Essays on Blockchain Business Ecosystems

Inaugural-Dissertation

to obtain the degree of Doctor of Business Administration
(doctor rerum politicarum – Dr. rer. pol.)

submitted to the

Faculty of Business Administration and Economics
Heinrich Heine University Düsseldorf

presented by

Xian Zhuo, M.Sc.
Scholar at Manchot Graduate School
"Competitiveness of Young Enterprises"
Heinrich Heine University Düsseldorf

First Supervisor: Prof. Dr. Denefa Bostandzic
Department of Management and Entrepreneurship
Chair of Corporate Finance
Witten/Herdecke University

June 27, 2024

Acknowledgements

When I was a teenager, I built up the vision of my life. I thought of being an artist or a poet, exploring the world in my own way. While I have never portrayed myself as a business student, my life has led me here in a surprisingly logical way.

Starting my PhD journey, I imagined how joyful it would be when I finally reached the final stage of my thesis and started writing my acknowledgment. Yet, at this very moment, when sitting here contemplating my feelings and to whom I should pay gratitude, I am equally overwhelmed and empty. Moments - happy, depressed, energetic, sleepy - come to me all at once. They are deeply intertwined, and I find it hard to disentangle them. But one thing is clear: I am not alone on this journey.

I am deeply grateful for all the support and guidance my supervisors, Denefa Bostandzic and Felix Irresberger, have given me. Whenever I needed help, they were always there, listening and trying to find solutions. As the German expression goes, they are father and mother to me on this journey. Without them, I would not have started, but most importantly, I would not have reached this stage. At this point, I would also like to thank Professor Christoph Börner and Professor Janine Maniora for taking the time to be my examiners for my viva.

I would like to express my gratitude to the Manchot Graduate School for offering me the position and generous financial support to pursue my doctoral research. My appreciation extends to everyone I met at Manchot and the faculty. I appreciate every great conversation we had and the support they gave.

I cannot forget to thank everyone I encountered at the conferences and workshops. I remember the shared-interest-sparked conversations, and more than that, the moments we talked about the research challenges and explored cities together.

Behind the doctoral facade, I am deeply surrounded by the all-around love of my family and friends. Whether or not they understand what I am doing, their constant listening, comfort, and encouragement are the backbones that keep me going. I thank my brother Yue Zhuo for his emotional support. I especially thank my husband Nikolas Feistkorn, who told me to believe in myself and gave me all the love and support I needed to achieve my goals.

As I am expecting a farewell to my PhD time, I would like to thank this journey and the time itself. I have learned and gained so much from this time. Finally, I deserve a “thank you” to myself. I thank the person four years ago who had the courage to start, the person two years ago who perservered, and the person now who has the confidence to face any challenge in her life in the future.

Düsseldorf, June 27, 2024

Contents

Acknowledgements **i**

1 Introduction **1**

1.1 Motivation . . . 2
1.2 Introduction to Blockchain Technology . . . 5
1.2.1 Concepts . . . 5
1.2.2 Applications . . . 7
1.3 Overview Of the Studies In the Dissertation . . . 9
1.3.1 Study 1: Systematic Literature Review . . . 9
1.3.2 Study 2: LDA for Patent Data . . . 11
1.3.3 Study 3: Blockchain in Cross-border Payments . . . 12
1.3.4 Study 4: DeFi Exploit Event Study with DID Models . . . 13
1.4 Additional Remarks . . . 15

2 How Are Texts Analyzed in Blockchain Research?
A Systematic Literature Review **19**

2.1 Introduction . . . 20
2.2 Research Methodology . . . 22
2.2.1 Definition of Research Questions . . . 22
2.2.2 Literature Search and Selection . . . 23
2.3 Descriptive Results . . . 25
2.3.1 Publication Trend . . . 25
2.3.2 Keyword Network . . . 27
2.3.3 Citation Ranking . . . 27
2.4 Discussion of Research Questions . . . 31

2.4.1 RQ1: Which research scope, text data, and methodology are used to conduct text analysis in the blockchain area? 31
2.4.2 RQ2: What topics are addressed using text analysis in current literature? 47
2.4.3 RQ3: What are the research gaps and promising future research topics? 53
2.5 Conclusion 55

3 The Landscape of Blockchain Innovation: A Textual Analysis of Patent Data 58
3.1 Introduction 59
3.2 Literature 61
3.3 Data and Methodology 62
3.3.1 Patent Data 62
3.3.2 LDA 63
Data Preprocessing 63
LDA Topic Modeling 64
3.4 Results and Discussions 67
3.4.1 Descriptive Results 67
3.4.2 Topic Modeling Results 67
Topic Distribution and Trends 67
Topic Focus of Enterprise Applications 78
3.4.3 Summary and Discussions 79
The Growing Number of Patent Filings 79
The Changing Trend of Topics 80
Differences Among Applicants 80
Research Gaps and Future Research 81
3.5 Conclusion 82

4 Blockchain for Cross-border Payments and Financial Inclusion in Africa: The Case of Stellar Network 88
4.1 Blockchain as a Solution for Cross-border Payment 89
4.2 Correspondent Banking and Its Drawbacks 90

4.3 Financial Inclusion and Blockchain in Africa 91
4.4 Introduction to Stellar Network 92
4.4.1 Consensus Protocol 92
4.4.2 Cross-border money transfer 93
4.5 Stablecoin: The Core of Blockchain-based Cross-border Payments 96
4.5.1 Classification of Stablecoins Depending on Stability Mechanism . 96
4.5.2 Stablecoins and cross-border payment 98
4.5.3 Adoption Challenges 99
Design and Standardization 99
Regulatory Framework 100
Social and Cultural Aspect 102
4.6 Localized stablecoin projects in African countries 104
4.6.1 Cowrie 104
4.6.2 Afreum 105
4.6.3 Uhuru 106
4.6.4 Leveraging Blockchain for Cross-border Payments 107
4.7 Concluding Remarks 108

5 Who cares about Decentralized Finance?
Evidence from DeFi Attacks **110**
5.1 Introduction 111
5.2 Categories of DeFi Exploit Events 113
5.3 Data and Methodology 116
5.3.1 Data 116
5.3.2 Methodology 118
5.4 Results and Discussion 119
5.4.1 TWFE DID regressions 119
5.4.2 DID event study regressions 119
5.4.3 Discussion 121
5.5 Conclusion 122

6 Concluding Remarks **127**
6.1 Main Findings and Contributions 128

6.2 Limitations and Future Research 130

A List of keywords for the query **132**

B Original outputs of the LDA model and author-generated topic labels **134**

Bibliography **136**

Eidesstattliche Erklärung **163**

List of Figures

1.1 Structure and relationship within the frame of the dissertation. 10

2.1 Flowchart of the literature selection phases. 25

2.2 Types and research areas of the publications in each year. 26

2.3 Keyword frequency and co-occurrence networks. 28

2.4 Connections among research scope, text data, and the methodology. . . . 33

3.1 Structure and relationship among four essays within the frame of the dissertation. 64

3.2 Coherence score of the LDA model from K=10 to K=100. 66

3.3 Distribution of patent filings in terms of applicant type, industry, and country. 69

3.4 Two-dimensional visualization of the blockchain-related topics from LDA topic modeling. 71

3.5 Trends of topics in years between 2009 and 2018 (Patent data for 2018 up to October 1, 2018). 72

3.6 Distribution of topics in pre-2015 periods. 84

3.7 Distribution of topics in post-2015 periods. 85

3.8 Trends in patent filed in years from companies in China and the United States between 2009 and 2018 (Patent data for 2018 up to October 1, 2018). 86

3.9 Topic distribution of startups and non-startups in China and the United States. 87

4.1 An Example of simplified cross-border payment process through correspondent banks. 91

4.2 An Example of simplified cross-border payment process on Stellar. . . . 96

5.1 Categories of DeFi Exploits. 114
5.2 Timeline of Crypto-related Exploit Events between 2012 and 2023. 117
5.3 DID event study regressions for all-event sample with estimation windows of 10 days and 15 days. 124
5.4 DID event study regressions for all-event sample with estimation windows of 10 days and 15 days. 125
5.5 DID event study regressions for all-event sample with estimation windows of 10 days and 15 days. 126

List of Tables

1.1 Summary of each study included in this dissertation. 16
1.2 Authorship, conference presentation, and publication status of each study included in this dissertation. 18
2.1 Top 10 most cited papers by total citation and citation per year. 29
2.2 Detailed information of text data sources used in the literature. 34
2.3 Detailed information of text analysis techniques used in the literature. . 39
3.1 Distribution of topics, categories, and relevant examples in each topic. . 70
5.1 TWFE DID regressions with estimation windows of 10 days and 15 days. 120

List of Abbreviations

Acronym	**Meaning**
AI	**A**rtificial **I**ntelligence
AML	**A**nti-**M**oney **L**aundering
BERT	**B**idirectional **E**ncoder **R**epresentations from **T**ransformers
BoW	**B**ag-**o**f-**W**ords
CBOW	**C**ontinuous **B**ag **O**f **W**ords
CeFi	**C**entralized **F**inance
CPPI	**C**ommittee on **P**ayments and **M**arket **I**nfrastructures
CPY	**C**itation **P**er **Y**ear
CFT	**C**ountering the **F**inancing of **T**errorism
DBDC	**D**igital **B**ank **D**igital **C**urrency
DeFi	**D**ecentralized **F**inance
DID	**D**ifference-**i**n-**D**ifferences
DLTs	**D**istributed **L**edger **T**echnologie**s**
DPoS	**D**elegated **P**roof-**o**f-**S**take
dApps	**D**ecentralized **A**pplications
DTM	**D**ynamic **T**opic **M**odels
EPO	**E**uropean **P**atent **O**ffice
ESMA	**E**uropean **S**ecurities and **M**arkets **A**uthority
ETH	**E**thereum
FBA	**F**ederated **B**yzantine **A**greement
Fintech	**Fin**ancial **Tech**nology
FSB	**F**inancial **S**tability **B**oard
GDPR	**G**eneral **D**ata **P**rotection **R**egulation
ICO	**I**nitial **C**oin **O**ffering
IEO	**I**nitial **E**xchange **O**ffering
IoT	**I**nternet **o**f **T**hings
IOSCO	**I**nternational **O**rganization of **S**ecurities **Co**mmissions
IPO	**I**nitial **P**ublic **O**ffering
ISO	**I**nternational **O**rganization for **S**tandardization
KYC	**K**now **Y**our **C**ustomer
LDA	**L**atent **D**irichlet **A**llocation
LM lexicon	**L**oughran and **M**cDonald Sentiment **Lexicon**
LSTM	**L**ong **S**hort-**T**erm **M**emory
MCAP	**M**arket **Cap**italization
MiCA	**M**arket **i**n **C**rypto-**A**ssets
NGN	**N**igerian **N**aira
NGNT	**N**igerian **N**aira **T**oken
NGO	**N**on-**G**overnmental **O**rganization
NFT	**N**on-**F**ungible **T**oken
NIBSS	**N**igerian **I**nterbank **S**ettlement **S**ystem
NLP	**N**atural **L**anguage **P**rocessing

OECD	**O**rganization for **E**conomic **C**ooperation and **D**evelopment
pBFT	**p**ractical **B**yzantine **F**ault **T**olerance
PoS	**P**roof-**o**f-**S**take
PoW	**P**roof-**o**f-**W**ork
PRISMA	**P**referred **R**eporting **I**tems for **S**ystematic Reviews and **M**eta-Analyses
RFID	**R**adio-**F**requency **Id**entification
SCP	**S**tellar **C**onsensus **P**rotocol
SEC	U.S. **S**ecurities and **E**xchange **C**ommission
SDEX	**S**tellar **D**ecentralized **Ex**change
SMEs	**S**mall and **M**edium-sized **E**nterprises
SSI	**S**elf-**S**overeign **I**dentity
STO	**S**ecurity **T**oken **O**ffering
SWIFT	**S**ociety for **W**orldwide **I**nterbank **F**inancial **T**elecommunication
TF-IDF	**T**erm **F**requency-**I**nverse **D**ocument **F**requency
TVL	**T**otal **V**alue **L**ocked
TWFE	**T**wo-**w**ay **F**ixed **E**ffects
UN	**U**nited **N**ations
VADER	**V**alence **A**ware **D**ictionary for S**E**ntiment **R**easoning
WoS	**W**eb **o**f **S**cience
XLM	Lumen
ZAR	South African Rand
ZCL	**ZCL**assic

Chapter 1

Introduction

1.1 Motivation

Since its introduction in 2008, blockchain has become a technology that goes far beyond bitcoin and cryptocurrency and is beginning to impact many sectors and aspects of society, including manufacturing, healthcare, government functions, and others. Its unique characteristics, such as decentralization, immutability, and transparency, have captured the attention of researchers and practitioners alike, as it has the potential to disrupt the status quo and transform industries in numerous sectors.

In the banking sector specifically, established intermediaries and banks dominate the market. Due to the extremely high entry barrier, it is almost impossible for new players to enter the market (Patel, Migliavacca, and Oriani, 2022). This preserves the credibility of the institutions. However, it causes limited competition and innovation, and most importantly, the control of personal information by large organizations. Blockchain enables more Fintech startups to leverage the technology to provide various financial services to customers, especially to those excluded from the formal banking system (Rühmann et al., 2020). Another example is in the healthcare sector, where the files of patients are established in an organization-centric manner, which means each healthcare facility (e.g., hospital and clinic) holds a portion of the medical history of the patient, and the systems are not interoperable with each other. This leads to various issues, including incomplete medical history, over-prescription, and medical data insecurity (Zhang et al., 2018). Blockchain technology offers a solution by allowing for decentralization and empowering individual patients to take control of their personal information, thereby promoting a more fair and user-centric approach (Gordon and Catalini, 2018).

Due to the novelty and rapid development of blockchain technology, despite the increasing amount of literature dedicated to this subject, there still remains a lack of comprehensive guidance for navigating the complex world of blockchain applications. Many areas within this field remain understudied, both from a holistic standpoint and within specific business sectors. The gaps are partially attributed to the lengthy publication processes but also come from the limited availability of data and methodology and a lack of understanding of the discrepancy between business and research agenda.

Attention should also be drawn from the innovator's point of view, especially young enterprises (i.e., startups), who are the backbone of blockchain development. On the one hand, they benefit from the technology and get the opportunities to enter the market with novel business ideas. On the other hand, they are the driving force behind technological innovation, pushing for progress and introducing new business models to the market. According to PitchBook (2023) and Financial Times (2023), the total amount of venture capital flew into crypto startups has increased dramatically from $2.2 Billion in 2017 to $3,112 Billion in 2022, and the investors are selecting projects more circumspect, searching for projects that facilitate the real-world business, rather than those that are solely speculative and virtual.

Established technology companies such as IBM and Amazon are indeed leveraging their resources and funding to develop comprehensive blockchain platforms. However, young enterprises are the pioneers in the process, taking the risks of immature new technology and transforming the business. Limited funding and resources also give these startups a unique approach to building solutions compared to established companies, as they tend to focus on niche markets and narrow market segmentation. They are the primary power of innovation, and their solutions tend to be more disruptive to the market (Barraza, 2019; Chen et al., 2019b). The market has long recognized the value of young firms in blockchain innovation, yet there has not been much research in the literature on this topic. Studies have touched on the topic of blockchain startups by identifying the most blockchain-innovative sectors and developing a taxonomy of the startup business models (Fiedler and Sandner, 2017; Friedlmaier et al., 2018; Beinke, Nguyen, and Teuteberg, 2018; Park and Sung, 2020), but more studies laying explicit focus on young firms are needed to address the essential roles of them in this area.

The objective of this dissertation is to provide a comprehensive overview of the blockchain business ecosystem landscape by addressing the research gaps through four distinctive studies, which apply different methodologies from different perspectives. The first study is a literature review that explores the possibilities of using text analysis techniques for blockchain-related studies. It widens the accessibility of using omnipresent text data as a source of information for blockchain-related research, thereby enabling more diverse studies. The second study employs one of the text analysis techniques to examine blockchain-related patent data to provide an overview of the innovation landscape, and identify the discrepancy of focus between business innovation cases and literature. Potential research directions are underlined to guide the subsequent two studies focusing on specific under-studied blockchain applications, namely the innovative applications of blockchain in financial sectors: cross-border payments and DeFi applications. The third study analyzes blockchain cross-border payment systems from a socio-technical perspective. It points out the opportunities and challenges of blockchain to be used in cross-border payment in African countries as a tool to increase financial inclusion. Specifically, case studies of Africa-based startups are conducted to identify their business approaches to address local challenges. The final study focuses on the security of the decentralized finance (DeFi) platform. DeFi applications are one of the most innovative adoption of blockchain in the financial sector, mostly developed by startups that are the driving force in the open-financial blockchain ecosystem. However, they suffer from inherent vulnerabilities due to financial exploits and attacks which presumably reduce their credibility. This study applies stacked difference-in-differences (DID) regressions to DeFi exploit events to examine the impacts of malicious attacks and security failures on the native assets of the platforms. The study considers attacks on DeFi protocols as a potential source of systematic risk to the overall blockchain-specific ecosystem, especially if its native asset prices move after a single DeFi protocol attack.

This dissertation represents a modest attempt to explore and facilitate the research agenda development of the newly established field of blockchain research from a business perspective. Given that blockchain research lies at the intersection of multiple disciplines, it is imperative to employ diverse methodologies and integrate both technical and business/economic understanding to fully comprehend the technology and its impact on society. Drawing from industry and literature, this dissertation provides novel perspectives to approach unresolved questions and contributes to advancing the academic understanding of the rapidly evolving blockchain ecosystem. A detailed introduction and contributions of each study included in this dissertation are provided in Section 1.3 Overview Of the Studies In the Dissertation.

1.2 Introduction to Blockchain Technology

Blockchain is considered one of the most disruptive technological innovations after the Internet (Swan, 2015). Brought to prominence by Nakamoto (2008) through Bitcoin, it combines several pre-existing technologies and provides a secure and transparent method of recording and managing information over a peer-to-peer network. This section provides the fundamental background of blockchain technology and its applications[1].

