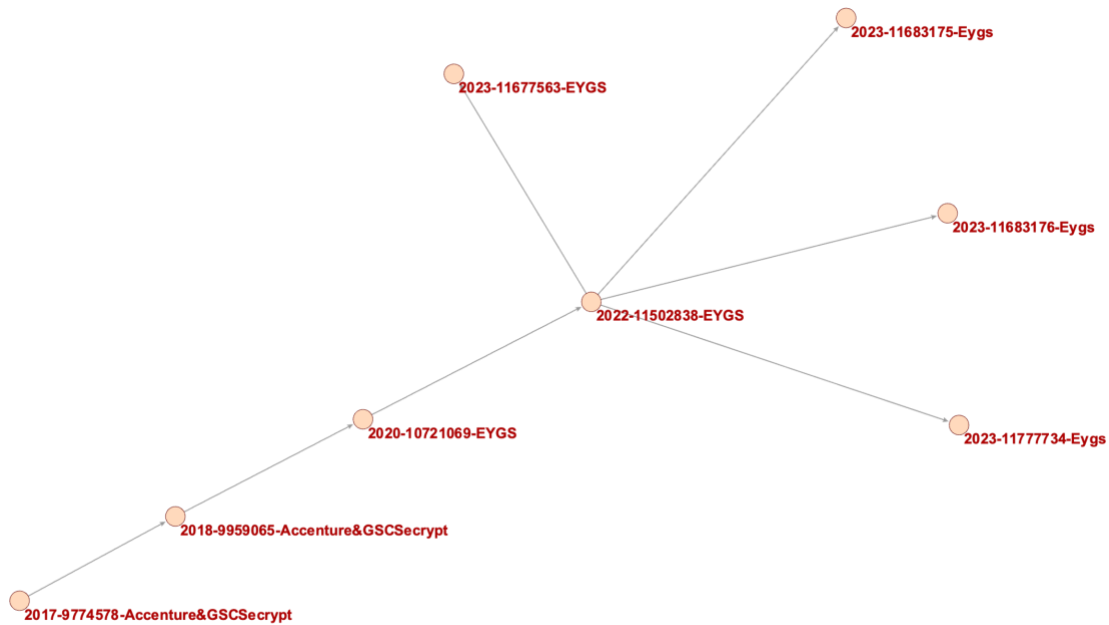


Figure 4.3 Main path analysis for topic 3

Moreover, EYGS LLP's patents from 2022 and 2023 focus on the practical application of these technologies for tracking and recovering stolen assets on distributed ledger networks, further bolstering security. The introduction of systems for local state storage of distributed ledger data without data cloning in 2023 further optimizes storage efficiency. Collectively, these patents reflect a comprehensive approach to securing blockchain networks and authenticating users, highlighting significant advancements in maintaining privacy, preventing tampering, and efficiently managing data. In this topic also same as previous topic, the path starts with well-known companies and follows by a startup. Another notable thing is that all companies are from Europe, and American companies, despite having the large portion of patents in the dataset, didn't play an important role in the security and privacy in blockchain.

4.2 Knowledge flow among dominant themes

Analyzing each topic separately provides valuable insights into specific areas of blockchain technology, such as Smart Contract Execution, Decentralized Data Management, and Network Security. However, this approach falls short in revealing how these domains influence and enrich each other. For a comprehensive understanding, Figure 4.4 illustrates the interconnectedness between these topics. This interconnected analysis uncovers the synergies and dependencies between different technological advancements. For instance, innovations in smart contract execution might drive new developments in decentralized data management, or improvements in network security might enhance the reliability and trustworthiness of both smart contracts and data storage solutions.

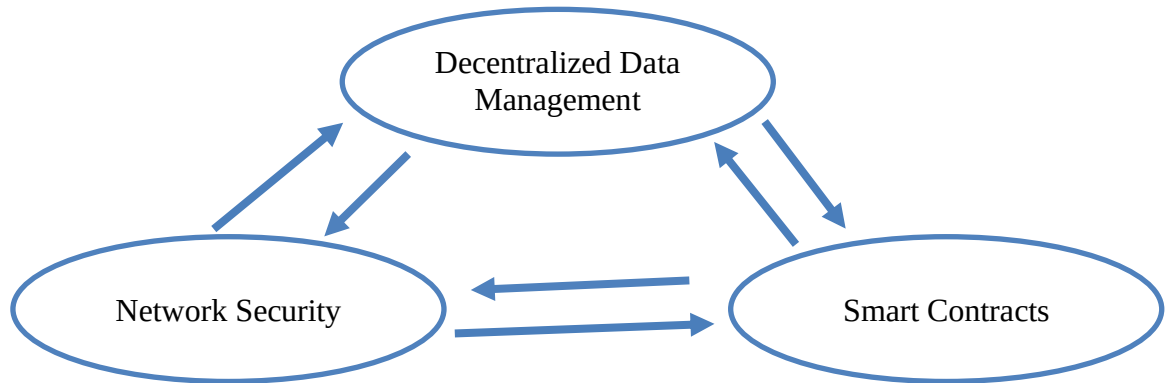


Figure 4.4 knowledge flow model analysis

4.2.1 Knowledge flow from data management

This section examines the knowledge flow from data management to smart contracts and network security. The route began with Manifold and ShoCard in 2016 and 2017 respectively. The Manifold Technology's patent (9397985) introduces a cryptographic platform for secure information exchange via a distributed ledger, while ShoCard's patent (9722790) focuses on identity management using blockchain to certify transactions between devices. Both cases show a transition to new topics following these initial patents, indicating their significant influence on the direction of knowledge dissemination. Another noteworthy similarity between

the two cases is the proliferation of branches towards the end. This implies that there is no singular endpoint in the patent landscape, and there are numerous opportunities for further growth and dissemination of knowledge. The abundance of branches underscores the dynamic nature of blockchain innovation, with ample room for exploration and expansion.

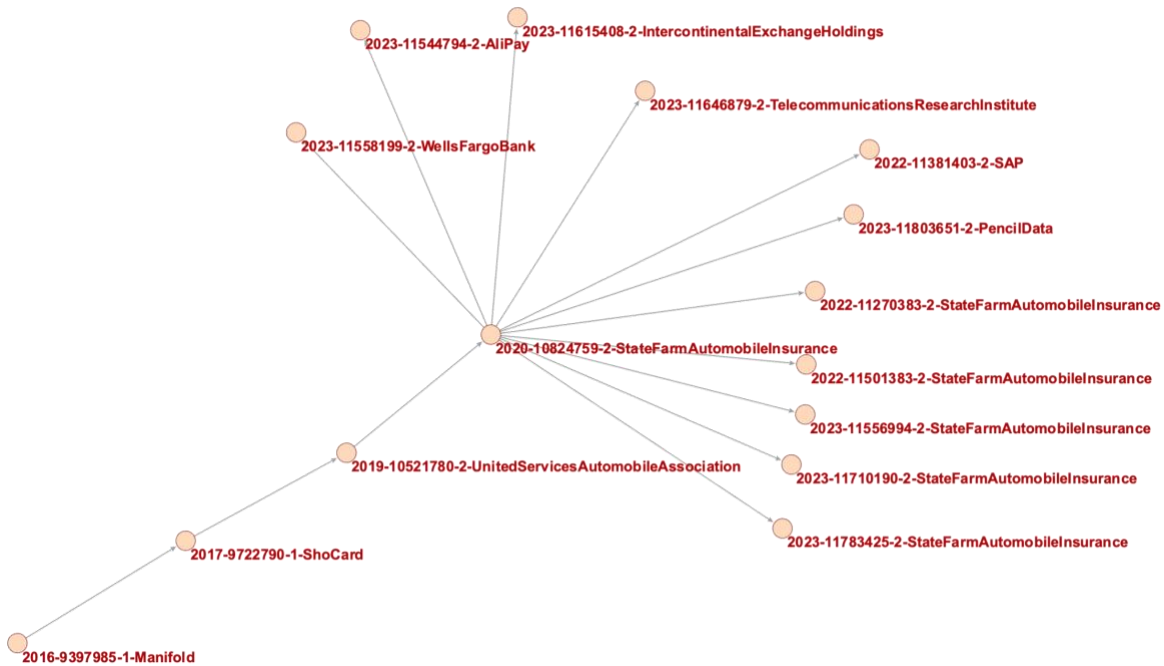


Figure 4.5 knowledge flow from data management to smart contract

Fig 4.5 (toward smart contracts) shows several patents originating from notable entities such as SAP SE, State Farm, AliPay, Wells Fargo Bank, Intercontinental Exchange Holdings, Electronics and Telecommunications Research Institute, and PencilData. These patents cover a diverse array of applications, ranging from insurance payments and claims to innovative methods for achieving consensus and generating dynamic smart contracts. The presence of patents from these entities underscores the breadth of applications and the innovative approaches being explored in the realm of smart contracts.

In Fig 4.6 (toward network security), the landscape is marked by the participation of established industry giants like Alipay, IBM, and Bank of America, alongside newer players like Dapper and Aneco Technologies. This juxtaposition highlights a burgeoning competitive

environment within network security. Unlike traditional sectors dominated solely by established players, the blockchain network security domain offers space for both established entities and emerging innovators to thrive. This coexistence fosters a dynamic ecosystem where competition and collaboration drive advancements in blockchain technology.

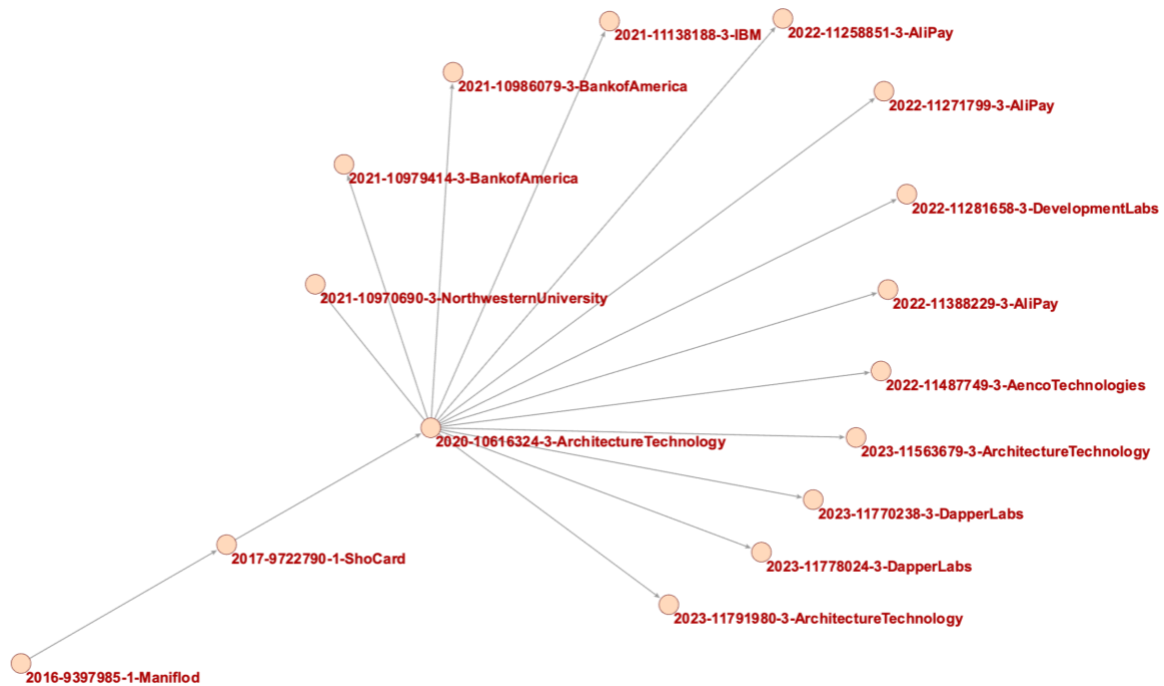


Figure 4.6 knowledge flow from data management to network security and privacy

This shift not only enhances existing digital infrastructures but also lays the groundwork for future breakthroughs in secure and decentralized technologies. The collaboration between established and emerging entities enriches the current landscape and paves the way for a more resilient and innovative future in blockchain-powered network security.

4.2.2 Knowledge flow from smart contract

This section examines the knowledge flow from smart contracts towards data management and security, noting two notable differences compared to the knowledge flow from network security. First, unlike the path analysis towards network security which tends to converge

towards a few patents, the path towards data management and security resulted in a multitude of branches. Second, in the knowledge flow towards data management, there is a discernible limitation. Most patents within topic 2 focus predominantly on advancements specific to smart contract execution, with only a handful extending into decentralized data management and storage solutions.

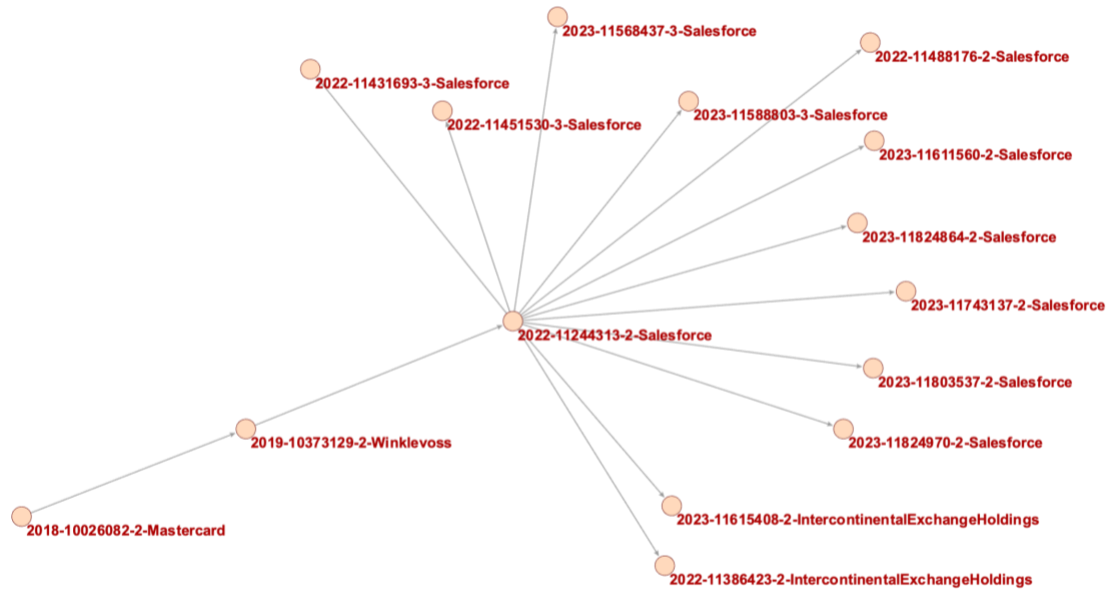


Figure 4.7 knowledge flow from smart contract to network security and privacy

In Fig 4.7, key patents at the beginning include 10026082 by Mastercard and 10373129 by Winklevoss IP, which concentrate on fundamental aspects of blockchain technology. These patents emphasize integrating blockchain-based assets with traditional financial systems and developing stable value digital assets. The main path is largely populated by patents from Salesforce, indicating a significant emphasis on leveraging blockchain technology within their ecosystem. This strategic focus by Salesforce underscores their dedication to exploring and implementing blockchain solutions to enhance network security and potentially transform various facets of their operations.

Turning to the data management topic in Fig 4.8, while the path commences with patents from established entity like Mastercard, it follows a distinct pattern. There is minimal

evidence of knowledge transition towards data management. Only two patents from Securrency originate from HP in the last part of the path. This suggests that innovations in data management largely evolve independently from developments in smart contract technologies. Furthermore, the limited transitions primarily revolve around secure data transfer and multi-lateral data exchange over decentralized networks, underscoring a gap in integrating smart contract technologies with broader data management solutions.

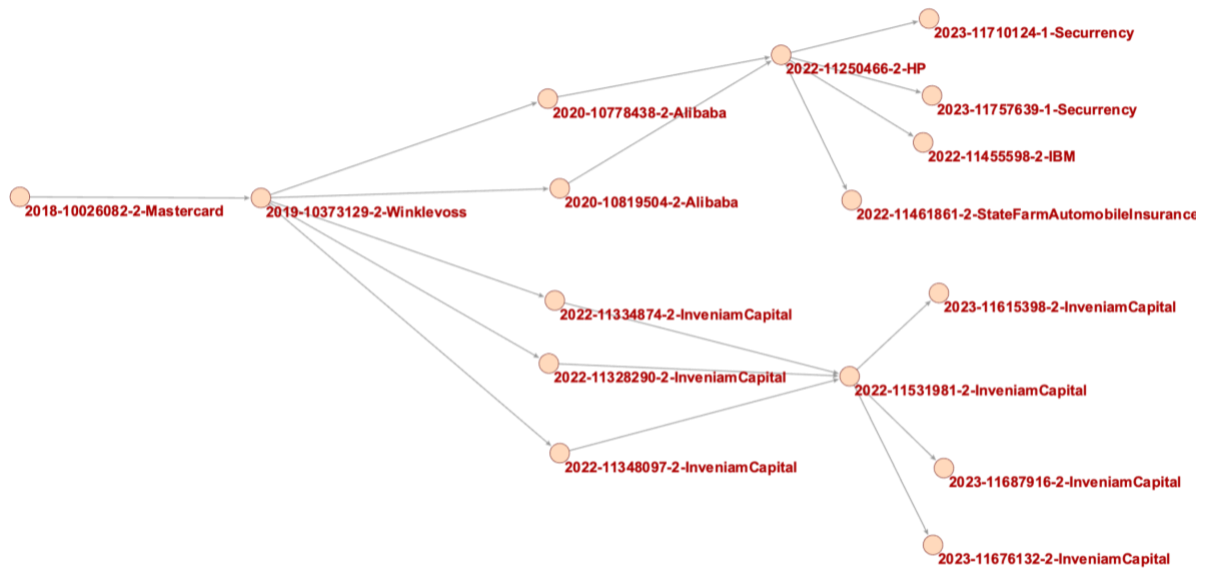


Figure 4.8 knowledge flow from smart contract to data management

4.2.3 Knowledge flow from network security

This section examines knowledge flows from network security towards smart contracts and data management to uncover any influence or knowledge exchange. Both Fig 4.9 and Fig 4.10 reveal a common starting point with two patents from new entrants: Coinlab in 2016 and Saed Muftic in 2017. Coinlab's patent (9298806) focuses on enhancing security through transaction analysis in a distributed ledger system. This foundation is further developed by Saed Muftic's patent (9635000) in 2017, which introduces a blockchain identity management system ensuring secure creation, validation, and storage of digital identities on a public ledger, thus enhancing user authentication and privacy.