1.2.1 Concepts

Blockchain is a type of *Distributed Ledger Technologies (DLTs)*, which is fundamentally a decentralized database that allows all network participants to read and write data without a central, controlling authority. Every member can input data at any time, which then undergoes a confirmation process. The latest data state is always shared with all network participants. A blockchain consists of four fundamental elements that give it its unique characteristics.

Node Network participants are called *nodes* on a blockchain. They are responsible for diverse tasks that maintain the functioning of the network. First, they validate the transactions, deciding if they should be accepted and added to the block (distributed ledger). Second, they maintain the records on the blockchain. Third, they broadcast and distribute the accepted transactions to other nodes on the network to keep all the participants with updated information. The methods nodes use to conduct the tasks are determined by the consensus protocols of the specific networks. For example, miners are a particular type of node in the Proof-of-Work (PoW) protocol (i.e., Bitcoin network) that adds new records by solving computational problems.

Distributed ledger A *distributed ledger* is a database that compiles all transactions created, maintained, and updated by nodes in a blockchain network and is shared and synchronized across the network (Pustišek, Živic, and Kos, 2022). The distributed nature ensures that the process is done through the joint work of the nodes without a

[1] A thorough introduction to blockchain technology is beyond the scope of this introductory chapter. Please refer to Antonopoulos (2017), Swan (2015), or many other internet sources and developer documentations for more information.

central controlling authority. In a blockchain, blocks make up the ledgers, and each block includes a list of transactions. Once the capacity of the block is reached, it is sealed and then linked to the previous block through a cryptographic hash pointer, which ensures that the information in the block is interconnected with the previous blockchain and cannot be tampered with. This ongoing process ultimately creates a chain of blocks, giving the blockchain its name.

Hash A *hash* is a random sequence of characters that typically appears as the output of a hash function. A hash function takes input (i.e., transaction data) and returns a hash. The function is specifically designed to be one-way and deterministic, meaning that the same input will always produce the same output. However, it is virtually impossible to reverse-engineer the hash output to derive the original input. It is also highly unlikely that different inputs will produce the same hash output.

In a blockchain, the hash ensures the authenticity of the data within a block. In between the blocks, hash pointers are used to prevent the data from being tampered with. A hash pointer is a data structure that includes a reference to the previous block along with the hash of the contents in the current block. This design ensures that the blocks are connected in a chain, and any modifications in a block will alter the hash and break the chain.

Consensus Protocol *Consensus Protocol* is a very crucial component of blockchain because it establishes the underlying rules that determine how nodes reach agreements, including which transactions are added to the distributed ledger, without the need for any central authority (Pilkington, 2016). The design of the consensus protocol is specific to each blockchain (Irresberger et al., 2023). There are many types of Consensus Protocols, the most common ones being: a) Proof of Work (PoW). Used by Bitcoin, it involves solving complex mathematical puzzles to validate transactions. It requires significant computing power and is energy-intensive. b) Proof of Stake (PoS). Used by Ethereum, where nodes are selected to validate transactions based on the number of coins they hold. c) Delegated Proof of Stake (DPoS). It is a variation of PoS where coin holders vote for a selected number of delegates to validate transactions.

Despite the standard structures of blockchains, their component-based designs can be different. The basic designs of blockchain can all be constructed with distinct characteristics to fit the specific implementations (Tasca and Tessone, 2019). Blockchain can be categorized by two criteria (Peters and Panayi, 2016; Nærland et al., 2017): (1) Access to transactions. Public blockchains allow all nodes to read and submit transactions; private blockchains restrict this right to authorized nodes. (2) Access to transaction validation. Permissionless blockchains allow all nodes to validate transactions; permissioned blockchains have pre-selected nodes for validation.

Another categorization like the following is more commonly used (Mougayar, 2016). It could also be explained by the two criteria mentioned above:

- **Public Blockchain**: It corresponds to the public permissionless blockchain. All nodes have the right to read and validate transactions. Bitcoin and Ethereum are public blockchains.

- **Consortium (Federated) Blockchain**: It is permissioned, but can be public or private depending on the settings. In a consortium blockchain, only authorized parties can validate transactions, and it can be open to the public or restricted to certain parties. Many consortium blockchain networks are built to bring together entities in the same industry, allowing for more interaction and collaboration.

- **Private Blockchain**: This type of blockchain is a strict form of private permissioned blockchain, where typically only one party has the right to validate transactions and only certain parties have access to transactions. The level of decentralization in private blockchains is limited, but comes with higher efficiency. Private blockchains are mainly used as internal systems within a company.

1.2.2 Applications

Based on the developing process, blockchain applications can be divided into three phases, Blockchain 1.0, 2.0, and 3.0, each characterized by their use cases and capabilities (Pustišek, Živic, and Kos, 2022). A more specific and detailed introduction to blockchain applications will be presented in the studies included in this thesis.

Blockchain 1.0 The first generation of blockchain focuses primarily on cryptocurrencies. It involves the deployment of cryptocurrencies related to digital payment, currency transfer, and remittance. It is the most straightforward implementation of blockchain technology.

Blockchain 2.0 The second generation of blockchain is characterized by its integration with smart contracts. It is essentially a computer-coded contract on blockchain which is automatically executed when the contract terms are met. This increases the enforceability of business contracts without the involvement of a trusted third party (Cong and He, 2019). The programmability of smart contracts allows for the automation of complex transaction processes, which greatly expands the range of blockchain applications. In particular, smart contracts allow developers to build decentralized applications (DApps) on the blockchain, including DeFi platforms and supply chain management, many of which go beyond currency exchange. Additionally, smart contracts facilitate tokenization, which involves the creation of digital tokens that represent various assets or utilities. This includes the use of non-fungible tokens (NFTs).

Blockchain 3.0 Blockchain 3.0 is an evolutionary generation of blockchain that aims to build integrated blockchain ecosystems and tackle some major challenges of previous generations. By incorporating other cutting-edge technologies such as the Internet of Things (IoT) and Artificial Intelligence (AI), Blockchain 3.0 is able to expand its application into diverse sectors such as healthcare, culture, and government functions, resulting in more comprehensive and sophisticated systems.

It also addresses several challenges that hinder blockchain adoption. Unlike centralized systems, blockchains lack standardization due to variations in several parameters. This lack of standardization often impedes information exchange among systems, limiting blockchain's ability to facilitate collaboration among enterprises or industries that use different blockchains (Buterin, 2016). Increasing *interoperability* enables cross-blockchain communication, thereby extending blockchain functionality and boosting cross-sector blockchain adoption.

Scalability is another essential challenge. Scalability describes a system's ability to handle increasing numbers of elements and smoothly process the growing workload

(Bondi, 2000). Various solutions have been proposed to address scalability issues, including data propagation, on-chain, and off-chain solutions. Data propagation solutions focus on optimizing information broadcasting. The on-chain solutions mainly focus on blockchain data, consensus mechanism, and sharding (Zhou et al., 2020). In contrast, off-chain solutions solve the problem by adding other channels or communicating two or more blockchains (Kim et al., 2018), which requires interoperability to allow heterogeneous blockchains to work together.

1.3 Overview Of the Studies In the Dissertation

This dissertation consists of four distinct studies that focus on different aspects of the blockchain innovation ecosystem. Yet, they are cohesively interconnected. Figure 1.1 illustrates the structure and the relationship of the four studies within the framework. It begins with Study 1, a systematic literature review that explores the possibilities of using computer-based text analysis for blockchain-related research. Using the methodology and research gaps identified in the literature review, Study 2 applies the Latent Dirichlet Allocation (LDA) topic model to blockchain-related patent data to pinpoint the topics discussed in blockchain innovation and provides an overview of the ecosystem. Based on the research gaps identified in Study 2, Studies 3 and 4 explore blockchain applications in cross-border payments and the impact of DeFi exploit events on blockchain platforms, respectively. The detailed objectives, research questions, and their interrelationships are elaborated in this section. A summary of each study can be found in Table 1.4.

1.3.1 Study 1: Systematic Literature Review

While empirical research primarily relies on numerical data like network metrics and cryptoasset prices from public blockchains, this approach faces challenges in consortium and private blockchains due to limited data accessibility.

Text-based analysis in blockchain contexts allows the researchers to delve into both the metadata and the actual content of blockchain data and make inferences that could not be made before with numbers alone. The main challenge in using text analysis is the substantial manual effort and time required to process large volumes of text,

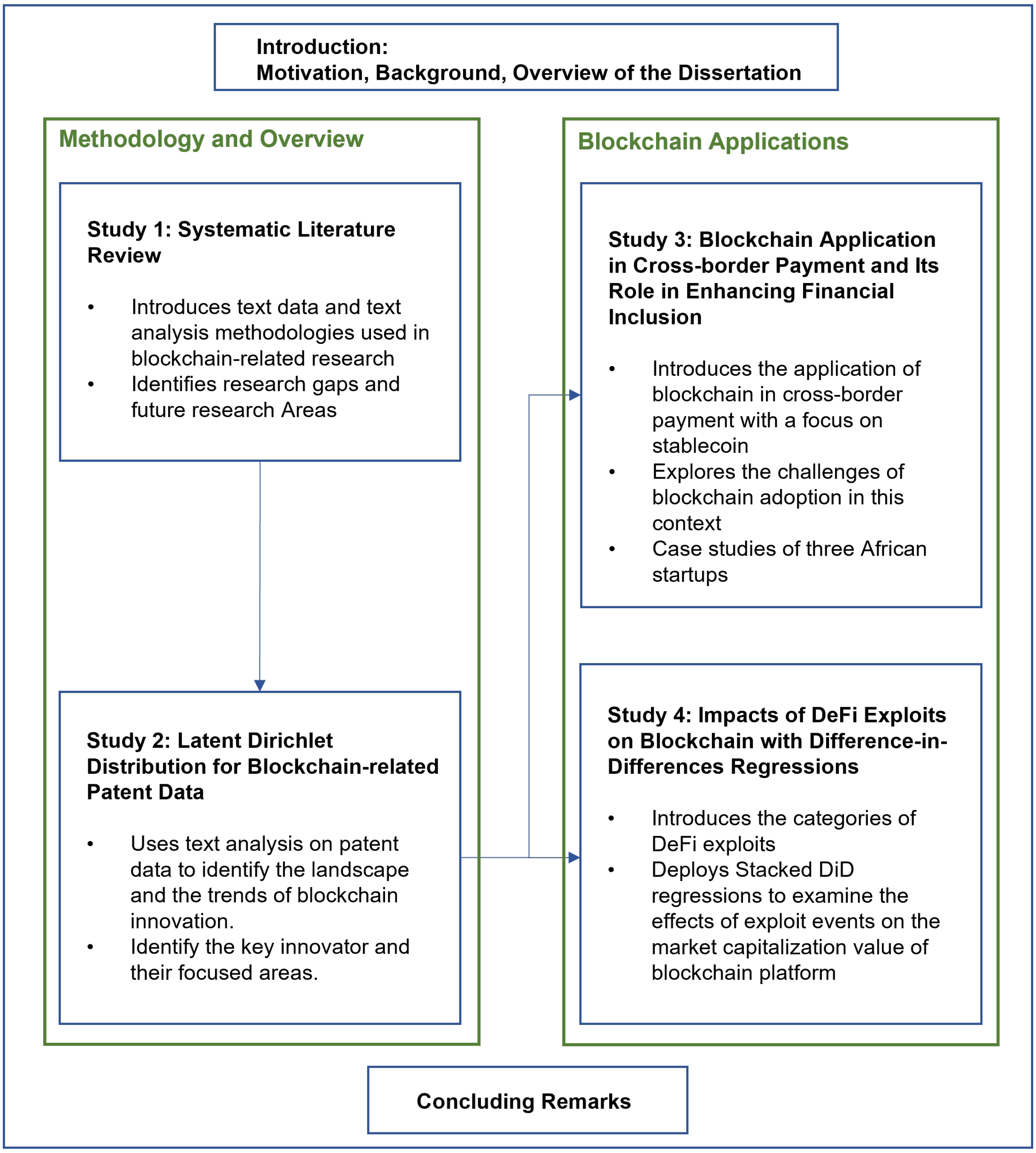


FIGURE 1.1: Structure and relationship within the frame of the dissertation.

which can be solved by computer-based text analysis. Despite the potential to solve the problems, such reviews have not been conducted in blockchain research. This review, therefore, aims to synthesize the current knowledge in the literature by examining published and unpublished academic literature on text analysis related to blockchain topics across disciplines to understand the relevance and potential of text analysis.

Specifically, it addresses the following research questions:

- *Which research scope, text data, and methodology are used to conduct text analysis in the blockchain area?*

- *What topics are addressed using text analysis in current literature?*
- *What are the research gaps and promising future research topics?*

This review makes several contributions to the literature by addressing the above research questions. First, it provides comprehensive summaries of the research scope, text data sources, and text analysis methodologies in the existing literature. Second, it goes beyond individual elements and exhibits the connections between them. It underscores the importance of selecting suitable combinations of data characteristics and research questions from various angles. Finally, it integrates blockchain-related research areas and text analysis approaches into a joint framework. It underlines five key research areas identified in the literature: relationship discovery, cryptocurrency performance prediction, classification and trend, crime and regulation, and the perception of blockchain. The review also introduces three emerging research directions: improvement of data preparation, studies with underused data and growing areas, and regulation-related research. It is helpful for researchers from various disciplines interested in leveraging large-scale text data for blockchain-related studies.

1.3.2 Study 2: LDA for Patent Data

Despite the broad potential adoptions of blockchain across numerous sectors, there is a lack of standardization in the design of blockchain - they vary widely in economic design and technical implementation, making it difficult to draw a guiding framework to navigate the landscape of blockchain-related innovation. Such a framework could inform business research and innovation agendas by highlighting topics of practical relevance or the lack thereof.

The findings of Study 1 suggest that an unsupervised machine learning such as LDA is an ideal methodology for analyzing the text of emerging technologies, and patent data is one of the insightful yet understudied text data used in the literature. Leveraging these insights, this paper aims to draw a landscape of blockchain innovation by applying LDA to the text of blockchain-related patent filings to create topic models and identify subtopics within the broader context of blockchain innovation.

This study extends the analysis beyond a static overview of blockchain innovation, focusing on several research questions:

- *What does the blockchain innovation landscape look like from a business perspective, and how has it changed over time?*
- *Who are the innovators, and what are the different approaches to blockchain technology among the innovators?*
- *What are the discrepancies in the literature and business regarding blockchain innovation, and where are the gaps?*

In addressing these questions, the study sheds light on overlooked areas in current blockchain literature and proposes multiple directions for future research. First, while blockchain design is a prominent theme in applications, its integration with business issues is often overlooked in business and management literature. Future studies could take a more integrative approach to exploring business applications. Second, the literature underscores the cryptocurrency applications of blockchain in the financial sector, overlooking its broader range of business applications. This opens avenues for exploring the diverse uses in financial sector. Finally, the role of startups in driving blockchain innovation deserves deeper investigation, particularly their business models and their impact on the progress of blockchain adoption in business contexts. Overall, this study offers valuable guidance for researchers in pinpointing future research areas and provides practical insights for practitioners about their business and innovation prospects in blockchain-associated sectors.

1.3.3 Study 3: Blockchain in Cross-border Payments

The results of Study 2 suggest that 1) blockchain applications in the financial sector have reached numerous areas, but the literature has been focused on cryptocurrency, and 2) blockchain design and business applications can be examined more integrated.

One of the system-relevant applications of blockchain, which has the potential to benefit many underprivileged people, especially in developing countries, yet is less explored, is the application in cross-border payments. The heterogeneous payment schemes in different jurisdictions, opaque processes, limited access to financial services, and a lack of competition have kept the conventional cross-border payment system with high cost and long settlement time for many developing countries with exotic currencies (Rice, Peter, and Boar, 2020). Blockchain can disrupt this sector by offering connectivity, transparency, and lower barriers to entry. This enables technology startups to provide more accessible and affordable financial services, helping individuals and businesses in developing countries, often excluded from formal financial systems, to securely manage and transfer their wealth internationally.

However, the current literature in the business area focuses more on the overall benefits and implications of blockchain for cross-border payments, rarely examining specific mechanisms or business models. Additionally, discussions about blockchain and stablecoins are often skewed towards developed countries and major currencies like the US dollar.

This study aims to bridge these gaps by concentrating on African countries with the case of Stellar network to shed light on the following topics:

- *Introducing a blockchain consensus mechanism tailored for cross-border payment applications.*
- *Classifying stablecoin and identifying the challenges in adopting stablecoin for cross-border payments.*
- *Providing business insights through case studies of three startups offering localized solutions for regional needs.*

The results of this study highlight that achieving the full potential of blockchain in enhancing financial inclusion requires innovative approaches to address the unique regional challenges. Such outcomes can only be realized through the collaborative efforts of governments, financial institutions, educational institutions, NGOs, and other stakeholders. Overcoming the technical, regulatory, and social challenges is key to unlocking the transformative power of blockchain in cross-border payments.

1.3.4 Study 4: DeFi Exploit Event Study with DID Models

Based on the findings of Study 2, we further explore blockchain applications within the financial sector, with a particular emphasis on the DeFi ecosystem and platform security. It is one of the focuses of the burgeoning generation of blockchain applications, which has gained economic relevance and attracted substantial capital deployment (Irresberger et al., 2023).

However, the rapid growth of DeFi applications has also led to an increase in attempted attacks for financial gain. Although the decentralized structure of DeFi enables the non-custodial access to complex financial protocols, giving DeFi advantages over centralized institutions, it also renders the underlying codes vulnerable to exploitation. These exploits, if not appropriately addressed, can undermine trust in the security of blockchain and adversely affect the value of its native cryptoasset. Despite being a crucial issue for blockchain, few studies have explored the effect of exploit events at the native token level. While there exist studies elaborating the categories and potential influences of DeFi attacks (Hornuf et al., 2023; Qin et al., 2021; Zhou et al., 2023), the quantitative measurement of the impact of the events at the native token level is scarce.