The main path towards smart contracts is linear, with one patent at each step, initially led by established companies but concluding with two patents from new entrants. Vijay Madiseti's patent (11556928) in 2023 introduces a method for filtering and updating transactions using smart contracts, emphasizing blockchain's role in automating and securing complex transaction processes. Zhiji Automotive Technology's patent (11570000) in 2023 illustrates advanced blockchain applications for processing and storing driving data, highlighting its utility in managing real-time data in a decentralized network.

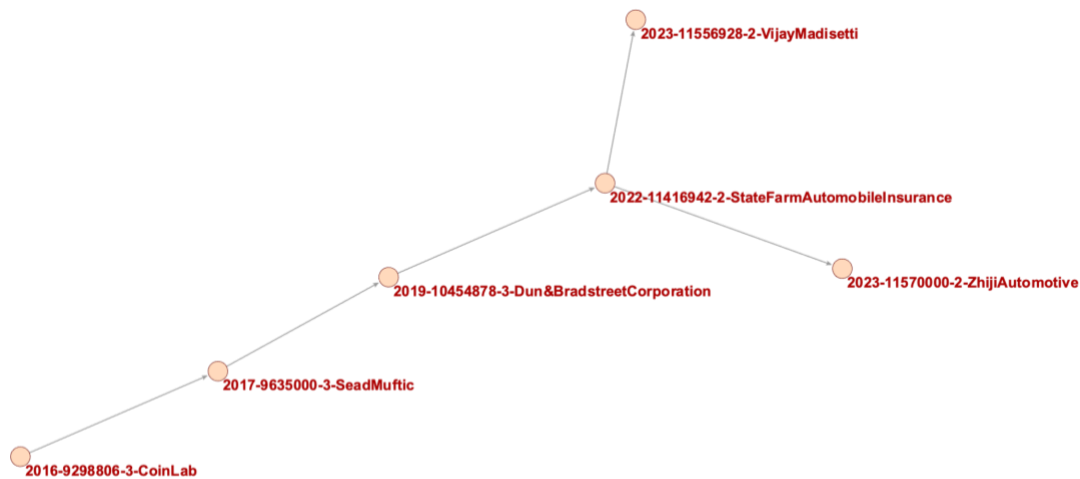


Figure 4.9 knowledge flow from network security and privacy to smart contract

In contrast, the path from network security towards decentralized data management exhibits a branching pattern before converging to a single patent. At the end, it branches out again, resulting in three different patents. A noteworthy observation is the inspiration drawn from patents by new entrants in network security, followed by a new entrant in the data management topic, with subsequent emergence of well-known Chinese companies playing a pivotal role in blockchain data management development. Overall, these inventions signify significant progress in leveraging blockchain technology for identity management and transaction analysis, presenting innovative solutions to real-world challenges in digital ecosystems.

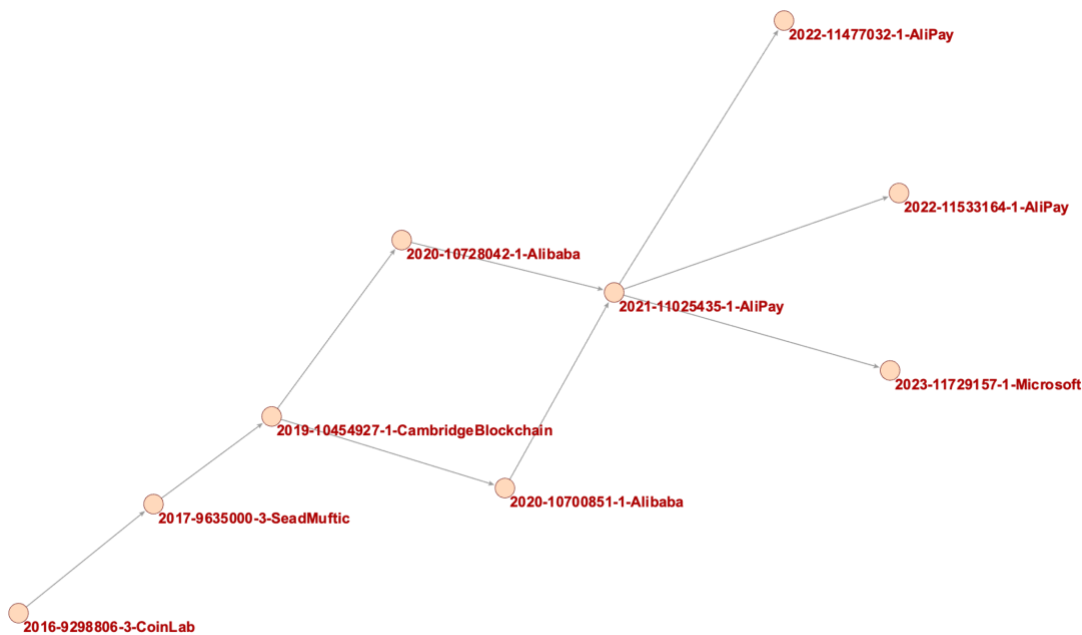


Figure 4.10 knowledge flow from network security and privacy to data management

The knowledge flows across different facets of the blockchain landscape reveal distinct trajectories, shedding light on the dynamics of innovation within the ecosystem. Notably, the pathways from network security towards smart contracts and data management revealed nuanced patterns. While the progression from network security displays a linear evolution, featuring contributions from both established industry players and emerging innovators, the route from data management showcases a branching pattern, indicative of the multifaceted approaches and applications within this sphere.

Central to our findings were key patents that acts as catalysts for knowledge dissemination, shaping the trajectory of innovation. Examples include Coinlab's emphasis on security enhancement through transaction analysis and Sead Mufic's introduction of a blockchain identity management system. Moreover, patents from industry stalwarts like Mastercard and Salesforce, alongside those from burgeoning entrants, underscored the collaborative and competitive dynamics driving blockchain advancement. Our exploration into

specific topics within each knowledge flow further underscored the breadth and depth of applications under exploration, from smart contracts facilitating insurance payments to network security solutions developed by industry giants like Alipay and IBM. This holistic understanding offers valuable insights into the potential of blockchain technology to revolutionize diverse sectors and drive transformative innovation in the digital landscape.

4.3 Assignees' positions and capabilities

The plots in Fig 4.11 illustrate the positions of assignees within each dominant topic. As depicted in the charts, a small number of assignees exhibit higher betweenness centrality scores. This indicates that these assignees play a crucial role in bridging various nodes within the network. Additionally, the concentration of assignees actively producing knowledge across all three topics is significantly higher than those which are merely absorbing knowledge. This demonstrates a dynamic innovation landscape where a few key players act as central hubs, facilitating the flow of information and fostering advancements. These leading assignees not only generate new ideas but also connect disparate parts of the network, thereby enhancing overall knowledge dissemination and integration within the blockchain domain.

The first largest company in terms of patent registrations is Ali Pay. Nonetheless, the charts reveal that it does not play a major role in the creation of new knowledge. Instead, this company primarily focuses on absorbing existing knowledge. Specifically, in topic 1 - Decentralized Data Management and Storage, it exhibits high betweenness centrality, functioning as a knowledge broker. Which also emerged substantially in the topic 1 main path analysis.

While the second largest company, IBM, doesn't play a role in the main paths, it plays a crucial role in the knowledge flow network with the highest betweenness centrality. This high centrality signifies IBM's role as a critical intermediary and facilitator, efficiently disseminating knowledge to other nodes. However, despite its pivotal position in knowledge distribution, IBM does not significantly contribute to the generation of new knowledge which can be explain why IBM didn't show in the main paths.