This study aims to shed light on the impact of DeFi exploit events on the valuation of native tokens using DID models. *It seeks to answer the question of whether these attacks lead to a positive or negative impact on the underlying native cryptoasset of a blockchain.* Instead of analyzing individual DeFi tokens, it focuses on the valuation of the native cryptoasset to understand the impact of vulnerabilities in decentralized financial applications (protocols) on the underlying smart contract blockchain.

The key finding is that when major DeFi protocols on a blockchain experience attacks that lead to financial losses for their users, the valuation of the native cryptoassets increases after the events. This suggests that the blockchain infrastructure views these incidents as positive developments, potentially attracting more new users, encouraging the development of more robust smart contract code, conducting thorough audits, and providing confidence that the developer community can effectively manage individual economic or technical exploits of blockchain-based applications.

This study can serve as a valuable foundation for future research in the area of DeFi attack and security, allowing researchers to delve deeper into the relationship between exploit incidents and the underlying blockchain infrastructure, as well as the attribution of their impact. In addition, the paper provides insight into how investors and the broader DeFi community perceive these exploits, highlighting implications for practitioners and regulators in terms of a nuanced understanding of how such incidents can influence market perceptions and stakeholder behavior.

1.4 Additional Remarks

The four studies included in this dissertation were crafted in separate research projects, so they are at different stages of development at the time of this writing. Table 1.2 lists the authorship, the conference presentation, and the publication status of each study.

TABLE 1.1: Summary of each study included in this dissertation.

Study	Research Questions	Data and Methodology	Results and Contributions
Study 1 (Chapter 2): How Are Texts Analyzed in Blockchain Research? A Systematic Literature Review	Which research scope, text data, and methodology are used to conduct text analysis in the blockchain area? What topics are addressed using text analysis in current literature? What are the research gaps and promising future research topics?	124 published or unpublished academic papers include specific keywords in the areas of blockchain and text analysis. Systematic Literature Review following the guidelines of Siddaway et al. (2019) and the PRISMA statement.	Provides comprehensive summaries of research scope, text data sources, and text analysis methodologies in the existing literature to guide researchers in finding pertinent resources. Goes beyond individual elements and exhibit the connections between them. We emphasize that it is crucial to choose appropriate combinations considering variable perspectives. Integrates blokcchain-related research areas and text analysis approaches into a joint framework. By not restricting our search to one discipline, we are able to capture the use of text analysis in non-technical blockchain studies across disciplines and provide multiple perspectives on the topic
Study 2 (Chapter 3): The Landscape of Blockchain Innovation: Evidence from Patent Data	What does the blockchain innovation landscape look like from a business perspective, and how has it changed over time? Who are the innovators, and what are the different approaches to blockchain technology among the innovators? What are the discrepancies in the literature and business regarding blockchain innovation, and where are the gaps?	Blockchain-related patent families from the European Patent Office (EPO) worldwide database between 2009 and 2018. Latent Dirichlet Allocation (LDA) text anlaysis topic model.	Uses novel text analysis methodology for business-relevant patent filings identify the sub-topics within blockchain to give an overview of the blockchain innovation. Additionally, I provide no only the static landscape, but the evolvements of the topic and the innovation characteristics of different innovators. The patterns of blockchain applications and research gaps I identify can be used as a starting point for researchers to explore the related potential research areas in the blockchain area and be used by practitioners to distinguish their potential business opportunities.

Study 3 (Chapter 4): A Connected Future: Blockchain for Cross-border Payments and Financial Inclusion in Africa	Introduces a blockchain consensus mechanism tailored for cross-border Classifes stablecoin and identifies the challenges in adopting stablecoin for cross-border payments. Provides business insights through case studies of three startups offering localized solutions for regional needs.	Case studies applied to three business cases on Stellar network.	Discusses a payment mechanism superior to traditional systems, detailing stablecoin concepts and associated challenges. Presents case studies of three Africa-centric blockchain startups, offering insights into practical applications and five key strategies for business model development: local stablecoin, easy-to-use channel, specific target market, special product, and corporations. Stresses the crucial role of collaborative efforts of governments, financial institutions, educational institutions, NGOs, and other stakeholders to overcome the technical, regulatory, and social challenges which will unlock the potential of blockchain in cross-border payments.
Study 4 (Chapter 5): DeFi Attacks and Blockchain Application Ecosystem	Seeks to answer the question of whether these attacks lead to a positive or negative impact on the underlying native cryptoasset of a blockchain.	All exploit events listed in the REKT database between September 1, 2012 and June 1, 2023 with more than $10 million in financial losses: 143 exploit events, including 56 CeFi events and 87 DeFi events affecting 11 chains. Two-way fixed effects (TWFE) DID regressions and (TWFE) DID event study regressions.	Discovers that when major DeFi protocols on a blockchain experience attacks, the valuation of the native cryptoassets increases after the events, suggests that the blockchain infrastructure views these incidents as positive development. Serves as a valuable foundation for future research in the area of DeFi attack and security. Provides insight into how investors and the broader DeFi community perceive these exploit.

TABLE 1.2: Authorship, conference presentation, and publication status of each study included in this dissertation.

Study	**Authorship and Distribution**	**Conference Presentations**	**Revision and Publication**
Study 1 (Chapter 2): How Are Texts Analyzed in Blockchain Research? A Systematic Literature Review	Xian Zhuo 60% Felix Irresberger 20% Denefa Bostandzic 20%	29. – 31. July 2022, 3rd International Symposium in Finance, Kissamos, Crete, Greece	Published at *Financial Innovation* Zhuo, X., Irresberger, F. & Bostandzic, D. How are texts analyzed in blockchain research? A systematic literature review. Financ Innov 10, 60 (2024). https://doi.org/10.1186/s40854-023-00501-6
Study 2 (Chapter 3): The Landscape of Blockchain Innovation: Evidence from Patent Data	Xian Zhuo 100%	9. April, 12. – 13. April 2021, British Accounting and Finance Association Annual Conference, Online 10. – 12. December 2021, European International Business Academy, 47th Annual Conference, Madrid, Spain 17. – 18. December 2021, World Finance and Banking Symposium, Budapest, Hungary	Unpublished working paper
Study 3 (Chapter 4): A Connected Future: Blockchain for Cross-border Payments and Financial Inclusion in Africa	Xian Zhuo 60% Felix Irresberger 20% Denefa Bostandzic 20%		Unpublished working paper, manuscript shared on SSRN Zhuo, X., Irresberger, F., Bostandzic, D., Building a Connected Future: The Role of Blockchain in Cross-border Payments and Financial Inclusion in Africa (May 20, 2023). http://dx.doi.org/10.2139/ssrn.4550837
Study 4 (Chapter 5): DeFi Attacks and Blockchain Application Ecosystem	Xian Zhuo 60% Felix Irresberger 20% Denefa Bostandzic 20%		To be included as a book chapter in *Handbook of Blockchain Analytics* (Springer).

Chapter 2

How Are Texts Analyzed in Blockchain Research? A Systematic Literature Review

This chapter is published at Financial Innovation: Zhuo, X., Irresberger, F. & Bostandzic, D. How are texts analyzed in blockchain research? A systematic literature review. Financ Innov 10, 60 (2024). https://doi.org/10.1186/s40854-023-00501-6

2.1 Introduction

Blockchain technology and its economics have attracted considerable attention from academic researchers. The total volume of research has increased dramatically, with the proportion of empirical studies growing gradually in recent years (Casino, Dasaklis, and Patsakis, 2019; Xu, Chen, and Kou, 2019; Frizzo-Barker et al., 2020). Data availability is often a primary obstacle in empirical studies in emerging research areas, such as blockchain, where it is not clear which alternative data sources should or can be used for quantitative analysis. Owing to its nature, a blockchain primarily comprises numerical data such as on-chain transactions by users or network (value) metrics, trading activity, price data of cryptoassets, or financial reports of the few available companies, most of which are readily available in the public blockchain. However, these datasets can be complemented by text data to obtain more data from consortiums and private blockchains, thus expanding the research span and deriving additional relevant insights.

Given the decentralized nature of the public blockchain ecosystem, there are limited compulsory disclosures or official platforms representing the comprehensive information of single blockchain projects that can serve as sources of blockchain-related information. Alternative sources of textual data play a vital role for different parties in gathering information and making decisions within a blockchain network. For example, the sentiments of a crowd (via news, social media, or other text sources) may be a more relevant reference for investment in the blockchain ecosystem than in corporations. Such data can affect the market, influence investors' decisions, and provide an impetus for blockchain development. Researchers can make use of texts in blockchain-related contexts to obtain information in the data from more perspectives (i.e., explore not only the metadata describing the data but also the actual content of the data) and make inferences that cannot be made before with only numbers.

Therefore, in this study, we focus on providing an overview of text analysis methodologies and data sources as they pertain to blockchains, which differ from the text-based analyses of corporations. There is no consensus on the type of text data that should or could be used to analyze a specific blockchain network or project; therefore, our systematic overview helps alleviate this concern.

Several types of blockchain-related text data are publicly available. First, blockchain is a frequent topic in news articles reporting, with subtopics including the performance of cryptocurrencies and the latest developments in the technology. Second, because of the technical nature of blockchain technology, online platforms or forums such as Twitter, GitHub, and Reddit have been actively used by different groups (e.g., investors and developers) to express their opinions and share and track new developments (Mendoza-Tello et al., 2018). Blockchain startups also use social media for marketing. Third, blockchain project whitepapers provide key information (e.g., technical and marketing) to potential investors and are the primary method for understanding project details (Cohney et al., 2019).

In all these cases, manual examination of large-scale text content is exceptionally labor-intensive and time-consuming, if not impossible. Hence, computer-based text analysis is essential. Researchers across disciplines have provided guidelines for using such type of approaches. Grimmer and Stewart, 2013, for example, illustrate the promise and the pitfalls of text analysis for political science. Günther and Quandt, 2016 give a comprehensive overview of text analysis methods useful in digital journalism research. Studies in economics and finance have addressed the advantages and disadvantages of different methodologies (Loughran and McDonald, 2016; Cong et al., 2021; Gentzkow, Kelly, and Taddy, 2019).

Such reviews have not been conducted in blockchain-related research areas, despite the close connection between blockchain technology and multiple text datasets. Therefore, we argue that it is necessary to use a transparent approach and an academic standpoint to synthesize the current knowledge in the literature to better understand the relevance and potential of text analysis. In this study, we conduct a systematic literature review by examining published and unpublished academic literature, focusing on text analysis associated with blockchain topics across disciplines. We provide the fundamental principles and relevant sources of text analysis methodologies and connect the relationships of research scopes, text data, and methodologies to provide researchers with a reference for choosing suitable combinations of the above elements with respect to their research question at hand. We then pinpoint the specific research topics studied in the literature and propose directions for future research. This review serves as a guide for researchers from different disciplines interested in conducting blockchain-related text analysis studies.

2.2 Research Methodology

We conduct a systematic review of the academic literature on blockchain-related research using text analysis. Research in this area has expanded because of the rapid development of blockchain technology. However, because of the interdisciplinary nature of blockchain research, research perspectives vary starkly, posing difficulties in searching for and gathering knowledge beyond a single field. We focus on computer-based text analysis used in blockchain research to comprehend and synthesize studies across disciplines that utilize text analysis as a primary or ancillary methodology. We aim to gain knowledge from the existing literature in this area and discover future research opportunities. We adopt the guidelines of Siddaway, Wood, and Hedges, 2019 and the PRISMA statement (Liberati et al., 2009; Moher et al., 2009; Page et al., 2021b; Page et al., 2021a).

2.2.1 Definition of Research Questions

The first stage of a systematic review involves defining research questions that guide subsequent actions. We propose the following research questions to achieve the objectives of our review:

RQ1: Which research scope, text data, and methodology are used to conduct text analysis in the blockchain area?

Both blockchain and text analysis are broad concepts. This question is designed to identify the specific scope of the studies (e.g., cryptocurrency[1], smart contract[2]), the text data being analyzed (e.g., social media posts and news), and specific methodologies or techniques used to perform the analyses (e.g., sentiment analysis). We aim to bridge

[1]The first use case for blockchains is the creation of cryptocurrencies (e.g., Bitcoin), where Nakamoto (2008) proposed a design for a decentralized payment system in which all transactions are stored in transparent blocks, and transactions are validated through a consensus protocol. The idea is to build trust through protocols and operate the system without authority (i.e., a trusted third party).

[2]A smart contract is essentially a computer-coded contract on blockchain that is automatically executed when the contract terms are met. This increases the enforceability of business contracts without the involvement of a trusted third party (Cong and He, 2019).

and highlight the connections between these elements in each study. This will assist researchers in selecting the appropriate data and methodologies for their research.

RQ2: What topics are addressed using text analysis in current literature?

The research questions determine how the research develops, and text analysis is one of the methods used to serve the purposes of a study. Regardless of whether text analysis is used alone or as part of a broader analysis, we intend to provide an interdisciplinary overview of the topics and research questions addressed in the existing literature, and illustrate how text analysis contributes to the study of these topics.

RQ3: What are the research gaps and promising future research topics?

Based on the findings of our review, we identify understudied areas and future research opportunities using text analysis in blockchain research. This allows researchers to recognize promising research topics and specify the methodologies (and data) they can use.

2.2.2 Literature Search and Selection

Initial keyword searches were conducted on May 24, 2022, followed by updated searches on August 23, 2022, to find relevant studies. We chose the Web of Science (WoS) and Scopus databases to cover publications indexed in academic databases. As text analysis in blockchain research is relatively new, some studies may not have been published. Therefore, we also performed a keyword search of the Social Science Research Network (SSRN) to distinguish unpublished papers (e.g., working and discussion papers) (Garanina, Ranta, and Dumay, 2021). Subsequently, backward snowballing of the articles obtained through keyword searches was performed to identify additional articles.

For a comprehensive result, our query keywords encompassed not only *blockchain* and *text analysis* but also synonyms and multiple specific topics relevant to the area. Relevant words from blockchain included *blockchain, cryptocurrency, stablecoin*[3]*, crypto token, smart contract, initial coin offering (ICO), security token offering (STO), and initial*

[3]Stablecoins are cryptocurrencies designed to be price-stable by pegging their values to a specific asset (or a basket of assets), making them a better medium of exchange than typical cryptocurrencies. The most common peg is to the US dollar.

exchange offering (IEO)[4], and *non-fungible token (NFT)*[5] Keywords from text analysis included *text analysis, textual analysis, text analytics, topic modeling, natural language processing (NLP), word embedding, sentence embedding, bag of words* , and *sentiment analysis*. We also used asterisks (*) and quotation marks ("") to eliminate the impacts of plural forms, hyphens, or spelling variations. A description of our keyword-selection process and a complete list of keywords are included in the Appendix.

Keywords were searched in the title, abstract, and keywords [6]. The exact query is as follows:

(blockchain* OR cryptocurrenc* OR stablecoin* OR "crypto token*" OR "smart contract*" OR "initial coin offering*" OR "security token offering*" OR "initial exchange offering*" OR "non*fungible token*") AND ("text* analysis" OR "text analytics" OR "topic model*" OR "natural language processing*" OR "word embedding*" OR "sentence embedding*" OR "bag of words" OR "sentiment analysis")

The details of the literature search and selection process are presented in Figure 2.1. Search queries in the two databases returned 517 records. First, we screened the metadata of the articles to remove articles that were 1) non-English articles, 2) notes, editorials, conference proceedings titles, and preliminary papers, 3) duplicates, and 4) without full-text access. We screened the titles and abstracts to remove articles based on our content-based exclusion criteria. To obtain relevant articles from multiple perspectives, we did not set inclusion/exclusion criteria by discipline. Alternatively, we checked the content of the articles and only excluded an article if 1) it did not contain information related to both blockchain and text analysis, 2) it focused purely on the technical aspect of blockchain, or 3) it did not specify the specific text analysis techniques used. After the above screening, 140 articles remained for full-text assessment, and we applied the exclusion criteria again and obtained 99 published articles.

[4]ICO is an alternative way of financing projects or startups by creating and issuing tokens on a blockchain and selling them to raise funds. IEOs can be seen as an ICO supervised by cryptocurrency exchange platforms: the project goes through due diligence before commencing the sale, which gives investors more assurance about the validity and success of the project. STOs are tokenized digital securities and are sold in security token exchanges. They are classified as securities and are subject to rigorous vetting before issuance.

[5]NFTs differ from other tokens by its non-fungibility. A token can represent ownership of a specific item (e.g., painting, land) and is not interchangeable with other tokens because it has unique (digital) properties encoded in the smart contract that creates it.

[6]For WoS, we also searched in Keywords Plus. It is a feature of WoS that returns the articles in results if the words or phrases in our search appear frequently in the titles of these articles' references, but not in the title of the article itself. By doing this, we also collected articles that have the potential to be relevant to our topic but did not have the keywords placed in the article.

Our search on SSRN initially returned 30 articles. We removed 24 articles based on our exclusion criteria, leaving six unpublished articles. Subsequently, we conducted backward snowballing on 105 articles included in the keyword searches (i.e., we went through the references of the included articles) to find additional articles that did not appear in the keyword searches. This process yielded nineteen additional 19 papers. A total of 124 studies were included in the literature review.

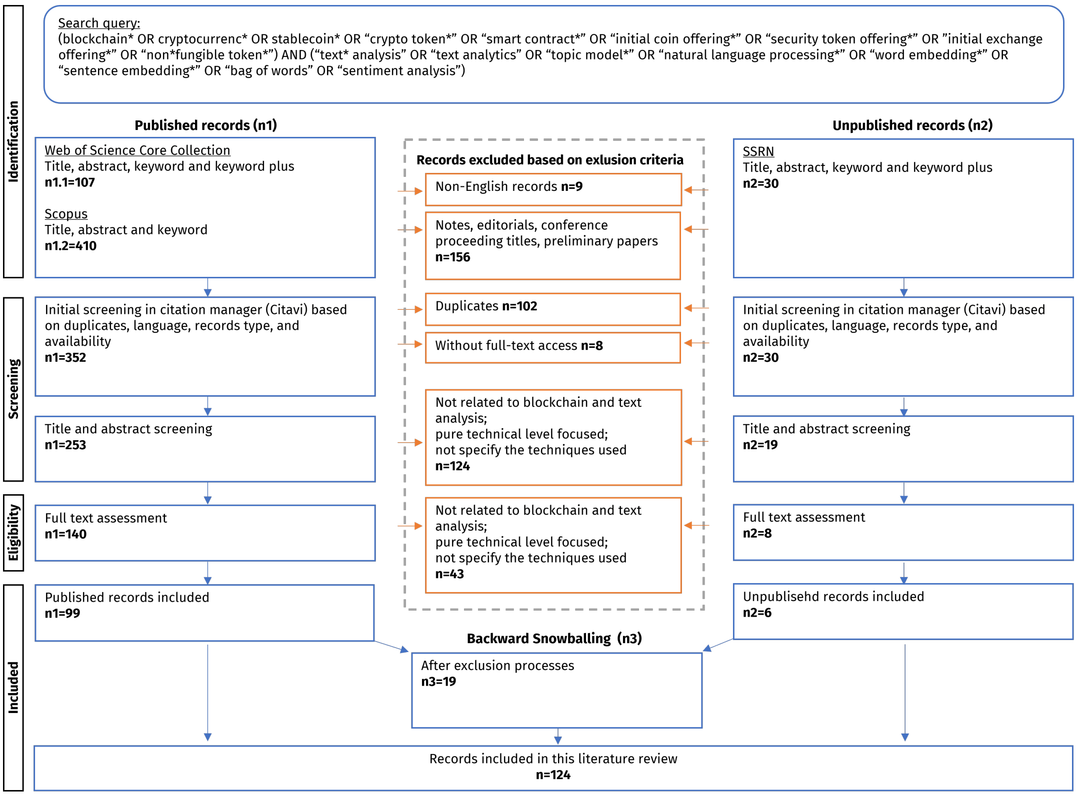


FIGURE 2.1: Flowchart of the literature selection phases.

2.3 Descriptive Results

This section reports the descriptive results of the papers, including publication trends, keyword networks, and citation rankings.

2.3.1 Publication Trend

Figure 2.2 depicts the number of papers on a yearly basis subject to article type and research area. Although we did not set any timeframe restrictions in our keyword search, the first blockchain paper using text analysis appeared in 2015, six years after the birth of the Bitcoin blockchain (Nakamoto, 2008). The total number of papers published annually has been increasing, indicating the growing interest in and recognition of text analysis as a methodology for blockchain-related research. Until 2019, conference proceedings were the main channels through which related papers were published; however, from 2020 onward, the number of papers published in journals began to increase. For several years, computer science papers have largely dominated the topic, which can be explained by the entry requirements for coding skills in many machine learning-based text analyses. Nevertheless, later years saw a growing number of papers from business-, economics-, and finance-related fields. Studies from other areas, such as social sciences and multidisciplinary studies, have also contributed to this topic. The number of papers in most of these areas remains limited. However, the growing diversification of research areas indicates that interest has begun to spread from computer science to these areas.

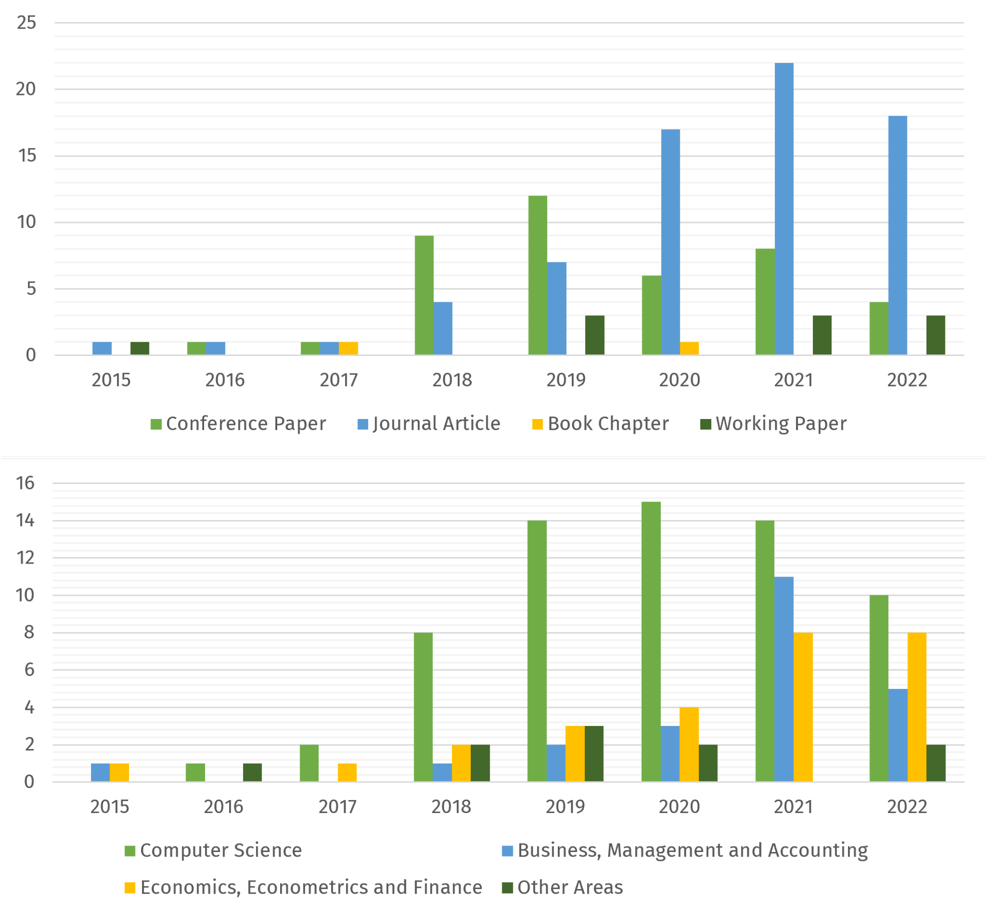


FIGURE 2.2: Types and research areas of the publications in each year.

2.3.2 Keyword Network

We analyzed the network of papers' keywords (see Figure 2.3) [7]. The size of the nodes reflects the frequency, the connection between the nodes indicates the co-occurrence of keywords in a paper, and the color of the nodes indicates the average year in which the keyword appears. The most common keywords are the three blockchain concepts: *Bitcoin*, *cryptocurrency*, and *blockchain*. Bitcoin had the earliest average occurrence and was associated with crime (e.g., crime, DarkNet market), social media (e.g., social networking, Twitter), and sentiment (e.g., opinion mining and sentiment analysis). Cryptocurrency is associated not only with crime but also with financial activities (e.g., financial services and investments), classification, and clustering (e.g., recurrent neural networks, deep learning, and topic modeling). The keyword blockchain tends to co-occur with specific applications (e.g., commerce and FinTech), topic modeling, and relationship analysis (e.g., network and trend analyses). Different keyword associations imply that the different scopes of topics within a blockchain are related to distinct economic activities and analyses. Individual text analysis-related keywords are mentioned less frequently; however, they appear in each blockchain scope. Sentiment analysis tends to go together with Bitcoin and cryptocurrency, whereas topic modeling and the corresponding keywords connect closely to cryptocurrency and blockchain.

2.3.3 Citation Ranking

Citation analysis helps identify the impact and common concerns of papers. However, one problem with using citations as an indicator of impact is that older papers have longer periods of citation accumulation. Thus, to offset this problem, we ranked the papers in terms of both total citations and citations per year (CPY) (Dumay and Cai, 2014) and considered the top ten papers from both criteria. Table 2.1 lists these papers and summarizes their text data, sample period, text analysis techniques, and brief abstracts of the papers.

Nine papers appeared on both lists; one older paper (Georgoula et al., 2015) fell short of CPY and was surpassed by a newer paper (Kim, Park, and Lee, 2020). The topics of high-impact papers tended to concentrate on a narrow range. Ten studies

[7] We cleaned the keywords of the papers before conducting the network analysis to eliminate the effects of the plural form, abbreviation, and spelling variation, etc.

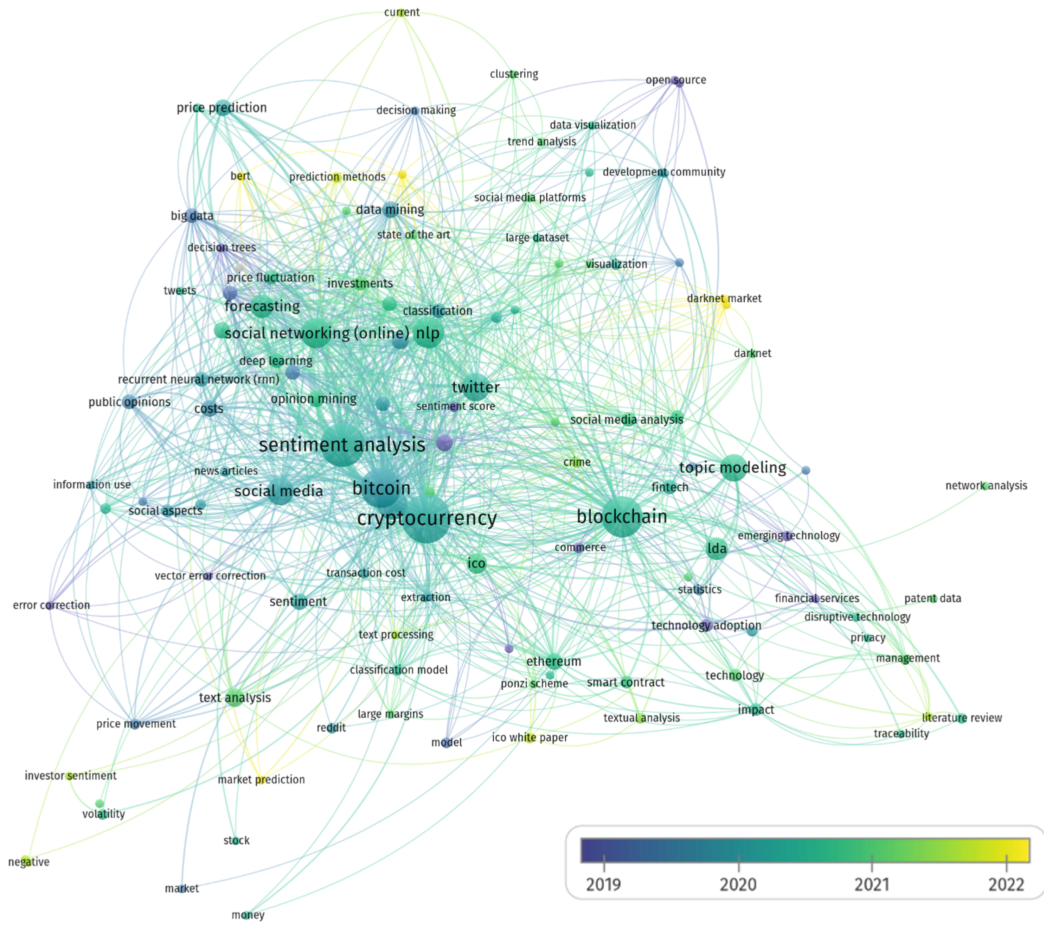


FIGURE 2.3: Keyword frequency and co-occurrence networks.

applied sentiment analysis and nine explored the predictive power of sentiment from social media platforms/news for cryptocurrency prices. Most studies focused on Bitcoin or a few altcoins with large market caps, while Kraaijeveld and Smedt (2020) included nine cryptocurrencies, and Li et al. (2019) studied a smaller cryptocurrency called ZClassic (ZCL). One study examined the sentiments of blockchain-related tweets and found that blockchain benefits were discussed more than its drawbacks (Grover et al., 2019). The study by Kim, Park, and Lee (2020) proposed a new topic modeling method and applied it to conduct a literature review on blockchain research to discover research trends. A detailed discussion is provided in Table 2.1.

TABLE 2.1: Top 10 most cited papers by total citation and citation per year.

Paper	Citation	CPY	Data and Period	Methodology	Summary
Polasik et al., 2015	431	54	Nexis database: English-language news mentioning Bitcoin 04.2011 - 03.2014	Lexicon-based sentiment analysis (Henry's finance-specific dictionary)	This study examines the determinants of Bitcoin price and the drivers of its success. Using sentiment in newspaper articles as one of the variables, it discovers that the negative mentions of Bitcoin lead to a price drop, while exhortatory pieces increase the Bitcoin price.
Kim et al., 2016	268	38	Comments and relevant replies in three cryptocurrency online communities: Bitcointalk, Forum ethereum, Xrpchat 12.2013 - 02.2016	Lexicon-based sentiment analysis (VADER)	This study uses the contents on three cryptocurrency communities to predict the price and number of transaction fluctuations. The sentiment of posts, the number of posts/replies, and the number of views of posts are used to perform Granger causality test on each currency for a time lag of 1 to 13 days. The results show that positive comments affect the price fluctuations of Bitcoin, whereas Ethereum and Ripple are influenced by negative comments.
Mai et al., 2018	208	42	Bitcointalk 01.2012 - 12.2014 Twitter: hastag Bitcoin 09.2014 - 12.2014	Lexicon-based sentiment analysis (LM lexicon)	This study investigates the impacts of social media on Bitcoin price. It separates the users into two groups, 1) the silent majority of users and 2) the vocal minority, and examines the impacts of these two groups, respectively. It finds that Bitcointalk has a more substantial impact than Twitter, and the silent minority exerts a more significant effect on future Bitcoin prices.
Georgoula et al., 2015	170	21	Twitter: keywords and hashtags Bitcoin, BTC, and Bitcoins 10.2014 - 01.2015	Machine leaning-based sentiment analysis (Support Vector Machines)	This study sheds light on the factors determining the price of Bitcoin in the short- and long-run. It adds Twitter sentiment into conventional prediction model. Specifically, it constructs a Twitter sentiment measure using SVMs and finds that sentiments have a positive short-run impact on Bitcoin prices.
Abraham et al., 2018	164	33	Twitter: hashtags Bitcoin and Ethereum 03.2018 - 05.2018	Lexicon-based sentiment analysis (VADER)	This study uses a linear model for predicting price changes of Bitcoin and Ethereum utilizing Twitter sentiment, tweet volume and Google Trends data. The results indicate that Twitter sentiment tends to be positive regardless of price direction and is, therefore, not a feasible predictor of price changes.
Kraaijeveld and Smedt, 2020	132	44	Twitter: hashtags including following nine cryptocurrencies: Bitcoin, Ethereum, XRP, Bitcoin Cash, EOS, Litecoin, Cardano, Stellar and TRON 06.2018 - 08.2018	Lexicon-based sentiment analysis (VADER, LM lexicon, and manually complied cryptocurrency-related words)	This study tests to what extent Twitter sentiment can be used to predict price returns for nine cryptocurrencies. It measures sentiments using a self-constructed lexicon and performs bilateral Granger-causality testing to find the causality. It finds the predictive power of Twitter sentiment for several cryptocurrencies.

Grover et al., 2019	122	31	Twitter: hashtag Blockchain 01.2018 - 02.2018	Lexicon-based sentiment analysis (Bing)	This study explores blockchain acceptance by examining the tweet information. It combines manual content analysis and lexicon-based sentiment analysis to distinguish the topics discussed and the user opinion. The analysis shows that users are attracted by security, privacy, transparency, trust and traceability. Furthermore, blockchain benefits are more frequently discussed than its drawback.
Karalevicius, Degrande, and Weerdt, 2018	120	24	Expert media news from CoinDesk, Cointelegraph, NewsBTC 05.2013 - 02.2016	Lexicon-based sentiment analysis (Harvard-IV General Purpose Psychological Dictionary and LM lexicon)	This study utilizes Bitcoin-related news articles to predict semi-short-term Bitcoin price movement. Integrating the sentiments of such news shows that the market initially overreacted to the news articles, resulting in multiple corrections.
Li et al., 2019	93	23	Twitter: keywords and hashtags ZClassic, ZCL, and BTCP 01.2019 - 02.2019	Lexicon-based sentiment analysis (Textblob)	This study analyzes Twitter signals as a medium for user sentiment to predict the hourly price fluctuations of ZClassic. It compiles the tweets into an hourly sentiment index, creating a weighted index giving larger weight to retweets. These two indices and the raw sentiment are used as input for Extreme Gradient Boosting Regression Tree Model for prediction.
Valencia, Gómez-Espinosa, and Valdés-Aguirre, 2019	90	23	Twitter: keywords and hashtags including following four cryptocurrencies: Bitcoin, Ethereum, XRP, Litecoin 02.2018 - 04.2018	Lexicon-based sentiment analysis (VADER)	This study uses sentiments on Twitter as input features for multiple machine learning algorithms to predict the price movement of four cryptocurrencies. It shows that Twitter data alone can be used to predict certain cryptocurrencies.
Kim, Park, and Lee, 2020	79	26	Academic papers: keyword or abstract contain "Blockchain", "Block chain", and "Block-chain" in six databases: Scopus, ScienceDirect, Web of Science, IEEE Xplore, Google Scholar, and Korean Citation Index. 01.2014 - 08.2018	Topic modeling (W2V-LSA)	This study proposes an improved method for topic modeling (W2V-LSA) and performs an annual trend analysis of blockchain-related literature. The experimental results confirmed the usefulness of W2V-LSA in terms of the accuracy and diversity of topics by quantitative and qualitative evaluation, and it can be an option for researchers using topic modeling for technology trend analysis.

2.4 Discussion of Research Questions

2.4.1 RQ1: Which research scope, text data, and methodology are used to conduct text analysis in the blockchain area?

In this section, we briefly introduce the scope, text data, and methodologies used in the papers and bridge the elements to identify the most used combinations. Figure 2.4 displays the connections among research scopes, text data, and methodologies in proportion to the number of papers [8].

Research Scope

'Specific cryptocurrency' (72 papers, 58%) is the most frequently used scope and Bitcoin in particular is the most studied cryptocurrency. To better recognize the importance of Bitcoin, we separate studies that focus exclusively on Bitcoin (40 papers, 32%) from the others. Other studies examine cryptocurrencies with large market caps, special small cryptocurrencies (Li et al., 2019; Mnif, Lacombe, and Jarboui, 2021; Vacca et al., 2021), or a large number of cryptocurrencies to represent the market (Steinert and Herff, 2018; Schwenkler and Zheng, 2021).