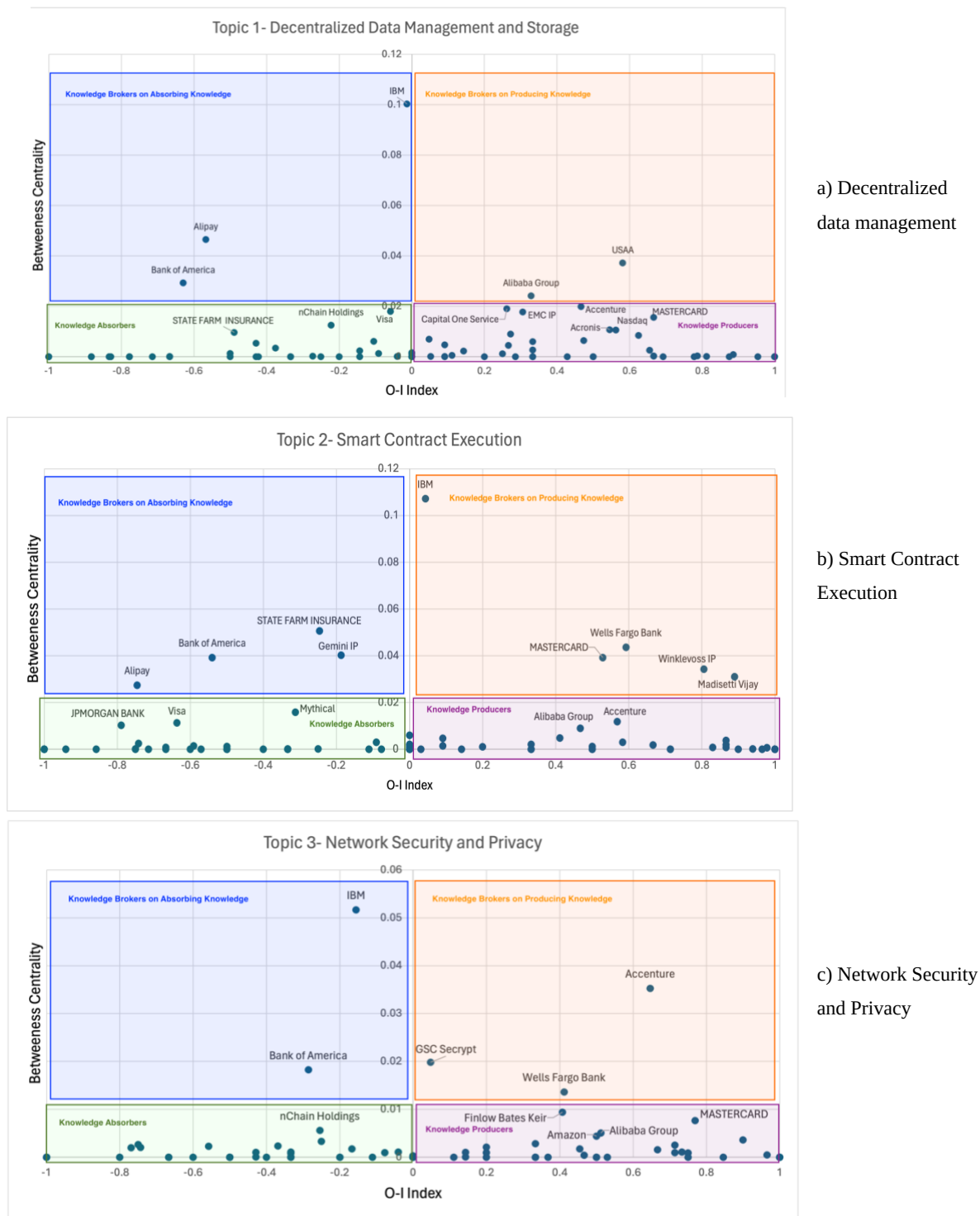


Figure 4.11 Assignees' positions based on network citation attributes

Similarly, the Alibaba Group, ranking third in patent registrations, demonstrates a dual role. In both topic 1 and topic 3, Alibaba serves as a producer of knowledge. Particularly in Topic 1, it operates as both a knowledge producer and a broker, facilitating the flow of

information. An important observation in Topic 1 and Topic 3 is the predominance of large companies, often referred to as incubators, in the role of brokers. These include established corporations such as IBM, Bank of America, USAA, Alibaba Group, and Accenture. However, Topic 3 also highlights the emergence of new entrants and venture capital firms in the knowledge-sharing domain. Companies like Gemini and Madiestti Vijay are new entrants, while Winklevoss IP, a venture capital firm, plays a substantial role in disseminating knowledge.

It is noteworthy that the majority of entities with a knowledge production score of 1 are new companies and individual innovators. This observation aligns with the evolution theory, which suggests that in the realm of new and disruptive technologies, it is often the newcomers who excel in knowledge creation and dominate the technological landscape. Established firms then strive to catch up and surpass these innovative entrants. The dynamics observed in the charts reflect this evolutionary process, where new entrants and venture capital-backed firms drive innovation and knowledge creation, while larger, established companies primarily facilitate knowledge dissemination and occasionally adopt and integrate these new innovations.

Chapter 5 Discussion

5.1 Interpretation of Results

This study attempts to tie together MPA and LDA topic modeling. The research questions are discussed further with theoretical and practical implications.

RQ1a: How many distinct dominant topics can blockchain technology be classified?

The LDA analysis reveals that the optimal number of topics is three. The three identified topics—Decentralized Data Management and Storage, Smart Contract Execution, and Network Security and User Authentication—represent distinct yet interrelated facets of blockchain innovation. This categorization is objective, differing from other methods that rely on researcher intuition and classify blockchain technology into broad categories like cryptocurrencies and distributed ledgers (B. T. S. Kim & Hyun, 2023). By focusing on these three dominant topics, we can gain deeper insights into the primary drivers of blockchain technology and better understand its multifaceted applications and potential impacts across various sectors. It is reported that several challenges, including security, privacy, consensus, and latency, pose significant hurdles for blockchain technology have mentioned (Bamakan et al., 2021a; Yli-Huumo et al., 2016). The analysis reveals that these challenges have indeed emerged as dominant topics within the blockchain innovation landscape, indicating a concerted effort to devise solutions and drive innovation in these areas.

However, the challenges like usability and resource wastage do not emerge prominently within the dominant topics identified in our study. This absence may suggest that either these challenges are not as pressing within the current discourse of blockchain innovation, or that existing solutions are perceived as adequate and thus do not warrant extensive exploration or discussion within patent filings. Alternatively, it could indicate a gap in innovation focus, where certain critical challenges are not receiving sufficient attention within the patent landscape despite their importance in broader discussions within the academic and industry communities. Further research and analysis would be needed to delve

deeper into this aspect and explore the implications for the ongoing development and adoption of blockchain technology.

RQ1b: what their evolutionary trajectory will reveal?

In the path analysis, the evolutionary trajectories reveal several significant insights. Regarding the similarities between the three topics, the analysis indicates a transition from fundamental principles to more intricate and specialized applications. For instance, early patents in Topic 1 focused on foundational aspects of data integrity and decentralized storage, which evolved into more complex solutions for seamless data management across distributed systems. Topic 2's trajectory showcases an initial emphasis on linking blockchain assets with traditional financial systems, progressing towards advanced smart contract deployment mechanisms that optimize transaction efficiency and security. Meanwhile, Topic 3 evolved from basic cryptographic methods to intricate security frameworks ensuring robust user authentication and system integrity. This diversity and evolutionary progression illustrate blockchain's expansive potential and its adaptability to various technological needs and challenges.

In terms of the path structure, the evolutionary trajectories of blockchain topics—data management, smart contracts, and network security—exhibit distinct patterns, reflecting their unique research focuses. Network Security follows a linear progression, indicating sequential advancements building on each other, with a significant breakthrough in 2022 influencing subsequent studies. In contrast, smart contracts and data management display complex, interconnected structures, suggesting parallel advancements and a dynamic research environment. Also, the observed patterns indicate that the evolution of blockchain technology was not characterized by a consistent year-after-year progression. In examining the main paths, it becomes apparent that there are one or two-year gaps present in all topics at certain points. This gap suggests pauses or periods of slower development within each topic, possibly due to shifts in research focus, technological limitations, or external factors influencing the trajectory of innovation. This intermittent nature of advancement underscores the complexity and non-linear nature of blockchain evolution.

Furthermore, concerning the assignees in each topic, a significant portion of the key assignees identified in the main paths of Topic 1 - data management and Topic 3 - network security and privacy comprises established companies. In contrast, assignees associated with Topic 2 - smart contracts predominantly consist of new entrants and smaller companies. This poses a challenge, as the Disruptive Innovation (DI) theory proposed by Christensen (2006) suggests that established incumbents often struggle to respond to disruptions brought about by new entrants in mainstream markets. It appears that, in the realm of blockchain, only smart contracts exhibit the characteristics of disruptive innovations, while the others tend to follow the path of accumulating knowledge within established firms. Furthermore, it's important to note the distinctive characteristics of pioneers and significant patents within each topic. In Topic 1, which predominantly focuses on data management, particularly within fintech, notable Chinese companies such as Alibaba and AliPay emerge as pivotal figures along the path. On the other hand, Topic 2 showcases a prevalence of lesser-known companies, signaling a landscape that remains relatively unclaimed by incumbents, thereby providing opportunities for new entrants to carve out a niche within the blockchain ecosystem.

Topic 3 delves into network security and privacy, and two noteworthy aspects emerge. First, the companies featured along this path are predominantly large and well-established entities based in Europe. Second, unlike fintech or major IT corporations, these entities primarily comprise consulting firms. This observation underscores the increasingly critical role of consultancy firms in propelling advancements in network security and privacy. Leveraging their strategic insights and advisory expertise, consulting firms are actively shaping the development and implementation of technology in these pivotal areas. This trend underscores the growing significance of multidisciplinary approaches and the indispensable value of consultancy in navigating intricate, technology-driven challenges.

RQ2: How does the flow of knowledge among dominant topics within the blockchain industry influence and stimulate each other's growth?

The flow of knowledge among dominant topics within the blockchain industry is pivotal in fostering mutual growth and innovation. For example, advancements in network security

(Topic 3) have significantly bolstered decentralized data management (Topic 1), as secure data handling is integral to effective decentralized systems. Conversely, improvements in data management techniques have facilitated the development of more efficient smart contracts (Topic 2), which rely on secure and reliable data inputs.

To provide further details, it is important to delve into the intricate knowledge flow observed within each blockchain domain. For example, in the transition from Topic 1 to subsequent topics, we discern a notable pattern. Two emerging companies, Manifold Technology and ShoCard, have emerged as influential figures, holding fundamental patents that drive growth in Topic 2 - smart contracts, and Topic 3 - network security. This illustrating how new players can significantly influence and drive innovation within other topics which shows the disruptive potential of blockchain. However, as we trace the trajectory, the path is predominantly shaped by established companies. This highlights a critical dynamic in technological evolution: while new entrants can introduce ground-breaking innovations and challenge existing norms, the sustained development and widespread adoption of these innovations often rely on the resources and expertise of established firms.

A similar pattern emerges in the knowledge flow from Topic 3 - network security, to other areas. Here, we observe the participation of both new entrants and individual inventors, such as Coinlab and Sead Muftic, influencing the evolution of knowledge in Topics 1 and 2. Moreover, in the transition from network security and privacy to data management, while initiated by new companies, the progression is notably fostered by incubators. This underscores how incubators leverage fresh resources and combine them with their capabilities to assert dominance in the emerging era of blockchain. Conversely, in the knowledge flow from network security to smart contracts, we witness a blend of new and established companies along the path. This finding matches with the significant presence of new entrants in the smart contracts topics which discussed in the previous research question. This synthesis reflects a spectrum of perspectives, where the agility of new entrants converges with the robust capabilities of established firms, driving advancements in smart contract technologies. The involvement of diverse entities across these knowledge flows underscores the collaborative

ecosystem inherent in blockchain innovation. In the knowledge flow originating from Topic 2 - smart contracts, to other topics, a distinct pattern emerges. A significant knowledge transfer pathway exists from smart contracts to network security (Topic 3), whereas there is no notable path from smart contracts to data management (Topic 1). The transition from Topic 2 to Topic 3 is characterized by a blend of contributions from both established and new companies. Notably, within this transition, Salesforce stands out as the sole prominent established company. This shows that in the realm of smart contracts influencing network security, both innovative startups and established firms play critical roles in driving technological advancements. However, the presence of Salesforce in this transition highlights how leading companies are crucial for integrating these innovations into mainstream applications, leveraging their extensive resources and market reach to enhance security frameworks. This mix of new entrants and industry giants underscores the collaborative effort required to advance complex technological fields like smart contracts and network security.

The cross-referencing of patents and the common technological developments that underpin a variety of themes demonstrate this connection. The mutually beneficial interaction between these subjects suggests that innovations in one field often spur advancements in others, resulting in a dynamic and ever evolving blockchain ecosystem. These interdependent knowledge flows foster creativity and aid in the development of increasingly robust and complex blockchain applications.

RQ3: How can we evaluate assignees' relative positions and capabilities in blockchain domains?

Evaluating the relative positions and capabilities of assignees in blockchain domains can be effectively captured by analyzing citation network metrics, such as betweenness centrality and the O_I index. Betweenness centrality indicates how often an assignee acts as an intermediary between other assignees, while the O_I index reveals whether an assignee is more involved in producing or absorbing knowledge. Based on these metrics, it is evident that established

companies play a significant role in spreading knowledge in Topic 1 (data management) and Topic 3 (network security and privacy). As previously mentioned, these companies are crucial in facilitating and sharing knowledge within the industry. In contrast, for Topic 2 (smart contracts), new entrant companies are also seen as key facilitators of knowledge flow.

These observations align with the main path analysis of each topic. In Topic 1 and 3, the main paths are dominated by established companies, whereas in Topic 2, new entrants are the primary influencers. This supports the theory that the smart contracts sector is an area where established companies are not as dominant, allowing new companies to secure important positions. Moreover, it is notable that in all topics, the relative density of new entrants in producing knowledge is higher than in absorbing knowledge. This suggests that there is ample opportunity for new entrants to create knowledge and significantly contribute to the development of blockchain technology.

Overall, established companies, with their extensive experience and resources, often act as central hubs for knowledge exchange. They leverage their positions to disseminate information and shape industry standards. Conversely, new entrants bring fresh perspectives and innovations to the discourse, often pioneering novel approaches that challenge existing paradigms. The interplay between these two types of actors highlights the collaborative and competitive nature of blockchain development. While established companies maintain their influence through extensive networks and established reputations, new entrants disrupt the status quo by pushing boundaries and driving innovation. This dynamic tension fosters a rich ecosystem where ideas are continuously challenged, refined, and disseminated, ultimately propelling the field forward.

5.2 Theoretical implications

This study integrates the LDA topic modeling and main path analysis in blockchain technology and provides a robust theoretical framework for understanding knowledge flow and the

development of blockchain technology across different dominant topics. Previous research has explored the blockchain landscape based on patents or articles to identify trends, forecast future paths, or analyze the blockchain life cycle. While a few researchers have clustered blockchain applications based on industry sectors like healthcare or IoT or used the key-route methods to interpret clusters based on citations, they often relied on subjective methods for clustering. This research is exploratory in its objective clustering, which facilitates a thorough analysis of blockchain development, the influence of each cluster on the others, and the recognition of the assignees' roles. By integrating LDA topic modeling with the main path analysis and examining the roles of assignees, this study provides a comprehensive mapping of technological innovations, identifying key themes and their interconnections, thereby illustrating the evolution path, and revealing significant assignees and their role in each topic.

Furthermore, this study underscores the dynamic and multifaceted nature of blockchain development, emphasizing the importance of cross-domain innovations. The interconnectedness of different themes suggests that advancements in one area can drive progress in another, indicating a synergistic relationship between various technological domains. This holistic view is crucial for understanding how different aspects of blockchain technology evolve and influence each other, and it highlights the role of both established companies and new entrants in the ecosystem.

5.3 Practical implication

This study offers practical implications that are highly valuable for companies' managers and decision-makers engaged in blockchain development. It underscores the significance of understanding the path-dependent nature of industrial evolution. Scholars suggest that once a specific developmental path is chosen, organizations may become reliant on it, shaping their future innovations and strategies (M. B. Jaffe, 1986). Therefore, analyzing development paths becomes essential for organizations to enhance their technical capabilities and secure a competitive edge (Arthur, 2007).

By pinpointing the primary topics in blockchain technology, tracing their developmental paths, and recognizing the assignee's role in each sector, companies gain valuable insights into emerging trends and potential innovation areas. Analyzing knowledge flow helps companies understand which organizations own specific technologies. This understanding enables companies to focus their resources and efforts on areas with the greatest impact and most significant growth opportunities. Moreover, by identifying key players and comprehending their relative positions and capabilities within the blockchain field, as well as the geographical distribution of knowledge, companies can strategically align with potential collaborators or identify competitors and establish their market position. This strategic insight allows companies to form partnerships that enhance their strengths and fill any capability gaps, thereby improving their competitive advantage in the rapidly evolving blockchain industry.

The analysis of knowledge flow among these dominant topics underscores the significance of fostering cross-domain collaborations. By recognizing the interconnectedness of various technological areas within blockchain, companies can harness synergies between different domains, thereby accelerating their pace of innovation. Collaborating with experts from diverse fields can result in the creation of more comprehensive and robust solutions, ultimately enhancing market competitiveness.

The insights derived from this study empower managers and decision-makers to navigate the complexities of the blockchain ecosystem with greater clarity and purpose. Understanding the intricate relationships between different technological domains allows companies to optimize their innovation strategies. By leveraging cross-domain collaborations and strategically aligning themselves with key players, organizations can better position themselves for success in the dynamic and rapidly evolving world of blockchain technology. This strategic approach helps managers decide whether to continue focusing on their niche, adopt competitive strategies, or pursue collaborative efforts. Ultimately, these strategies enable organizations to stay ahead of industry trends, drive innovation, and maintain a competitive edge in the market.