Another substantial scope is the general concept of blockchain (26 studies, 21%). These studies treat blockchain technology and its applications as a whole and discover its uses in particular fields (e.g., supply chain management (Medhi, 2020; Hirata, Lambrou, and Watanabe, 2021; Xu and He, 2022), banking (Daluwathumullagamage and Sims, 2020), and accounting (Garanina, Ranta, and Dumay, 2021)) and how blockchain-related topics evolve (over time) (Zhang, Daim, and Zhang, 2021; Chousein et al., 2020; Medhi, 2020; Silva and Moro, 2021; Zeng et al., 2018; Shahid and Jungpil, 2020; Perdana et al., 2021).

The literature also covers the scope of the cryptocurrency market as a whole (11 papers, 8.9%) (Caliskan, 2020; Siu, Collier, and Hutchings, 2021), ICO projects (13 papers, 10.5%) (Toma and Cerchiello, 2020; Liu, Sheng, and Wang, 2021; Sapkota and Grobys, 2021), and smart contract (two papers, 1.6%) (Ibba, Ortu, and Tonelli, 2021; Zhang, Daim, and Zhang, 2021).

[8] Some of the papers use various types of text data and methodologies; therefore, the sums of text data and methodology exceed the number of papers.

It is worth noting that, in our search, the keywords also included stablecoin, NFT, and STO, but we found no papers that used text analysis to examine these scopes. This may have resulted from the late development of these blockchain use cases. However, increasing growth in such applications has been observed in recent years (Lambert, Liebau, and Roosenboom, 2021; Wang et al., 2021b), thus creating opportunities and the needs to address relevant research questions using text analysis.

Text Data

Table 2.2 summarizes the text data and corresponding data sources we identify from the papers, which helps researchers navigate to the sources of their target data. We categorize texts into four groups: 1) corporate-produced documents, 2) user-generated content, 3) news, and 4) academic papers.

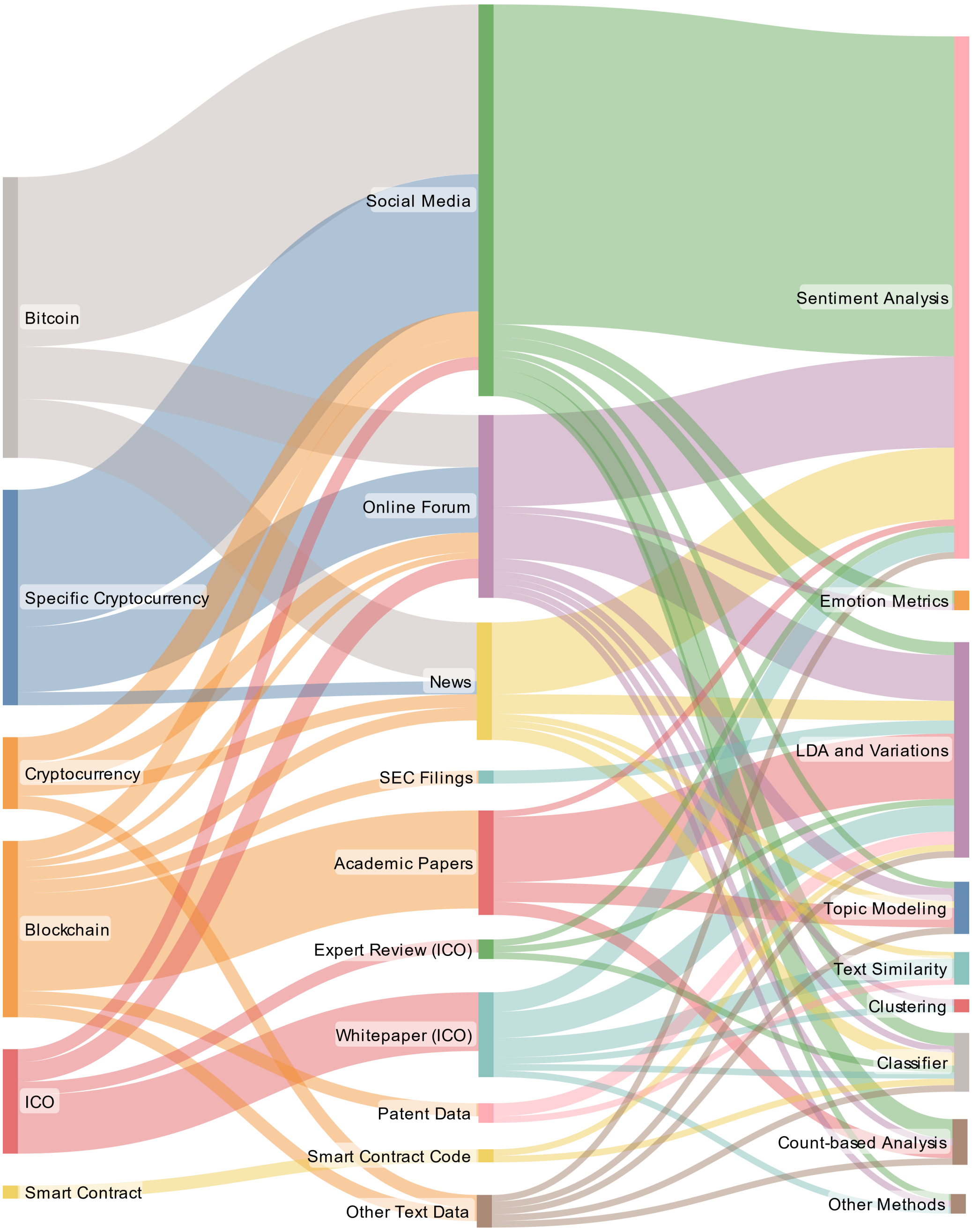


FIGURE 2.4: Connections among research scope, text data, and the methodology.

TABLE 2.2: Detailed information of text data sources used in the literature.

	Text Information	Data Source	Data Type *	Example Paper
Unique text data in blockchain	Smart contract code	Solidity	1	Ibba, Ortu, and Tonelli, 2021; Zhang, Daim, and Zhang, 2021
	Whitepaper	https://ICOHolder.com https://ICOMarks.com https://ICORatings.com https://ICODrops.com https://FoundICO.com https://CryptoCompare.com	1	Sapkota and Grobys, 2021; Thewissen et al., 2022
	Whitepaper and Expert report	https://ICOBench.com	1, 2	Xu et al., 2021
	Cryptocurrency community: technical and economic topics related to Bitcoin or other altcoins	https://bitcointalk.org https://www.xrpchat.com https://forum.ethereum.org	2	Kim et al., 2016; Gurdgiev and O'Loughlin, 2020
	Bitcoin abuse report	https://Bitcoinabuse.org	2	Choi et al., 2022
	Cryptocurrency news	https://www.coindesk.com https://www.newsbtc.com https://www.fxstreet.com/cryptocurrencies/news https://cointelegraph.com https://www.cryptocompare.com https://cryptocoin.news	3	Karalevicius, Degrande, and Weerdt, 2018; Farimani et al., 2022
General data sources contain blockchain topics	Firm disclosures including 10-Ks, conference calls, etc.	SEC Filings Full list: https://www.sec.gov/forms	1	Yen and Wang, 2021; Stratopoulos, Wang, and Ye, 2022
	Patent filing	Patent database: USPTO, EPO	1	Wang et al., 2021a; Zhang, Daim, and Zhang, 2021
	The required skills for job applicant	Online recruitment website	1	Ge et al., 2021
	Terms of Services Agreement	Company website	1	Caliskan, 2020
	Social media platform: contain blockchain-related messages	Twitter, Sina Weibo, Stocktwits	2	Chen et al., 2019a; Pan, Feng, and Jiayin, 2020; Huang et al., 2021

	Online forums or groups: contain blockchain-related posts	Reddit, Github, Telegram, Stack-Exchange, Discord	2	Alahi et al., 2019; Hinds-Charles et al., 2019; Nizzoli et al., 2020
	Online forum: contains criminal/illicit topics	HackForums	2	Siu, Collier, and Hutchings, 2021
	Users' review about a specific product/service	App store	2	Voskobojnikov et al., 2021
	Web data (news, social networking websites, forums, etc,.)	Web data monitoring: Webz.io, Notified, OpView Social Listening Platform	2, 3	Lu et al., 2017; Inamdar et al., 2019; Grassman et al., 2021
General data sources contain blockchain topics	News articles	Newspaper channels: The Financial Times, The Economist, The Economic Times, Business Insider, The Wall Street Journal	3	Azqueta-Gavaldón, 2020
	News articles from multiple channels	News Terminals: Nexis, Refinitiv Eikon, NewsAPI, RavenPack	3	Polasik et al., 2015; Rognone, Hyde, and Zhang, 2020; Anamika and Subramaniam, 2022
	Academic Papers/industry articles	WoS, Scopus, Google Scholar, Science Direct, IEEE Xplore Digital Library, ACM Digital Library, JSTOR, SSRN, Business Source Premier	4	Shahid and Jungpil, 2020; Silva and Moro, 2021

* 1) corporate-produced documents; 2) user-generated content; 3) news; and 4) academic papers.

Corporate-produced document Corporate-produced documents utilize formal and technical languages to provide detailed information about the company or specific products and services. Despite the precise information provided by these documents, we found only 18 studies that used such texts. ICO whitepaper, which pitches the project idea and outlines the business plan, is a voluntary disclosure by the ICO project team to attract potential investors (Florysiak and Schandlbauer, 2022; Thewissen et al., 2022). Another example of such document is smart contract code. Although the code does not strictly belong to human language, its fixed format enables researchers to obtain information regarding the subject of the contract (Ibba, Ortu, and Tonelli, 2021; Zhang, Daim, and Zhang, 2021). Blockchain-related texts can also be extracted from corporate documents, such as SEC and patent filings, through keyword searches and used to examine blockchain adoption (Yen and Wang, 2021; Wang et al., 2021a; Zhang, Daim, and Zhang, 2021; Stratopoulos, Wang, and Ye, 2022).

User-generated content Among all text data, user-generated content was the most frequently used (85 times, 64%). This type of text features a shorter length and informal language, and generally expresses the opinions of users on a particular topic. Social media platforms offer rich resources for such texts (56 times, 42%). Specifically, most studies chose Twitter to extract text data for conducting the analyses (Patil, Akarsh, and Parkavi, 2018; Huynh, 2021; Mareddy and Gupta, 2022), while others used Sina Weibo (a Chinese microblogging website) or Stocktwits (a social media platform focused on financial topics) (Chen et al., 2019a; Pan, Feng, and Jiayin, 2020; Huang et al., 2021).

Compared with social media platforms, online forums often have a specific focus and attract users with shared interests; therefore, they tend to offer deeper discussions. Cryptocurrency-specific forums, such as bitcointalk, XRPChat, and Ethereum Community Forum (Kim et al., 2016; Gurdgiev and O'Loughlin, 2020), have sections with distinctive topics. User discussions on topic-focused forums, such as GitHub, Reddit, and StackExchange have provided insights into the development of blockchain (Hinds-Charles et al., 2019; Bahamazava and Reznik, 2022; Ortu et al., 2022). There are numerous cummnities (i.e., subreddits) within the crypocurreny framework of Reddit (e.g., r/CryptoMarkets, r/Bitcoin), and users can join the communities to share up-to-date news or express their opinions on topics. In contrast, HackForums contains posts on illicit activities (Siu, Collier, and Hutchings, 2021).

News News articles are one of the most widespread and accessible textual data. They provide up-to-date factual information on events and commentaries/opinions on a topic. Analyzing blockchain news on a scale allows researchers to identify the evolution and public sentiment toward the technology. For instance, multiple news channels report the upcoming Ethereum Shanghai Hard Fork, but they contain different sentiments toward the event: FXStreet (2023) neutrally introduces the updates it would bring; U.Today (2023) illustrates multiple reasons for developers to be concerned about the hard fork, while Bloomberg (2023) is comparatively optimistic about it by emphasizing that "Shanghai is expected to push more people and institutional investors to stake their coins to support the Ethereum network and earn yield."

Many studies use cryptocurrency-specific news channels (e.g., Coindesk and Cointelegraph) as their primary news data sources (Karalevicius, Degrande, and Weerdt, 2018; Farimani et al., 2022), whereas others search for blockchain-related news from financial newspapers (e.g., The Financial Times and The Economist) through keyword searches (Azqueta-Gavaldón, 2020).

Academic paper Literature reviews assist researchers in understanding the current status of research, identifying research gaps, and guiding future research (Chakkarwar and Tamane, 2019; Shahid and Jungpil, 2020; Garanina, Ranta, and Dumay, 2021). Unlike the standard literature, in which researchers spend time manually examining papers, the automated processing of text-analysis-assisted literature reviews enables researchers to acquire insights into a large number of papers in a specific area in a short time.

Methodology

Choosing a suitable methodology depends not only on the data characteristics but also on the research questions of the study. Our goal is not to provide a systematic classification of the methodologies, but to provide a big picture of the methodologies used in blockchain-related literature. Therefore, the methodologies presented in this section may overlap. For example, the underlying methodology of sentiment analysis can be a machine-learning-based classifier. This section outlines the principal methodologies most directly related to the research questions. In addition, we summarize the specific text analysis techniques used in the papers in Table 2.3 to provide supplementary details[9].

[9]The mathematical principles of the methodologies are beyond the scope of this review, but for each methodology, interested readers can refer to the cited studies for details.

TABLE 2.3: Detailed information of text analysis techniques used in the literature.

Analysis Type	Sub-category	Specific Technique	Reference	Example Papers
Feature extraction	Count-based	BoW	Zhang, Jin, and Zhou, 2010	Yen, Wang, and Chen, 2021
		N-Gram	Cavnar, Trenkle, et al., 1994	El-Masri and Hussain, 2021
		TF-IDF	Ramos, 2003	Pan, Feng, and Jiayin, 2020
		DDPWI	Proposed in the paper	Burnie and Yilmaz, 2019
	Word/Sentence embedding	Word2vec	Mikolov et al., 2013	Kilimci, 2020, Kim, Park, and Lee, 2020, Liu, Sheng, and Wang, 2021
		Doc2vec	Le and Mikolov, 2014	
		GloVe	Pennington, Socher, and Manning, 2014	
		FastText	Bojanowski et al., 2017	
		Affective Tweet	https://affectivetweets.cms.waikato.ac.nz	Balfagih and Keselj, 2019
		A-BiRNN	Proposed in the paper	Xu et al., 2021
Sentiment analysis	Lexicon/rule-based	VADER	Hutto and Gilbert, 2014	Kim et al., 2016; Abraham et al., 2018
		TextBlob	https://textblob.readthedocs.io	Jain et al., 2018; Li et al., 2019
		Sentistrength	http://sentistrength.wlv.ac.uk	Caviggioli et al., 2020
		SentiWordNet	Baccianella, Esuli, and Sebastiani, 2010	Cheuque Cerda and L. Reutter, 2019
		Alex Davies word list	Christie and Huang, 1995	Stratopoulos, Wang, and Ye, 2022
		Bing	https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html	Grover et al., 2019; Hassan, Hudaefi, and Caraka, 2021
		AFINN	Nielsen, 2011	Ayvaz and Shiha, 2018; Toma and Cerchiello, 2020
		LM lexicon	Loughran and McDonald, 2011	Mai et al., 2018; Dittmar and Wu, 2019
		Harvard-IV General Purpose Psychological Dictionary	Stone, Dunphy, and Smith, 1966	Karalevicius, Degrande, and Weerdt, 2018
		Quantitative Discourse Analysis Package	https://www.rdocumentation.org/packages/qdap/versions/2.4.3	Sapkota and Grobys, 2021

Sentiment analysis	Lexicon/rule-based	Henry's finance-specific dictionary	Henry, 2008	Mnif, Lacombe, and Jarboui, 2021; Anamika and Subramaniam, 2022
		Pattern library	https://github.com/clips/pattern	Galeshchuk, Vasylchyshyn, and Krysovatyy, 2018
		SentimentR	https://github.com/trinker/sentimentr	Rahman et al., 2018; Chiarello et al., 2021
		Ethical and unethical words dictionary	Constructed in the paper	Barth et al., 2020
		63 cryptocurrency words and abbreviations	Constructed in the paper	Kraaijeveld and Smedt, 2020
		Crypto-specific sentiment dictionary (in Chinese)	Constructed in the paper	Huang et al., 2021
		Crypto-specific lexicon (words, emojis, informal langugage)	Constructed in the paper	Chen et al., 2019a
	Machine learning-based	Long short-term memory (LSTM)	Hochreiter and Schmidhuber, 1997	Inamdar et al., 2019; Şaşmaz and Tek, 2021
		Recurrent neural network	Goldberg, 2017	
		Random forest	Ho, 1995	
		Naïve Bayes	Jurafsky and Martin, 2017	
		Support vector machine	Boser, Guyon, and Vapnik, 1992	
		Gradient boosting	Friedman, 2001	
		BERT	Devlin et al., 2018	Bashchenko, 2022; Ortu et al., 2022
		Bidirectional LSTM	Mousa and Schuller, 2017	Han, Ye, and Zhang, 2020
		Voting-included Algorithm	Constructed in the paper	Pant et al., 2018
		Sentiment Graph	Constructed in the paper	Yao, Xu, and Li, 2019
	Analytics Tool	Crimson Hexagon social sentiment	https://www.carahsoft.com/crimson-hexagon	Stanley, 2019
		Semantria	https://www.lexalytics.com	Caviggioli et al., 2020
		Meaningcloud	https://www.meaningcloud.com	