Chapter 6 Conclusion

6.1 Summary of Key Findings

Blockchain technology has garnered significant attention due to its rapid evolution and the disruptive innovations it brings to various industries. In our study, we utilized blockchain patents from the US Patent and Trademark Office (USPTO) and employed advanced techniques such as LDA topic modeling, MPA, and citation network analysis to uncover key themes, their progression over time, and identify significant growth paths and influential companies in the field. Our findings, derived from the LDA machine learning model, led us to identify three dominant topics within blockchain research: data management, network security and privacy, and smart contracts.

Before addressing the contribution, this thesis highlights the key findings of the research. Regarding our first research question, initially our analysis revealed that inventors predominantly focus on addressing challenges related to network security and privacy, consensus systems, and data management within the blockchain space, rather than other challenges like energy consumption and usability. Also, our examination highlighted interesting trends within the three dominant topics. Specifically, we found that US inventors only play a pivotal role in shaping advancements in smart contracts, while well-known Chinese companies such as Alibaba Group and AliPay are pioneers in the data management topic. Furthermore, European consulting companies like Accenture and EGYs emerged as key players in the network security and privacy domain. These findings suggest that the dominance of blockchain innovation across different topics is geographically diverse, aligning with the theory that knowledge flows often occur within specific geographic regions (Aldieri & Vinci, 2016; Kaiser, 2002; Nemet, 2012; Singh, 2005).

Regarding the second research question, the knowledge flow relationships between topics reveal interesting dynamics. Knowledge flows from topic 1 and topic 3 to other topics originate from new entrants but are continued and concluded by large companies, which contradicts the disruptive innovation theory, suggesting that large companies cannot surpass

new entrants (Christensen, 2006). This indicates that although new entrants initiate and inspire the path, large and established companies leverage their accumulated knowledge and organizational expertise to catch up.

And for the last research question, in topics 3 and 1, most knowledge brokers or producers are from established companies, suggesting that these companies receive more citations and attention due to their prominence compared to new companies. However, in topic 2, some new entrants can be observed, aligning with the main path analysis of topic 2, which predominantly consists of new entrants. This suggests that the smart contract topic offers more opportunities for new entrants, as it may not require as much accumulated knowledge or path dependency compared to the other two topics.

This research makes several contributions. To best of our knowledge this is the first study to utilize topic modeling techniques for objective clustering of blockchain, applied in knowledge flow analysis. Unlike the traditional approaches, which often rely on subjective interpretations or simple keyword searches, our methodology integrates an objective framework, allowing for a more precise assessment of the state of the blockchain field by analyzing the relationships between words and contexts among patents. This approach helps mitigate potential biases stemming from human subjectivity and offers a complementary perspective to existing research.

Furthermore, this classification empowers us to conduct main path analyses not only enables a more granular exploration within each topic but also facilitates a broader examination of knowledge dynamics across different facets of blockchain research. While prior studies often rely on subjective judgments, they also overlook exploring the relationships between topics and their potential influence on each other. This thesis gains valuable insights into how advancements in one area may influence or catalyze developments in another, thereby offering a more holistic understanding of blockchain's evolution. By recognizing that these roles may differ based on the specific topic under consideration, we can avoid oversimplification and better appreciate the nuanced contributions of each participant. Ultimately, this approach not

only enhances our analytical capabilities but also sheds light on the multifaceted nature of blockchain innovation.

In conclusion, this novel approach, which merges LDA topic modeling with MPA to explore the intricate channels of knowledge flow within dominant topics, can significantly enhance our understanding of the blockchain ecosystem. By objective classification and systematically examining how advancements in one area influence others, we gain a holistic understanding of blockchain's evolution. Also, this combined approach enables us to identify the diverse roles of entities in the blockchain ecosystem, acknowledging that these roles can differ across specific topics. This recognition prevents oversimplification and allows for a better understanding of each participant's nuanced contributions.

Both established and new entrants can benefit from the results of this study. Established companies can use this information to identify innovative firms and bolster their capabilities, thereby enhancing their R&D efforts and maintaining competitiveness in rapidly evolving markets. Meanwhile, new market entrants can leverage this research to understand the competitive landscape strategically. This understanding enables them to decide whether to follow existing market leaders or carve out unique positions where they can gain a competitive advantage, as seen in the realm of smart contracts. Additionally, insights from patent data enable new firms to develop effective R&D strategies, ensuring successful competition or even disruption of markets. Ultimately, the insights from this study empower companies to make well-informed strategic decisions, fostering a culture of continuous innovation and sustaining or attaining technological leadership.

6.2 Limitations of the Study

This study has several limitations. Firstly, it relies on patent data, which may not capture all innovations in the rapidly evolving blockchain field. Many significant advancements might not be patented, as they could be protected by trade secrets or other means (Cohen Richard R Nelson John P Walsh et al., 2000). Additionally, in the realm of blockchain, open-source projects play a crucial role in the ecosystem's development and might not be reflected in patent data.

Second, the study assumes that patents from the United States Patent and Trademark Office (USPTO) adequately represent blockchain technology. However, this assumption overlooks significant international contributions and other patent databases. Blockchain is a global technology, and including international patents would provide a more comprehensive view of its development and innovation.

Third, regarding the citation network, there is a debate about the accuracy of patent citations. Scholars proved many citations are made by patent examiners rather than inventors, with estimates suggesting that around 40% of citations are examiner-added (Nelson, 2009). Additionally, researchers might intentionally avoid citing competitors' patents, which can lead to incomplete or biased citation networks. This potential inaccuracy affects the reliability of results derived from citation network analysis. However, in this study most blockchain citations have been made by inventors rather than examiners, which may reduce the inaccuracies typically associated with citation data. This mitigates some concerns about the validity of the citation network analysis in this particular study. Despite these limitations, the study provides valuable insights into the blockchain technology landscape and highlights areas for further research and improvement.

6.3 Future Recommendations

Future research could expand the understanding of blockchain technology's development in several ways. First, researchers could broaden their analysis by including international patents and other databases for a more global perspective on blockchain innovation. Second, non-patented innovations, such as open-source projects and proprietary technologies protected by means other than patents, could provide additional insights on the blockchain ecosystem. Third, exploring the influence of regulatory changes and market dynamics on blockchain development is essential to shape technological advancements and adoption rates by addressing these areas, future studies can offer a more holistic understanding of blockchain innovation and its implications for development and adoption. This comprehensive approach will enable stakeholders to navigate the complexities of the blockchain ecosystem more effectively and capitalize on emerging opportunities.

References

- Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, 3–13.
<https://doi.org/10.1016/J.WPI.2013.12.006>
- Acemoglu, D., & Akcigit, U. (2012). Intellectual Property Rights Policy, Competition and Innovation. *Journal of the European Economic Association*, 10(1), 1–42.
<https://doi.org/10.1111/J.1542-4774.2011.01053.X>
- Adamuthe, A. C., & Thampi, G. T. (2019). *Technology forecasting: A case study of computational technologies*. <https://doi.org/10.1016/j.techfore.2019.03.002>
- Aghion, P., & Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, 60(2), 323. <https://doi.org/10.2307/2951599>
- Aldieri, L., & Vinci, C. P. (2016). TECHNOLOGICAL SPILLOVERS THROUGH A PATENT CITATION ANALYSIS. *International Journal of Innovation Management*, 20(2). <https://doi.org/10.1142/S1363919616500286>
- Allen, D. W. E., Berg, C., Markey-Towler, B., Novak, M., & Potts, J. (2020). Blockchain and the evolution of institutional technologies: Implications for innovation policy. *Research Policy*, 49(1). <https://doi.org/10.1016/j.respol.2019.103865>
- Altuntas, S., Dereli, T., & Kusiak, A. (2015). *Forecasting technology success based on patent data*. <https://doi.org/10.1016/j.techfore.2015.03.011>
- Arthur, W. B. (2007). The structure of invention. *Research Policy*, 36(2), 274–287.
<https://doi.org/10.1016/J.RESPOL.2006.11.005>
- Bamakan, S. M. H., Babaei Bondarti, A., Babaei Bondarti, P., & Qu, Q. (2021). Blockchain technology forecasting by patent analytics and text mining. *Blockchain: Research and Applications*, 2(2). <https://doi.org/10.1016/j.bcra.2021.100019>
- Batagelj, V. (2003). *Efficient Algorithms for Citation Network Analysis*.
<https://arxiv.org/abs/cs/0309023v1>
- Bhatt, P. C., Lai, K. K., Drave, V. A., Lu, T. C., & Kumar, V. (2023). Patent analysis based technology innovation assessment with the lens of disruptive innovation theory: A case of blockchain technological trajectories. *Technological Forecasting and Social Change*, 196. <https://doi.org/10.1016/j.techfore.2023.122864>
- Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. In *Journal of Machine Learning Research* (Vol. 3).
- Campbell, J. C., Hindle, A., & Stroulia, E. (2015). Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data. *The Art and Science of Analyzing Software Data*, 139–159. <https://doi.org/10.1016/B978-0-12-411519-4.00006-9>
- Cavalheiro, G. M. do C., & Cavalheiro, M. B. (2022). Assessing technological trends through patent landscaping: The case of Bitcoin. *Journal of World Intellectual Property*, 25(1), 206–219. <https://doi.org/10.1111/jwip.12216>
- Chang, S. E., & Chen, Y. C. (2020). Blockchain in health care innovation: Literature review and case study from a business ecosystem perspective. In *Journal of Medical Internet Research* (Vol. 22, Issue 8). JMIR Publications Inc. <https://doi.org/10.2196/19480>