Sentiment analysis	Analytics Tool	StanfordCoreNLP	https://stanfordnlp.github.io/CoreNLP	Moustafa, Malli, and Hazimeh, 2022
		OPView	https://www.opview.com.tw	Lu et al., 2017
		RavenPack	https://www.ravenpack.com/products/edge/data/news-analytics	Rognone, Hyde, and Zhang, 2020
Emotion metrics		NRC-VAD Emotion Lexicon	https://saifmohammad.com/WebPages/nrc-vad.html	Toma and Cerchiello, 2020
		NRC Word-Emotion Association Lexicon	https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm	Chursook et al., 2022
		Text2Emotion	https://shivamsharma26.github.io/text2emotion	Aslam et al., 2022
Topic modeling	Topic modeling algorithm	LDA	Blei, Ng, and Jordan, 2003	Fu, Koh, and Griffin, 2019; Hirata, Lambrou, and Watanabe, 2021; Laturnus, 2023
		DTM	Blei and Lafferty, 2006	Linton et al., 2017; Lee, Zo, and Steinberger, 2022
		SentLDA	Bao and Datta, 2014	Thewissen et al., 2022
		Joint/sentiment topic model	Lin and He, 2009	Loginova et al., 2021
		Topic sentiment latent dirichlet allocation	Nguyen and Shirai, 2015	
		Nonnegative Matrix Factorization	Lee and Seung, 1999; Lee and Seung, 2000	Kang, Choo, and Kim, 2020
		Anchored Correlation Explanation	Gallagher et al., 2017	Nizzoli et al., 2020
		Word2vec-based Latent Semantic Analysis (W2V-LSA)	Proposed in the paper	Kim, Park, and Lee, 2020
	Analytics tool	Leximancer	https://www.leximancer.com	Daluwathumullagamage and Sims, 2020; Perdana et al., 2021
Text Similarity		Cosine Similarity	Kwon and Lee, 2003	Yen, Wang, and Chen, 2021
		Jaccard Similarity Coefficient	Jaccard, 1912	Sapkota and Grobys, 2021
		SBERT	Reimers and Gurevych, 2020	Bashchenko, 2022
Clustering		K-means clustering	MacQueen, 1967	Choi et al., 2022
		DBSCAN clustering	Ester et al., 1996	

Classifier	Machine learning algorithm	Catboost	Prokhorenkova et al., 2018	Chousein et al., 2020, Schwenkler and Zheng, 2021
		Random Forest	Ho, 1995	
		XGBoost	Chen and Guestrin, 2016	
		Neural network	Hashimoto et al., 2016	
		Naïve Bayes	Jurafsky and Martin, 2017	
Readability		Flesch-Kincaid Readability	Flesch, 1979	Narman, Uulu, and Liu, 2018, Sapkota and Grobys, 2021
		Dale-Chall Readability	Dale and Chall, 1948	
		Gunning Fog Index	Gunning, 1952	
		Automated Readability Index	Senter and Smith, 1967	
		Simple Measure of Gobbledygook	McLaughlin, 1969	
		Coleman-Liau Index	Coleman and Liau, 1975	
		Linsear Write	Klare, 1974	
		AWS blockchain template	https://docs.aws.amazon.com/blockchain-templates	Stanley, 2019
Network Analysis		Google knowledge graph	https://developers.google.com/knowledge-graph	Pan, Feng, and Jiayin, 2020

Text preprocessing Before conducting the actual analysis, multiple cleaning procedures should be applied to the raw text to prepare it as the input material. The necessary steps vary depending on the text condition and planned analysis. However, we identified standard preprocessing steps suitable for the majority of texts: removing special characters and punctuation, removing numbers and stopwords, lower-casing, spelling corrections, tokenization, assigning part-of-speech tags, and stemmization/lemmatization. Some raw texts require more cleaning than others. For example, texts from social media and online forums usually use informal language and emojis which can lead to misinterpretation. Papers therefore conducted additional procedures (Birim and Sönmez, 2022; Critien, Gatt, and Ellul, 2022): remove # and @user, remove URL links, convert emojis to words, and convert vocabulary abbreviations to words. These procedures remove redundant text, convert unrecognizable characters into valuable information, and are vital preparation steps.

Feature extraction The cleaned texts should be transferred to number representations to allow the computer to read and use for further analyses. It can also reduce computational complexity, enhance performance, and avoid the overfitting problem, making it an essential procedure in text analysis (Kou et al., 2020). This representation per se can also provides information and insight. Count-based methods are straightforward to understand and interpret. The Bag-of-words (BoW) is one of the most widely used approaches. It represents words according to their frequency in the corpus, disregarding order and context. N-grams are extensions of BoW that breaks the corpus into a contiguous sequence of n words. It can capture more context around each word, but produces a sparser feature set than BoW. BoW and N-grams assume that words that occur more frequently are more relevant and do not always hold true. Term frequency-inverse document frequency (TF-IDF) (Salton, Yang, and Yu, 1975) adds another metric of how rarely a word occurs across the entire corpus and assigns rarer words a higher score. Although such representations are generally used as inputs for further analysis, we identify papers that highlight frequent words and interpret them as blockchain topics (Zeng et al., 2018; Burnie and Yilmaz, 2019; El-Masri and Hussain, 2021). However, this method can be misleading, because count-based methods discard linguistic structures and may miss crucial text information.

Word-embedding mitigates this problem by representing words in vectors to capture their semantic and syntactic contexts in a document (Cong et al., 2021). In the vector space, the shorter the distance between two word vectors, the higher is the similarity of the words. *Word2vec* (Mikolov et al., 2013) is one of the most frequently used word embedding methods. It includes two configurations: skip-gram and continuous bag of words (CBOW). A skip-gram uses the current word to predict the surrounding words, whereas CBOW predicts the current word using its surrounding words. A generalization of *word2vec* and *doc2vec* (Le and Mikolov, 2014) adds a document feature vector to the word vector to capture the semantics of the paragraphs and documents. Word-embedding techniques are not frequently used in the literature, but we found that Kim, Park, and Lee, 2020 and Liu, Sheng, and Wang, 2021 integrated these techniques when processing their texts. Two other word-embedding models, *GloVe* and *fastText*, were used by Kilimci (2020).

Analysis Sentiment analysis is the dominant text-analysis approach in the literature (80 times, 53%). There are two major types of sentiment analysis: lexicon/rule-based and machine learning-based (Vohra and Teraiya, 2013).

Lexicon-based sentiment analysis calculates the sentiment score of a text based on the polarity of each word (i.e., positive, negative, or neutral) from sentiment dictionaries in which each vocabulary is assigned a sentiment score. Examples of well-established sentiment dictionaries include Valence Aware Dictionary for Sentiment Reasoning (VADER) (Hutto and Gilbert, 2014), which is particularly suitable for social media contexts, and Loughran and McDonald sentiment lexicon (LM lexicon) (Loughran and McDonald, 2011) in the finance domain. However, off-the-shelf dictionaries can sometimes generate inaccurate results because of different sentiments of the same vocabulary in different contexts (Loughran and McDonald, 2011). Therefore, some researchers have developed new and additional dictionaries (e.g., new vocabularies and emojis) in blockchain contexts for higher accuracy of sentiment quantification (Chen et al., 2019a; Barth et al., 2020; Kraaijeveld and Smedt, 2020).

Machine learning-based sentiment analysis adopts machine learning classifiers to study the sentiments of texts and classify them into instinctive sentiment groups. Researchers can build a model and train their data or apply a pre-trained model (e.g., Bidirectional Encoder Representations from Transformers (BERT)) to their analysis. Compared to lexicon/rule-based sentiment analysis, it is dynamic and can better fit the research context. We identified 12 papers that adopted this approach (e.g., Patil, Akarsh, and Parkavi (2018), Balfagih and Keselj (2019), Inamdar et al. (2019), and Aslam et al. (2022)). In particular, Han, Ye, and Zhang (2020) and Akba et al. (2021) propose and assess new models for sentiment analysis.

Sentiment analysis tools have also been utilized in academic studies (Lu et al., 2017; Stanley, 2019; Caviggioli et al., 2020; Moustafa, Malli, and Hazimeh, 2022). Such tools develop unique algorithms and reduce the programming requirements for researchers. However, most of these tools are commercially oriented, incur high subscription fees, and lack transparency regarding their algorithms. Hence, albeit the convenience, researchers should be cautious when using such tools.

In some studies, emotion-detection metrics have been applied in conjunction with sentiment analysis to achieve more precise emotion separation. For example, the NRC-VAD Emotion lexicon has three dimensions: valence, arousal, and dominance (Mohammad, 2018). This provides another layer for sentiment and can increase the quality of the analysis.

The Latent Dirichlet Allocation (LDA) and its variations were frequently chosen (33 times, 22%) for text analysis. LDA is a topic-modeling algorithm developed by Blei, Ng, and Jordan (2003). Topic modeling can identify the patterns of vocabulary and phrases in documents (within the corpus of interest), detect the differences in their topics, and cluster the documents according to the topics discussed in the documents. LDA is one of the most popular topic-modeling algorithms. It assumes that each document in the corpus consists of a number of latent topics and that each topic is characterized by a word distribution. Each topic is presented with a list of words and their fitting possibilities. Its variations include dynamic topic models (DTM), which add temporal features to the model (Blei and Lafferty, 2006) and SentLDA, which considers the boundaries between sentences and assumes that all words in a sentence are sampled from the same topic (Bao and Datta, 2014). The texts used in LDA models are typically unlabeled, and the researchers' task is to choose the optimal number of topics, which is primarily determined by the perplexity and coherence scores (Blei, Ng, and Jordan, 2003; Newman et al., 2010). After narrowing down the choices for the optimal number of topics, researchers become involved and integrate their interpretations to choose the optimal number of topics for the model. Together with other topic modeling and clustering algorithms, they belong to unsupervised machine learning. Evaluations of unsupervised machine learning vary from model to model, and human judgment is often required to evaluate the model quality. Nevertheless, these models are valuable for exploring the underlying features of a text without establishing an upfront framework (Grimmer and Stewart, 2013). This is especially applicable to research in blockchain, which is still understudied and has few established classifications.

In contrast, supervised machine-learning classifiers are applied to pre-labeled texts, and the texts are classified into pre-specified groups. The idea is to first manually categorize a set of documents and then train a supervised model that automatically learns how to assign categories to documents using a training set (Bao and Datta, 2014). Owing to the training process, they are domain-specific and better fit the research context (Grimmer and Stewart, 2013). Multiple models are often applied to the same dataset and researchers can easily compare the performance of classifiers using certain metrics (e.g., precision, recall, accuracy, F1-score) to select the best-fitting model. Nevertheless, in blockchain-related research, they are utilized much less for text data (nine times, 6%).

Bridging the Elements

Figure 2.4 shows that the combinations of the elements are diversified depending on the purpose of the studies. Nevertheless, we observe two primarily adopted paths for text analysis in blockchain research: a) papers studying specific cryptocurrencies tend to apply sentiment analysis to instant user-generated content or news articles to discover the correlations between public opinions/emotions and cryptocurrency market behavior, and b) papers studying the broad concept of blockchain primarily choose official documents from companies (e.g., SEC and patent filings) and apply topic models to explore the classifications or trends in the sector.

The links among the above elements are not permanent; that is, researchers can choose combinations according to their requirements. To select effective combinations, researchers must understand the characteristics of the data, presumptions to use a particular methodology, and the questions they intend to investigate. The design should facilitate the generation of interpretable and meaningful results to answer the research questions.

2.4.2 RQ2: What topics are addressed using text analysis in current literature?

The data and methodologies are used to serve the purpose of the study and should be chosen depending on the research questions (Grimmer and Stewart, 2013). In the following section, we summarize blockchain-related topics discussed in the existing literature that involve text analyses.

Relationship Discovery

Researchers have used different text data (often combined with other variables) to identify correlations. The speculative nature and high volatility of cryptocurrencies have led to studies exploring the relationship between market fluctuations and information on online platforms. Different factors of online discussions, including the counts of specific keywords, discussions of different topics, and sentiment classes, are extracted. These factors are used as variables to test whether they are associated with cryptocurrency market activities, such as price changes and the co-movement of peer cryptocurrencies (Polasik et al., 2015; Phillips and Gorse, 2018; Barth et al., 2020; Schwenkler and Zheng, 2021). From more specific perspectives, studies distinguish different user groups and vocabularies and find that content from certain groups or the presence of certain words is more closely related to changes in the cryptocurrency market (Burnie and Yilmaz, 2019; Kang, Choo, and Kim, 2020). Xie, 2021 explores the relationships among online discussions and demonstrates that online communities' conflicting opinions and redundant discussions result in low trading volumes.

An ICO whitepaper, perceived as a prospectus for an initial public offering (IPO) in a less regulated way, provides information that can impact investors' decisions and, to some extent, determine the success of projects. Many dimensions of such texts influence the performance of ICO. For instance, ICO projects with higher technological sophistication shown in whitepapers are more likely to be successful and less likely to be delisted (Liu, Sheng, and Wang, 2021). Those whitepapers that are unique–that is, have more project-specific information and avoid borrowing common phrases from previous whitepapers–can lead to higher fundraising amounts and better post-ICO performance (Yen and Wang, 2021; Florysiak and Schandlbauer, 2022). The readability and sentiment expressed in whitepapers can also affect investors' decisions to invest in the described project (Stanley, 2019; Sapkota and Grobys, 2021).

For public companies that meet higher disclosure standards, blockchain-related information can be extracted from 10-K filings and used to investigate whether blockchain adoption brings value and efficiency to companies (Yen, Wang, and Chen, 2021).

Cryptocurrency Performance Prediction

Forecasting has always been an important topic in cryptocurrency studies. In addition to econometric methods and statistical models for price prediction, sentiment has also been used as a predictor of market movement (Mao, Counts, and Bollen, 2011; Fang et al., 2022). The effect of sentiment on the cryptocurrency market could be magnified by the lack of traditional financial fundamentals in valuation, and vocal and active investors on social media (Corbet et al., 2018; Gurdgiev and O'Loughlin, 2020). Machine learning models, especially supervised models, are often applied to use sentiment data for prediction. Sentiment is used as the sole input to a model or as a supplement to conventional variables (e.g., price, trading volume, blockchain metadata (Sebastião and Godinho, 2021)).

Texts from social media are extracted, and each document is assigned a sentiment score using a sentiment analysis technique (see Table 2.3 for details). The scores (along with other variables) are subsequently used as inputs for the prediction models. They have predictive power for the direction of price movement (Loginova et al., 2021; Critien, Gatt, and Ellul, 2022) and the short-term (e.g., hourly and daily) magnitude of price changes (Li et al., 2019; Farimani et al., 2022; Ortu et al., 2022).

The impact of social media content depends particularly on the level of information dissemination. Thus, celebrity or opinion leader posts (i.e., influencers) or discussions about them could have more power than other posts (Kang, Choo, and Kim, 2020). Huynh (2021; 2022) quantifies the tweet sentiments of Donald Trump and Elon Musk using LM lexicon and finds that negativity in Trump's tweets leads to higher returns on Bitcoin, whereas both pessimistic and optimistic expressions from Musk have a positive effect on Bitcoin returns. Cary (2021) analyzes the tweet sentiment about Elon Musk's performance on Saturday Night Live on 8 May 2021 and found that the negative opinion toward his performance led to the price decline of Dogecoin.

Prediction models have also been used in ICO studies. Text data variables (e.g., expert reviews and social media sentiment) and non-text variables (e.g., sale price, project duration, and expert ratings) are utilized simultaneously to predict the success of ICO projects (Xu et al., 2021; Chursook et al., 2022).

Overall, studies focusing on predicting market movements and project success constitute a large proportion of the papers in this review. However, the data and methodologies mainly follow a similar direction: applying sentiment analysis to Twitter posts and associating the respective sentiment metrics with high market capitalization cryptocurrencies.

Classification and Trend

One step in understanding large-scale texts containing multiple documents is to categorize the documents and create classifications. Using clustering/topic models or classifiers, content features (i.e., the topics discussed) in documents can be extracted and used to group documents into different classifications. By adding a temporal dimension to the static classification, the classification information can provide the trends of a particular group of topics.

Such models can be valuable when applied to academic papers in literature reviews to facilitate an understanding of existing studies and identify further research. Unlike standard literature reviews, in which researchers read through papers to derive results, topic modeling-based literature reviews extract the titles and abstracts of papers and rely on algorithms to extract topics from the texts. Classification algorithms are used to understand the current state and development of blockchain research (Chakkarwar and Tamane, 2019; Shahid and Jungpil, 2020; Lee, Zo, and Steinberger, 2022). Some studies have dived into blockchain applications within a sector (e.g., consumer trust, banking, and accounting) to facilitate researchers and practitioners in identifying future research areas and business opportunities (Silva and Moro, 2021; Daluwathumullagamage and Sims, 2021; Garanina, Ranta, and Dumay, 2021). Although it enables researchers to examine text content on a large scale without time-consuming manual reading, one of the drawbacks of using text analysis for literature reviews is the lack of an information screening process, during which irrelevant papers are excluded from the review.

Most papers included in this review (Xu and He (2022) is an exception) directly use all papers from the keyword search results as their input for topic models and further analyses. In this case, many irrelevant papers may be erroneously included in the models and the noise information they contain can be significant, leading to biased or inaccurate conclusions. To avoid undermining the advantages of topic modeling, researchers must carefully design the selection criteria for their dataset when performing such studies.

At a more technical level, the classification and trends of blockchain infrastructure and application design problems have also been addressed. Using texts from technique-oriented platforms (e.g., GitHub and StackExchange), some studies have observed a shift in developers' interests from mining to software development (Alahi et al., 2019; Hinds-Charles et al., 2019). A special case involves the use of a smart contract code as an input for topic models or classifiers. Researchers can then discover the most common uses of smart contracts and identify Ponzi schemes by analyzing the code (Ibba, Ortu, and Tonelli, 2021; Zhang et al., 2021). Despite the focus on technical information, such studies have implications not only for developers and computer scientists but also benefit researchers in finance and economics by, for instance, identifying investor interests and customer demands.

The evolution of the blockchain topic is often tied to unique events that affect market activity and trigger changes in investor behavior. Linton et al., 2017, for example, study how blockchain topics change during periods of significant events in the cryptocurrency world, such as the insolvency of the MtGox Bitcoin exchange in 2014 (Goldstein and Tabuchi, 2014) and the hack into Bitfinex in 2016 (Baldwin, 2016) (e.g., from sole 'Bitcoin trading' topics to 'security issues' or 'scams' as predominant topics in online forums). Other researchers (Daluwathumullagamage and Sims, 2020; Pan, Feng, and Jiayin, 2020; Bahamazava and Nanda, 2022) incorporate the influence of specific events (e.g., Bitcoin halving events, the introduction of regulations, and COVID-19) into their models to better interpret the change in interest during different periods.

Crime and Regulation

Illegal activities and crimes have always surrounded discussions on cryptocurrency. Many early users appraised the (pseudo)anonymity of cryptocurrency and used it as currency for illicit purchases on DarkNet. In the early stages, cryptocurrencies were suggested that cryptocurrencies contribute to improving black markets (Foley, Karlsen, and Putniņš, 2019).

Bahamazava and Reznik, 2022 and Bahamazava and Nanda, 2022 explore the posts from Reddit (subreddit DarkNet) to study the criminal topic evolution and the mainstream methods to trade cryptocurrencies illegally. Crime-related texts on other channels such as Twitter, Telegram, and HackForums are also used to identify the specific illegal activities discussed (Barth et al., 2020; Nizzoli et al., 2020; Siu, Collier, and Hutchings, 2021). One rich first-hand source for examining fraud from the victim's side is the reports from https://www.bitcoinabuse.com, where the victims of Bitcoin fraud share their experiences and post the original messages they received from the abusers. Choi et al., 2022 cluster these messages and find high similarity of a large number of messages, suggesting the existence of only slight modification of fraud messages and certain patterns of the language usages from Bitcoin fraud instigators. Zhang et al., 2021 apply an improved CatBoost classifier to smart contract codes to find the common characteristics of Ponzi schemes hidden in the lines.

Although studies inspecting illegal activities have accumulated, the number of studies exploring relevant regulations remains minimal. We identified only two studies that explicitly discussed regulatory issues. In the study by Bahamazava and Nanda, 2022, after discovering the preferred methods of buying cryptocurrencies for money laundering, they cross-examined anti-money laundering regulations in Italy and Russia to see if they have corresponding paragraphs to address such purchasing methods. Chousein et al., 2020 investigate how service providers of public blockchain systems communicate with their users about the influences of the EU General Data Protection Regulation (GDPR) on their services and find a shortage of communication and transparency on GDPR compliance issues.

There are two reasons for the lack of regulation-oriented text analysis studies. First, the time lag between the introduction of regulations in different jurisdictions limits the availability of data for regulatory studies. Second, analyzing the content of regulations requires a computer program to understand the legal terms. Therefore, context-specific dictionaries are required to correctly extract information. Researchers should also have domain knowledge to interpret the results accurately, which can be challenging in many areas. Nevertheless, because understanding regulatory frameworks is essential to advance our understanding, combat blockchain crimes, and promote blockchain adoption, more research is needed from the perspective of blockchain-related regulations.

Perception of Blockchain

The perception of (potential) users is crucial for the development of emerging technologies such as blockchain. Public acceptance does not merely rely on economic benefits, but also on other aspects. Studies have attempted to discover how the public perceives blockchain technology and the drivers of attitude construction. Such studies are closely associated with social and cultural factors and are, therefore, located in interdisciplinary studies, such as behavioral finance. The number of papers was not significant (seven papers) in this review; however, the questions discussed were diverse.

Blockchain was initially surrounded by suspicion and considered a questionable technology; however, its acceptance grew gradually. Users are attracted to the security, privacy, transparency, trust, and traceability offered by blockchain (Grover et al., 2019), but their adoption is still hindered by a lack of blockchain knowledge and distrust of blockchain (Yadav et al., 2021). Doubts can be removed by building channels for the public to gain knowledge about it:1) articles from the media help the public obtain more information about blockchain, which boosts further exploration of the technology and acceptance; 2) existing business problems motivate experimenting with blockchain and enhance trust (Perdana et al., 2021). Cultural background also helps shape the perceived value of blockchain. Grassman et al., 2021 conduct a comparative study between Sweden and Japan on the attitude towards autonomy that cryptocurrency brings. The principle of autonomy has a higher intrinsic value in Sweden, whereas Japan adopts a more pragmatic view of autonomy (i.e., facilitating investment prospects).

In broad-term blockchain, specific products with distinctive characteristics are viewed differently. Some studies (Caliskan, 2020; Mnif, Lacombe, and Jarboui, 2021;

Bashchenko, 2022) explore the perceptions of Bitcoin, Bitcoin Green, and cryptocurrency exchanges and explained the reasons for their interpretations.

2.4.3 RQ3: What are the research gaps and promising future research topics?

We now summarize the research gaps described in the papers and observed by us and develop future research topics to which future studies could address.

Improvement of Data Preparation

The quality of the input data largely determines the model output results; however, the complexity of text data makes it challenging to prepare. Many current studies merely conduct standard data preparation and omit the features of different types of text. To prevent "garbage-in-garbage-out", future research can look more deeply into the characteristics of specific texts and prepare the data in a way that fits the characteristics of the texts.

Data selection After text preprocessing, the text data should be further selected or weighted by considering the text features. This procedure is yet neglected by a substantial number of papers. For example, Twitter offers millions of short texts daily, but misinformation is omnipresent. Bots and fake accounts should not be ignored and should be separated from others (Burnie and Yilmaz, 2019; Kraaijeveld and Smedt, 2020). Bashchenko, 2022 divides news into two types: a) endogenous news, which describes the past price movement; b) fundamental news, which provides information that can have higher impacts. When using news for price prediction, endogenous news should be filtered out because it has a limited influence on future prices.

Another way to improve preparation can be achieved by setting relevance levels for the texts. Twitter accounts can be weighted according to their influence levels (e.g., number of followers, retweets, and user networks) (Jain et al., 2018; Li et al., 2019), and the influence of a patent is reflected by the number of citations.

Dictionary building Dictionaries are essential in text analysis models (e.g., sentiments and topics). However, they are generally only applicable to a specific context since vocabularies can change their meanings depending on discipline (Loughran and

McDonald, 2011). The impact of using an off-the-shelf dictionary in other areas can be a substance for blockchain studies, as new vocabularies and jargons have been invented in blockchain. Studies have indicated that designing a domain-specific lexicon for blockchain could potentially improve the accuracy of analysis (Balfagih and Keselj, 2019; Chen et al., 2019a; Sattarov et al., 2020). existing studies primarily adopt the VADER (Hutto and Gilbert, 2014) and LM lexicons (Loughran and McDonald, 2011), and only a few studies have developed or integrated blockchain-specific lexicons (Chen et al., 2019a; Barth et al., 2020; Kraaijeveld and Smedt, 2020; Huang et al., 2021).

Extension to Underused Data and Growing Areas

In this review, we find a concentration of text data uses from social media, online forums, and academic papers. Simultaneously, many other documents containing valuable information are underused. Corporate-generated documents (e.g., SEC and patent filings) are not frequently utilized despite their importance in revealing corporate-level information. For instance, in finance studies, patent filings are used to identify specific FinTech categories (Chen et al., 2019b; Chen et al., 2022a). Studies use 10-Ks for different purposes: product description sections for the new industry set according to product similarity (Hoberg and Phillips, 2016), business descriptions for company's asset specificity(Chen et al., 2022a), and risk disclosures for risk detection (Bao and Datta, 2014; Hanley and Hoberg, 2019). Corporate disclosures are versatile, and cater to multiple research purposes. One limitation of corporate disclosures is that blockchain startups have limited mandatory disclosures. Nevertheless, future research can make greater use of such documents to gain insights into blockchain adoption strategies of established companies.

Another gap in the review is the absence of papers related to the keywords NFT, STO, IEO, and stablecoin. These are relatively new concepts in blockchain and are largely understudied. Researchers investigating these areas will contribute to a better understanding of market mechanisms. For example, potential text data in NFTs include descriptions and social media discussions of NFT items. STOs are treated as traditional securities and adhere to all rights and obligations including approved prospectuses for public offerings. IEO project whitepapers were thoroughly vetted by exchange prior to launch. Therefore, the above documents are more standardized and can be used similarly as standard corporate disclosures. Stablecoin is connected to conventional financial systems and have drawn attention to financial stability issues. News (integrated with event studies) could provide coverage from this perspective.

Regulation

Given the increasing trend of cryptocurrency in the monetary system, government policies and regulations are essential for counteracting risks, restricting illicit activities, and protecting consumers (Chokor and Alfieri, 2021).

Many jurisdictions have updated or supplemented their regulatory frameworks to accommodate the existence of cryptocurrencies and other blockchain-based decentralized applications (e.g., Market in Crypto-Assets (MiCA) and Framework for International Engagement on Digital Assets). Issues such as money laundering, terrorist financing, and tax evasion have been extensively recognized and addressed. In addition, organizations such as the International Organization for Standardization (ISO) and the Financial Stability Board (FSB) are working to establish international rules and standards to promote collaboration among jurisdictions. Many proposed frameworks are still in their initial stages or awaiting implementation, and updates can be expected.

Texts used in regulation-related research are not limited to regulatory documents, but also include other texts, such as corporate disclosures related to blockchain or cryptocurrency (SEC, 2022), terms of service agreements, and online discussions about regulatory terms. Future research could integrate regulatory factors into the study, examine the impact of regulations on markets in different jurisdictions (Barth et al., 2020), and observe users' perceptions of and reactions to specific regulations. This could provide insightful implications for practitioners and policymakers regarding the implementation of relevant regulations and how takers of specific regulations will adopt them.

2.5 Conclusion

The uncomplicated access and rich information in blockchain-related texts make them ideal for complementing numerical data in research. However, a comprehensive review of this topic to provide guidance for researchers is lacking.

This study addresses this issue by making several contributions to the literature. First, we provide comprehensive summaries of research scope, text data sources, and text analysis methodologies in the existing literature to guide researchers in finding pertinent resources. Second, we go beyond individual elements and exhibit the connections between them. We conflate the above elements and display the two most frequently used combinations:1) papers focusing on cryptocurrencies conduct sentiment analysis on posts from instant user-generated content or news articles to find the correlations between sentiment and market behavior, and 2) papers examining the concept of blockchain use formal documents to apply topic modeling to discover classifications and trends. We emphasize that it is crucial to choose appropriate combinations considering variable perspectives, such as data characteristics and research questions. Finally, we integrate blokcchain-related research areas and text analysis approaches into a joint framework. By not restricting our search to one discipline, we are able to capture the use of text analysis in non-technical blockchain studies across disciplines and provide multiple perspectives on the topic. We highlight five major research topics discussed in the literature: relationship discovery, cryptocurrency performance prediction, classification and trend, crime and regulation, and the perception of blockchain. Furthermore, by referring to individual papers and aggregated information, we uncover three future research topics that researchers can explore: improvement of data preparation, studies with underused data and growing areas, and regulation-related research.

We are aware that this review shares publication bias of literature reviews. Studies with statistically significant results are more likely to be published, leading to a publication bias (Rosenthal, 1979). To alleviate the impact of bias, we searched the most comprehensive databases for peer-reviewed papers and chapters. We also included unpublished working papers on SSRN in keyword searches. Backward snowballing was conducted on the included papers to identify more papers that did not appear in the keyword searches. We believe that through our multiple procedures for identifying targeted papers, we obtained a comprehensive collection of papers for this literature review.

Despite this limitation, this study provides a timely academic-oriented review of the text analysis approaches used in blockchain research. Our detailed summaries will help researchers navigate specific text data types and methodologies. The findings of the current research landscape and suggested future directions could facilitate the selection of promising research topics and the implementation of suitable methodologies for their analyses. Overall, this review will be useful for researchers from various disciplines interested in exploring large-scale text data in blockchain-related research.

Chapter 3

The Landscape of Blockchain Innovation: A Textual Analysis of Patent Data

This chapter is an unpublished working paper presented at multiple international conferences. See Table 1.2 for details.

3.1 Introduction

Blockchain, as one of the distributed ledger technologies (DLTs), has been recognized as a key innovation with the potential to disrupt and improve many industries, including finance, supply chains and international trade, public services, and others (Zhao et al., 2016; Friedlmaier et al., 2018). Blockchains vary widely in their economic design, technical implementation, and business purpose to meet practical needs. Because of this variety of approaches, there is seldom a guiding framework to navigate the landscape of blockchain-related innovation. Such a framework could inform business research and innovation agendas by highlighting topics of practical relevance or the lack thereof. This paper draws such a nuanced landscape by empirically analyzing blockchain-related patent applications.

The first set of blockchain applications was the creation of cryptocurrencies. However, it has been expanding into other areas, leading to a surge in the quantity and variety of blockchain applications and academic research across disciplines, including computer science, business finance. Many studies focus on developing a specific implementation of blockchain (Nærland et al., 2017; Muzammal et al., 2019), exploring potential innovations in one sector (Sikorski et al., 2017; Gordon and Catalini, 2018; Chang et al., 2020), or using economic models to analyze the (user) dynamics upon blockchain networks (Huberman et al., 2017; Cao et al., 2019). Studies on the cross-section of blockchain applications (Zīle and Strazdiņa, 2018; Casino, Dasaklis, and Patsakis, 2019; Labazova et al., 2019; Ruoti et al., 2019) usually conduct reviews in which they summarize the extant academic literature and practitioner outlets (e.g., expert blogs, industry reports). Such approaches have two limitations. First, the small size of the manually-examined data sample restricts the scope of the analysis. Most studies do not discuss changes in applications over time and omit information on innovator characteristics in their analysis. Second, classifications based on academic or industry articles could deviate from practical applications as they may concentrate on the potential of the technology and thereby idealize its actual use. This paper overcomes these limitations by conducting a textual analysis of patent data to provide a comprehensive current state of blockchain innovation.

In order to close the research gaps and limitations of previous studies, this paper aims to address the following main research questions:

- What does the blockchain innovation landscape look like from a business perspective, and how has it changed over time?
- Who are the innovators, and what are the different approaches to blockchain technology among the innovators?
- What are the discrepancies in the literature and business regarding blockchain innovation, and where are the gaps?

To answer the research questions, I apply an unsupervised machine learning technique, Latent Dirichlet Allocation (LDA), to the text of blockchain-related patent filings to create topic models and identify nuanced themes of blockchain innovation. Patents are commonly used to indicate adoption or innovation activity and to reveal preferences in the business world for specific application areas (Daim et al., 2006). Despite some limitations (Nagaoka et al., 2010), it is one of the most comprehensive datasets that reflects the actual indication of technology development (Basberg, 1987) and identifies innovations with potential impact on the economy (Youtie et al., 2008). I apply LDA to the text (i.e., titles and abstracts) of patent filings to systematically screen the content and identify the specific topics in the blockchain application domain. The approach allows me to draw a static landscape of blockchain applications based on the distribution of topics and to illustrate the evolution of topics over time.

Subsequently, I match the identified topics from patent filings with applicants' information to discern the innovators and their focused topics within the blockchain space. Among the existing research, only a few studies associate the different innovation entities with specific blockchain applications. Many argue that startups are the primary source of blockchain innovation, and the innovations that come from them tend to be more disruptive to the market (Barraza, 2019; Chen et al., 2019b). However, only a few researchers have studied the role of startups in blockchain application. Fiedler and Sandner (2017), for instance, analyze the startups' social media impact to identify the top blockchain startups globally, noting that most of the top startups are active in the financial sector. Similarly, Friedlmaier et al. (2018) find that although blockchain-based startups are present across all sectors, the financial and communication sectors are the most prominent areas in the blockchain startup ecosystem. Given the dominance of the financial sector in the sector distribution among blockchain startups, Beinke, Nguyen, and Teuteberg (2018) develop a taxonomy to examine the business models of startups in the financial sector. The above studies analyze the startup ecosystem in blockchain but rarely compare startups and incumbents. This paper examines the applications of both types of enterprises and compares their innovation focus to associate the specific topics (not only industry) in the blockchain landscape to discover if they have distinctive focuses.

My paper complements the existing literature by empirically examining textual data from a sizable patent filing data sample to provide a comprehensive overview of current and historical blockchain applications. This framework can guide researchers in identifying future research areas and provide practical implications for practitioners regarding their business and innovation opportunities in blockchain-related sectors.

3.2 Literature

Blockchain applications are primarily linked to cryptocurrency. The adoption process of blockchain in other areas was relatively slow until the emergence of Ethereum in 2015, allowing developers to create a variety of decentralized applications on top of the platforms, thus boosting blockchain innovation (Zhao et al., 2016).

Literature reviews are the most common way to study and provide an overview of blockchain applications. Researchers have found different patterns in their studies. Abou Jaoude and Saade (2019) find the Internet of Things (IoT) to be the most frequently mentioned application area, while finance and healthcare are less frequently discussed. On the contrary, Frizzo-Barker et al. (2020) conclude that the most significant number of papers in their sample focus on banking and finance. The review of Casino, Dasaklis, and Patsakis (2019) indicates that business-oriented applications (i.e., applications used for a specific sector) dominate, followed by governance, IoT, and data management. Like Abou Jaoude & Saade, they find that only a few papers focus on finance, despite its essence in the early stages of blockchain development.

It can be argued that industry sources are more practice-oriented and better reflect the actual uses of blockchain technologies. Zīle and Strazdiņa (2018) utilize a mix of academic papers and industry articles, and their results show that data management and data verification are the most popular usages of blockchain, and most of these use cases are used to serve the financial sector. Ruoti et al. (2019) use only industry sources and performed a manual textual analysis to extract the use cases. They note that while the applications can be divided into different categories, they are interconnected both between and within categories. All components of the blockchain (e.g. technical and business) serve the purpose as a whole.

Despite these connections, the classifications in current business-related studies tend to consider only the applications and neglect the technical designs of blockchain. Recognizing this problem, Labazova et al. (2019) develop a taxonomy of blockchain that captures both applications in specific sectors and the technical dimensions of blockchain. Their study also notes that the different technical characteristics of blockchain are suitable for different types of application areas.

This study aims to give an overview of the blockchain innovation landscape beyond the limitation of different layers. Therefore it does not separate the applications that focus on the technical designs and those tailored to serve the specific sectors but presents a landscape of blockchain application, including both aspects. In doing so, this paper identifies the discrepancies in the importance of topics between literature and business applications and generates directions for further research.