- Chen, Y. L., & Hu, H. L. (2006). An overlapping cluster algorithm to provide non-exhaustive clustering. *European Journal of Operational Research*, 173(3), 762–780.
<https://doi.org/10.1016/J.EJOR.2005.06.056>
- Cho, R. L. T., Liu, J. S., & Ho, M. H. C. (2021). The development of autonomous driving technology: perspectives from patent citation analysis. *Transport Reviews*, 41(5), 685–711. <https://doi.org/10.1080/01441647.2021.1879310>
- Choe, H., Lee, D. H., Kim, H. D., & Seo, I. W. (2016). Structural properties and inter-organizational knowledge flows of patent citation network: The case of organic solar cells. In *Renewable and Sustainable Energy Reviews* (Vol. 55, pp. 361–370). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2015.10.150>
- Choi, D., & Song, B. (2018). Exploring Technological Trends in Logistics: Topic Modeling-Based Patent Analysis. *Sustainability 2018*, Vol. 10, Page 2810, 10(8), 2810.
<https://doi.org/10.3390/SU10082810>
- Christensen, C. M. (2006). The Ongoing Process of Building a Theory of Disruption. *Product Development & Management Association*, 23, 39–55.
- Clarke, N. S., Jürgens, B., & Herrero-Solana, V. (2020). Blockchain patent landscaping: An expert based methodology and search query. *World Patent Information*, 61.
<https://doi.org/10.1016/j.wpi.2020.101964>
- Cohen Richard R Nelson John P Walsh, W. M., Arora, A., Arundel, A., Baldwin, J., Ceccagnoli, M., Goto, A., Greenstein, S., Guellec, D., Hall, B., Kabla, I., Kiepper, S., Lanjouw, J., Lerner, J., Levinthal, D., Meyers, M., Pakes, A., Quillen, C., Scherer, M., Trajtenberg, M., ... Walsh Professor of Sociology, J. (2000). *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)*. <https://doi.org/10.3386/W7552>
- Dang, S., & Ahmad, P. H. (2014). Text Mining : Techniques and its Application Text Mining: Techniques and its Application. *IJETI International Journal of Engineering & Technology Innovations*, 1. www.ijeti.com
- Dosi, G. (1982). *Technological paradigms and technological trajectories A suggested interpretation of the determinants and directions of technical change*.
- Duguet, E., & MacGarvie, M. (2005). How well do patent citations measure flows of technology? Evidence from French innovation surveys. *Economics of Innovation and New Technology*, 14(5), 375–393. <https://doi.org/10.1080/1043859042000307347>
- Eligüzel, N., Çetinkaya, C., & Dereli, T. (2020). *Comparison of different machine learning techniques on location extraction by utilizing geo-tagged tweets: A case study*. <https://doi.org/10.1016/j.aei.2020.101151>
- Fernández, A. M., Ferrándiz, E., & Medina, J. (2022). The diffusion of energy technologies. Evidence from renewable, fossil, and nuclear energy patents. *Technological Forecasting and Social Change*, 178. <https://doi.org/10.1016/j.techfore.2022.121566>
- Ghaffari, M., Aliahmadi, A., Khalkhali, A., Zakeri, A., Daim, T. U., & Yalcin, H. (2023). Topic-based technology mapping using patent data analysis: A case study of vehicle tires. *Technological Forecasting and Social Change*, 193.
<https://doi.org/10.1016/j.techfore.2023.122576>
- Gilsing, V. A., Cloudt, M., & Bertrand-Cloudt, D. (2016). What makes you more central? Antecedents of changes in betweenness-centrality in technology-based alliance

- networks. *Technological Forecasting and Social Change*, 111, 209–221.
<https://doi.org/10.1016/j.techfore.2016.07.001>
- Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., & van den Oord, A. (2008). Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy*, 37(10), 1717–1731.
<https://doi.org/10.1016/J.RESPOL.2008.08.010>
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5228–5235.
<https://doi.org/10.1073/PNAS.0307752101/ASSET/A171028F-351B-4832-8A36-1B0244146E12/ASSETS/GRAPHIC/ZPQ0080438990006.JPEG>
- Grossman, G. M., & Helpman, E. (2018). Growth, Trade, and Inequality. *Econometrica*, 86(1), 37–83. <https://doi.org/10.3982/ECTA14518>
- Grün, B., Wu, K. H., & Wien, W. (2011). topicmodels: An R Package for Fitting Topic Models. In *JSS Journal of Statistical Software* (Vol. 40). <http://www.jstatsoft.org/>
- Hagedoorn, J., & Cloudt, M. (2003). Measuring innovative performance: is there an advantage in using multiple indicators? *Research Policy*, 32(8), 1365–1379.
[https://doi.org/10.1016/S0048-7333\(02\)00137-3](https://doi.org/10.1016/S0048-7333(02)00137-3)
- Hong Kim, K., Jae Han, Y., Lee, S., Won Cho, S., & Lee, C. (2019). *Text Mining for Patent Analysis to Forecast Emerging Technologies in Wireless Power Transfer*.
<https://doi.org/10.3390/su11226240>
- Hughes, L., Dwivedi, Y. K., Misra, S. K., Rana, N. P., Raghavan, V., & Akella, V. (2019). Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda. *International Journal of Information Management*, 49, 114–129. <https://doi.org/10.1016/j.ijinfomgt.2019.02.005>
- Hummon, N. P., & Dereian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11(1), 39–63. [https://doi.org/10.1016/0378-8733\(89\)90017-8](https://doi.org/10.1016/0378-8733(89)90017-8)
- Hwang, Y., Hyung, J., Kim, J., Hyung, M., Choi, J., Lee, J., Joonhwan, & Lee, J. (2020). *Exploring Abnormal Behavior Patterns of Online Users With Emotional Eating Behavior: Topic Modeling Study*. <https://doi.org/10.2196/15700>
- Jack, S. L. (2005). The Role, Use and Activation of Strong and Weak Network Ties: A Qualitative Analysis*. *Journal of Management Studies*, 42(6), 1233–1259.
<https://doi.org/10.1111/J.1467-6486.2005.00540.X>
- Jaffe, M. B. (1986). *Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value*.
- Kaiser, U. (2002). Measuring knowledge spillovers in manufacturing and services: an empirical assessment of alternative approaches. *Research Policy*, 31(1), 125–144.
[https://doi.org/10.1016/S0048-7333\(00\)00159-1](https://doi.org/10.1016/S0048-7333(00)00159-1)
- Kim, B. T. S., & Hyun, E. J. (2023). Mapping the Landscape of Blockchain Technology Knowledge: A Patent Co-Citation and Semantic Similarity Approach. *Systems*, 11(3).
<https://doi.org/10.3390/systems11030111>
- Kim, H., Ahn, S.-J., & Jung, W.-S. (2018). *Horizon scanning in policy research database with a probabilistic topic model*. <https://doi.org/10.1016/j.techfore.2018.02.007>

- Kim, S., Hwang, H., & Suh, E. (2003). A process-based approach to knowledge-flow analysis: A case study of a manufacturing firm: Knowledge and process management. *Knowledge and Process Management*, 10(4), 260–276. <https://doi.org/10.1002/kpm.186>
- Kim, S., Yoon, S., Raghavan, N., Le, N. T., & Park, H. (2021). Developmental Trajectories in Blockchain Technology Using Patent-Based Knowledge Network Analysis. *IEEE Access*, 9, 44704–44717. <https://doi.org/10.1109/ACCESS.2021.3066573>
- Kumari, R., Jeong, J. Y., Lee, B. H., Choi, K. N., & Choi, K. (2021). Topic modelling and social network analysis of publications and patents in humanoid robot technology. *Journal of Information Science*, 47(5), 658–676. <https://doi.org/10.1177/0165551519887878>
- Kwok, S. W. H., Vadde, S. K., & Wang, G. (2021). Tweet topics and sentiments relating to COVID-19 vaccination among Australian twitter users: Machine learning analysis. *Journal of Medical Internet Research*, 23(5), e26953. <https://doi.org/10.2196/26953>
- Kyebambe, M. N., Cheng, G., Huang, Y., He, C., & Zhang, Z. (2017). *Forecasting emerging technologies: A supervised learning approach through patent analysis*. <https://doi.org/10.1016/j.techfore.2017.08.002>
- Lai, K. K., Chen, H. C., Chang, Y. H., Kumar, V., & Bhatt, P. C. (2021). A structured MPA approach to explore technological core competence, knowledge flow, and technology development through social network patentometrics. *Journal of Knowledge Management*, 25(2), 402–432. <https://doi.org/10.1108/JKM-01-2020-0037>
- Lanjouw, J. O., & Schankerman, M. (2004). *PATENT QUALITY AND RESEARCH PRODUCTIVITY: MEASURING INNOVATION WITH MULTIPLE INDICATORS**.
- Lee, C., Kwon, O., Kim, M., & Kwon, D. (2017). *Early identification of emerging technologies: A machine learning approach using multiple patent indicators*. <https://doi.org/10.1016/j.techfore.2017.10.002>
- Lee, J., Zo, H., & Steinberger, T. (2023). Exploring trends in blockchain publications with topic modeling: Implications for forecasting the emergence of industry applications. *ETRI Journal*, 45(6), 982–995. <https://doi.org/10.4218/etrij.2022-0257>
- Lee, S., Han, E. J., & Sohn, S. Y. (2015). *Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents*. <https://doi.org/10.1016/j.techfore.2015.07.022>
- Liu, J. S., & Lu, L. Y. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology*, 63(3), 528–542. <https://doi.org/10.1002/asi.21692>
- Liu, J. S., Lu, L. Y. Y., & Ho, M. H. C. (2019). A few notes on main path analysis. *Scientometrics*, 119(1), 379–391. <https://doi.org/10.1007/s11192-019-03034-x>
- Mehri, S., Ammar, J., Sedighi, M., & Jalalimanesh, A. (2014). Mapping research trends in the field of knowledge management. *Malaysian Journal of Library and Information Science*, 19(1), 71–85. <https://mjlis.um.edu.my/index.php/MJLIS/article/view/1812>
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & Mccallum, A. (n.d.). *Optimizing Semantic Coherence in Topic Models*. Association for Computational Linguistics.
- Mo, Y., Kontonatsios, G., & Ananiadou, S. (2015). Supporting systematic reviews using LDA-based document representations. *Systematic Reviews*, 4(1). <https://doi.org/10.1186/s13643-015-0117-0>

- Montobbio, F., Staccioli, J., Virgillito, M. E., & Vivarelli, M. (2022). Robots and the origin of their labour-saving impact. *Technological Forecasting and Social Change*, 174. <https://doi.org/10.1016/j.techfore.2021.121122>
- Nakamoto, S. (2010). *Bitcoin: A Peer-to-Peer Electronic Cash System*. www.bitcoin.org
- Nelson, A. J. (2009). Measuring knowledge spillovers: What patents, licenses and publications reveal about innovation diffusion. *Research Policy*, 38(6), 994–1005. <https://doi.org/10.1016/j.respol.2009.01.023>
- Nemet, G. F. (2012). *Inter-technology knowledge spillovers for energy technologies*. <https://doi.org/10.1016/j.eneco.2012.06.002>
- Nie, B., & Sun, S. (2017). Using Text Mining Techniques to Identify Research Trends: A Case Study of Design Research. *Applied Sciences* 2017, Vol. 7, Page 401, 7(4), 401. <https://doi.org/10.3390/APP7040401>
- Noailly, J., & Shestalova, V. (2017). Knowledge spillovers from renewable energy technologies: Lessons from patent citations. *Environmental Innovation and Societal Transitions*, 22, 1–14. <https://doi.org/10.1016/j.eist.2016.07.004>
- Perdana, A., Lee, W. E., & Robb, A. (2021). From enfant terrible to problem-solver? Tracing the competing discourse to explain blockchain-related technological diffusion. *Telematics and Informatics*, 63. <https://doi.org/10.1016/j.tele.2021.101662>
- Qi, G., Wu, J., Zhou, Y., Du, Y., Jia, Y., Hounsell, N., & Stanton, N. A. (2018). *Recognizing driving styles based on topic models*. <https://doi.org/10.1016/j.trd.2018.05.002>
- Shen, L., Wang, S., Dai, W., & Zhang, Z. (2019). Detecting the interdisciplinary nature and topic hotspots of robotics in surgery: social network analysis and bibliometric study. *Journal of Medical Internet Research*, 21(3). <https://doi.org/10.2196/12625>
- Shih, H. Y., & Chang, T. L. S. (2009). International diffusion of embodied and disembodied technology: A network analysis approach. *Technological Forecasting and Social Change*, 76(6), 821–834. <https://doi.org/10.1016/J.TECHFORE.2008.09.001>
- Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science*, 51(5), 756–770. <https://doi.org/10.1287/mnsc.1040.0349>
- Siu, W. S., & Bao, Q. (2008). Network strategies of small Chinese high-technology firms: A qualitative study. *Journal of Product Innovation Management*, 25(1), 79–102. <https://doi.org/10.1111/j.1540-5885.2007.00284.x>
- Someda, H., Akagi, T., & Kajikawa, Y. (2022). An analysis of the spillover effects based on patents and inter-industrial transactions for an emerging blockchain technology. *Scientometrics*, 127(8), 4299–4314. <https://doi.org/10.1007/S11192-022-04457-9/TABLES/4>
- Sun, B., Mao, H., & Yin, C. (2020). Male and Female Users' Differences in Online Technology Community Based on Text Mining. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00806>
- Trappey, C. V., Wu, H. Y., Taghaboni-Dutta, F., & Trappey, A. J. C. (2011). Using patent data for technology forecasting: China RFID patent analysis. *Advanced Engineering Informatics*, 25(1), 53–64. <https://doi.org/10.1016/J.AEI.2010.05.007>
- Tseng, C. Y., & Ting, P. H. (2013). Patent analysis for technology development of artificial intelligence: A country-level comparative study. *Innovation: Management, Policy and Practice*, 15(4), 463–475. <https://doi.org/10.5172/impp.2013.15.4.463>

- Venugopalan, S., & Rai, V. (2015). Topic based classification and pattern identification in patents. *Technological Forecasting and Social Change*, 94, 236–250. <https://doi.org/10.1016/J.TECHFORE.2014.10.006>
- Wang, C. C., Sung, H. Y., Chen, D. Z., & Huang, M. H. (2017). Strong ties and weak ties of the knowledge spillover network in the semiconductor industry. *Technological Forecasting and Social Change*, 118, 114–127. <https://doi.org/10.1016/j.techfore.2017.02.011>
- Wang, Y. H., Trappey, A. J. C., & Trappey, C. V. (2015). Life Cycle Analysis of the Optical Disc Industry Market Innovation and Development. *Innovation*, 17(2), 196–216. <https://doi.org/10.1080/14479338.2015.1011061>
- Welbers, K. (2015). *Quantitative analysis of large amounts of journalistic texts using topic modelling*. <https://doi.org/10.1080/21670811.2015.1093271>
- Williams, C. (2011). Client-vendor knowledge transfer in IS offshore outsourcing: Insights from a survey of Indian software engineers. *Information Systems Journal*, 21(4), 335–356. <https://doi.org/10.1111/j.1365-2575.2010.00354.x>
- Yang, Y., & Qi, S. (2014). A new feature selection method for computational prediction of type III secreted effectors. *International Journal of Data Mining and Bioinformatics*, 10(4), 440–454. <https://doi.org/10.1504/IJDMB.2014.064894>
- Yli-Huomo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where Is Current Research on Blockchain Technology?—A Systematic Review. *PLOS ONE*, 11(10), e0163477. <https://doi.org/10.1371/JOURNAL.PONE.0163477>
- Yu, D., & Pan, T. (2021a). Identifying technological development trajectories in blockchain domain: a patent citation network analysis. *Technology Analysis and Strategic Management*, 33(12), 1484–1497. <https://doi.org/10.1080/09537325.2021.1879381>
- Yu, D., & Pan, T. (2021b). Tracing the main path of interdisciplinary research considering citation preference: A case from blockchain domain. *Journal of Informetrics*, 15(2). <https://doi.org/10.1016/j.joi.2021.101136>
- Yu, D., & Sheng, L. (2020). Knowledge diffusion paths of blockchain domain: the main path analysis. *Scientometrics*, 125(1), 471–497. <https://doi.org/10.1007/s11192-020-03650-y>
- Yuan, X., & Li, X. (2021). Mapping the technology diffusion of battery electric vehicle based on patent analysis: A perspective of global innovation systems. *Energy*, 222. <https://doi.org/10.1016/j.energy.2021.119897>
- Zhang, H., Daim, T., & Zhang, Y. (Peggy). (2021). Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain. *Technological Forecasting and Social Change*, 167, 120729. <https://doi.org/10.1016/J.TECHFORE.2021.120729>

Appendix A

Topic Modeling's Result

Topic 1:

Top words: data, system, block, ledger, device, distributed, transaction, information, storage, associated

Number of patents = 1403 out of 3810 (36.82%)

Topic 2:

Top words: transaction, contract, smart, network, digital, system, ledger, distributed, node, user

Number of patents = 1376 out of 3810 (36.12%)

Topic 3:

Top words: key, network, system, data, node, user, service, public, block, device

Number of patents = 1031 out of 3810 (27.06%)