3.3 Data and Methodology

3.3.1 Patent Data

I obtain patent filing data from the study of Clarke et al. (2020), in which they built unique search queries to obtain blockchain-related patent families from the European Patent Office (EPO) worldwide database (which contains patent filings from major patent offices around the globe) between 2009 and 2018. Due to the long processing time of patent grants, the data availability of granted patents for a new technology like blockchain is limited and could result in a truncation bias (Chen et al., 2019b). Hence, published patents are used in this study. To serve my purpose, I keep only the first-filed application for one patent family to avoid text duplication. The dataset consists of 4,034 blockchain-related patent filings between 2009 and 2018 from patent authorities worldwide. The title and abstract texts of patent filings are utilized in the textual analysis.

Among all the applicants, enterprises are of the essence in blockchain innovations. On the one hand, sizeable established companies, such as International Business Machines (IBM), hold many blockchain-related patents and have a significant stake in the blockchain application space. On the other hand, start-ups play a vital role in the global blockchain innovation map. To differentiate enterprises, I divide the enterprise applicants into start-ups and non-startups (subsidiaries are considered stand-alone). First, in this study, a start-up is defined as a company not founded more than eight years before the patent filing (Chen et al., 2019b). Second, it is not a subsidiary of established firms at the time of filings (Farre-Mensa et al., 2020). The company-specific information is retrieved from BvD Orbis and Crunchbase, supplemented by web search, if necessary.

3.3.2 LDA

The patent classification code gives a view of which technology category the patent pertains to but does not reflect the content of the patents in great detail. Blockchain-related patents tend to fall in a narrow range of classifications (Clarke et al., 2020), making it impossible to differentiate the blockchain-related patents into sub-categories accurately. Therefore, I apply the topic modeling technique LDA (Blei, Ng, and Jordan, 2003) on the texts of my patent filings to capture the actual contents of the patent files.

Data Preprocessing

Following the guidance of Bird et al. (2009), I use Natural Language Processing approaches to clean my textual data set. Since the patents are filed from patent offices in different countries, the texts are of multiple languages. Therefore, firstly, all non-English titles and abstracts are translated into English. Subsequently, I lowercase all the capital letters, remove punctuations, special symbols, and stop-words[1] from texts. I then tokenize all the texts by vocabulary and give part-of-speech tags (i.e., noun, verb, adjective, and adverb) to identify each vocabulary's property. Based on the part-of-speech tags of words, I lemmatize each word to its base form. This process avoids the impact of vocabulary variation (i.e., plural form, time tense, comparative, and superlative degree) on word recognition. After such processes, I get my corpus for further analysis, which contains 4,034 patent titles and abstracts with 14,144 unique terms. The most frequent words in the corpus are data, transaction, system, information, and node. The overall term frequency is presented by a word cloud in Figure 3.1.

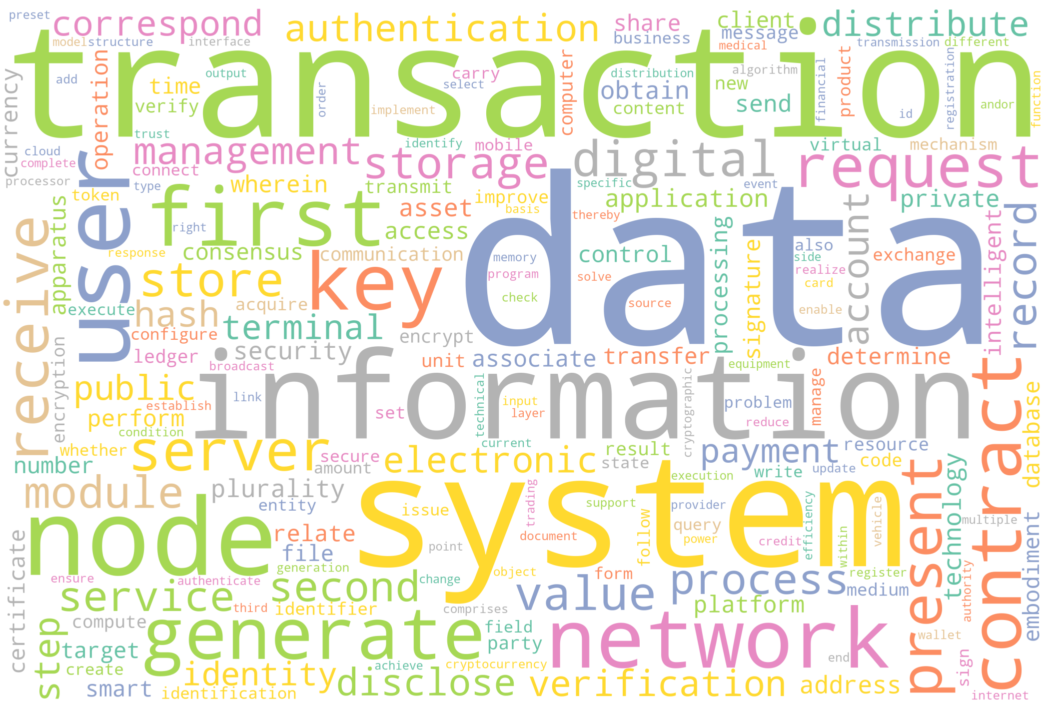


FIGURE 3.1: Structure and relationship among four essays within the frame of the dissertation.

LDA Topic Modeling

Concept LDA is a generative probabilistic model for text corpora developed by Blei, Ng, and Jordan (2003). It assumes that each document in the corpora consists of a number of latent topics, and each topic is characterized by a word distribution. Each topic is presented with a list of words and their possibilities for fitting into the topic. The operational process of LDA is as follows (Cong et al., 2021): (1) Researchers determine

[1] I added personal stop-words, which include words representing general concepts of blockchain or standard words in patent files (e.g., blockchain, invention).

the number of topics in the whole corpora, K; (2) A word is randomly assigned to one of the K topics; (3) If the word and topic do not match, the word will be assigned to another random topic. The previous steps are repeated in the corpora until the optimal match is found.

LDA is used extensively in computational linguistics but has also gained popularity in accounting and finance literature. Dyer et al. (2017) analyze the 10-K disclosure to understand the disclosure trend. Hansen et al. (2018) use Federal Open Market Committee meeting transcripts to examine the impact of transparency on communication measures. Huang et al. (2018) conduct a textual analysis on conference call transcripts and subsequent analyst reports to study analysts' role as an information intermediary.

In this study, LDA is chosen because the texts used in the model are not labeled. The algorithm itself constructs the texts' patterns, and the researchers' goal is to describe and interpret the patterns in a meaningful way (Hastie et al., 2009). The method allows me to examine the data in a more exploratory way. The classifications of blockchain applications in extant literature are constructed from various perspectives. Thus, fitting my documents into one of the classifications may cause confusion. Therefore, I use LDA to understand the corpus structure before creating my classification.

Model Implementation To construct my LDA model, I firstly build a dictionary of words in vector representation. Common words appearing in more than 80% of the patent filings are dropped since they offer general information and do not point to specific topics; rare words appearing in less than ten patent filings are also excluded due to their insignificance. The final dictionary has 1,941 unique words.

I use the coherence score as my primary measure of model performance. It measures how well the topics generated from the corpus are coherent: Topics are distinct yet support each other as a whole. Researchers have used co-occurrence frequencies of terms within a corpus to measure coherence, and it is shown to reach close to the quality of human interpretation (Newman et al., 2010; Mimno et al., 2011; Röder et al., 2015).

I then determine the best number of topics, K, for my model through the coherence score. I run the LDA model with K from 10 to 100 (with interval length 5) and calculate the coherence score (Röder et al., 2015) for each K. Figure 3.2 shows that K= 10, 15, and

20 generate the highest coherence scores among the number of topics I tested. For this reason, *K*=10, 15, and 20 are chosen as my primary options. Next, I run the model using *K*=10, 15, and 20, respectively, and use the results for further evaluations.

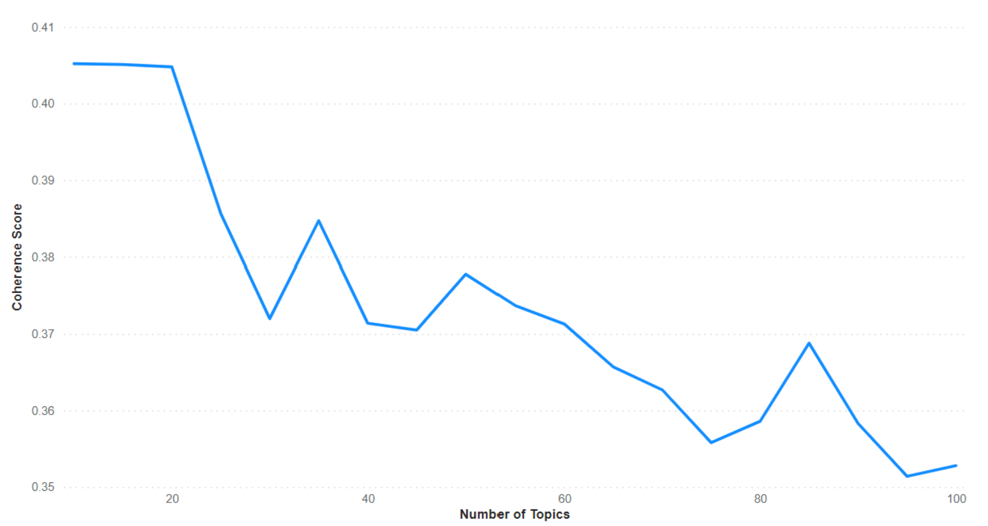


FIGURE 3.2: Coherence score of the LDA model from K=10 to K=100.

In the light of interpreting the outputs from the model, two additional researchers and I label the topics based on the keywords and domain knowledge (Huang et al., 2018). Comparing the labels of models with *K*=10, 15, and 20, we find that *K*=10 and 15 tend to tangle some topics into one and fail to distinguish topics frequently mentioned in the literature, such as IoT and cloud services. Based on the above criteria, I use 20 topics in my model.

Subsequently, two research assistants and I manually and independently examine ten documents on each topic to finalize my interpretation. For each topic, we choose five documents that are most significant in that topic and five random documents. We read through the documents independently to confirm or update the labels. We then discuss our results together and reach a consensus on the labels. The final topic names, information on their relevance in the sample, as well as examples for each topic, are presented in Table 3.1.

3.4 Results and Discussions

3.4.1 Descriptive Results

As shown in the first graph of Figure 3.3, of the 4,034 patent filings, 3,124 (77%) are from enterprises, and more than half of the enterprise patents are filed by startups. The non-startup applicants filing the most patents are large technology companies such as IBM, Alibaba, and BOE Technology, and the top three startup applicants are Coinplug, NChain, and Fuzamei Technology. The other 23% of the patents are filed by individuals, non-profit entities, and governments.

The second graph of Figure 3.3 shows the distribution of applicants by industry. Companies in the information and communication sector (e.g., telecommunication, computer programming) filed 1,790 (44%) patent applications. Enterprises in the financial and insurance sector applied for 518 (13%) patents, and enterprises in the manufacturing sector filed 374 (9%) patents. When referring to industry information, it should be noted that the industry information does not necessarily reflect the exact direction of a particular patent of the applicant. For example, a software company belongs to the information and communication sector, but it may file a patent to develop a payment platform that serves the financial sector.

The third graph of Figure 3.3 classifies the country (or geographical region) of origin of patent filings. Institutions are sorted by their headquarters and individuals by their nationality. China and the United States are the two countries with the highest number of patent filings, counting for 54% and 24% respectively. They are followed by South Korea (5%) and the United Kingdom (3%).

3.4.2 Topic Modeling Results

Topic Distribution and Trends

The results and topic labels from my LDA model, along with business examples for each topic are illustrated in Table 3.1. It should be noted that the topics identified are located in different layers of blockchains, which are all needed for a complete blockchain application. Therefore, they should be seen as guidance for the directions of the applications rather than a strictly exclusive classification. For a broad overview of the topics, I also group my twenty blockchain-related topics into four categories. The category principles are based on the combination of existing literature, domain knowledge, and two-dimensional virtualization of topic connectivity by *pyLDAvis* (see Figure 3.4[2]).

[2]Please refer to Appendix B for the numbering of topics.

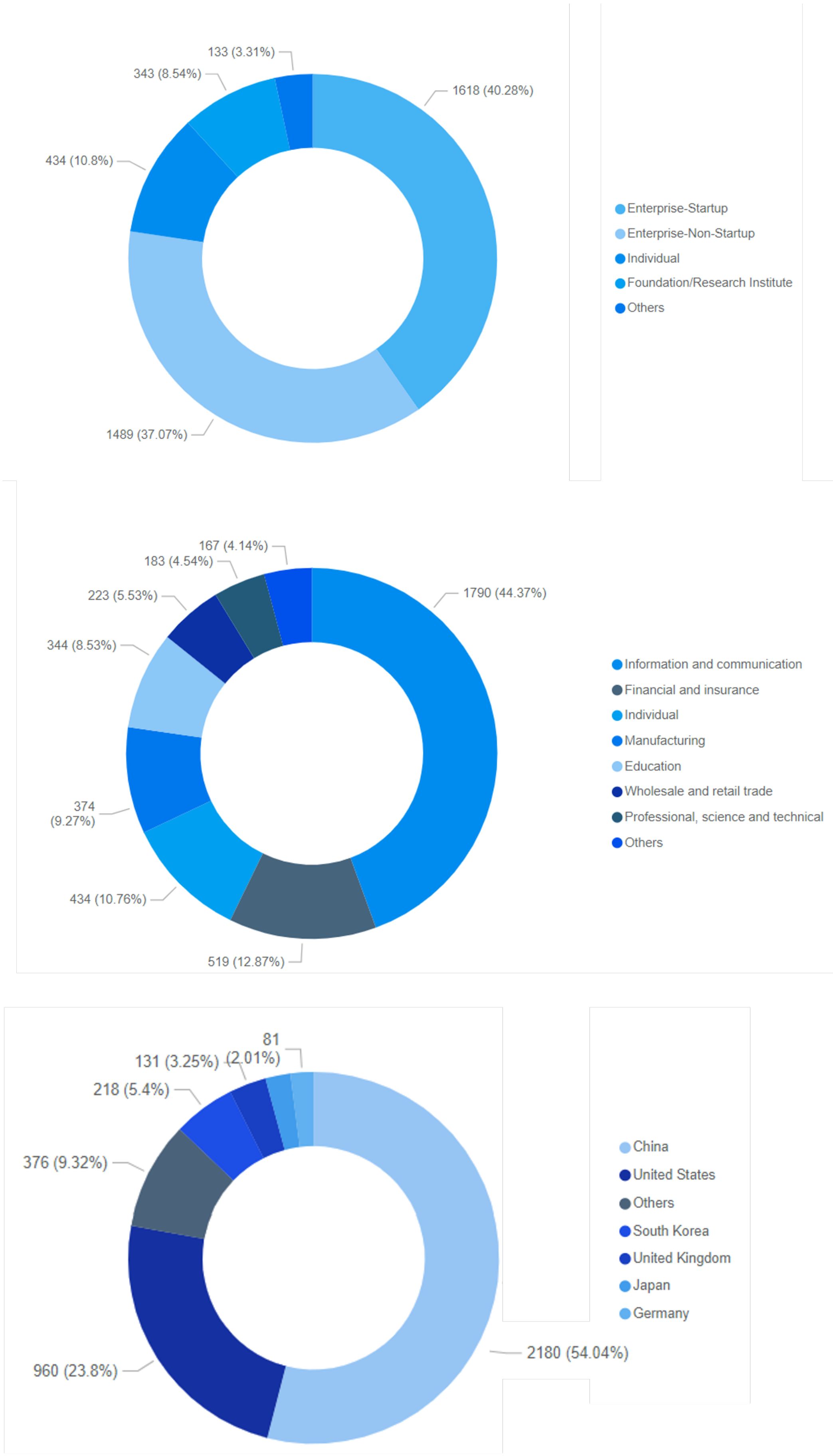


FIGURE 3.3: Distribution of patent filings in terms of applicant type, industry, and country.

TABLE 3.1: Distribution of topics, categories, and relevant examples in each topic.

Category	Topic	Top Five Keywords	# of Patents	Percentage	Examples
Blockchain Design	Consensus Mechanism	Consensus, Network, Time, Verification, Mechanism	265	6.57%	Protoblock
	Public Key Cryptography	Key, Public, Private, Signature, Encryption	222	5.50%	ChinaDB
	Data Processing	Request, Target, Processing, Process, Send	293	7.26%	Hithingschain, Ledger Domain
	Smart Contract	Contract, Smart, Intelligent, Execution, Execute	136	3.37%	Ethereum, Blossoom
	Network Infrastructure	Management, Platform, Technology, Share, Security	463	11.48%	ChromaWay, Livepeer
	Hash Function	Hash, Number, Generate, Random, Domain	162	4.02%	Bubi
Financial Uses	Trading Platform	Trading, Member, Trade, Bitcoin, Fund	130	3.22%	Oneconnect, Hashlynx
	Transaction Processing	Transaction, Address, Network, Account, Request	206	5.11%	Omnibazaar
	Cryptocurrency and Payment	Payment, Account, Token, Cryptocurrency, Financial	162	4.02%	Bitcoin, Litecoin
	Digital Asset Management	Digital, Asset, Currency, Virtual, Transfer	143	3.54%	Blockchains.com, Meridio
Data Management	Document Digitalization	Document, Output, Task, Input, Proof	133	3.30%	Docuseal
	Secure Storage	Storage, File, Medium, Content, Store	156	3.87%	Stampery
	Cloud Service	Service, Server, Client, Application, Network	133	3.30%	Wanglu Tech
	Healthcare	Electronic, Record, Medical, Audit, Time	114	2.83%	HSBlox, Rongzer
	Inter-organizational Data Management	Distribute, Associate, Ledger, Network, Resource	539	13.36%	Qbrics
	Identity Verification	Authentication, Identity, Verification, Terminal, Authenticate	186	4.61%	Bloom, Blockchains, Zamna
	Certificate Management	Certificate, Server, Register, Issue, Public	109	2.70%	Sony Global Education, Photochain, Blockchain for Art
	Multimedia	Vehicle, Power, Source, Energy, Plurality	144	3.57%	Verifi.media, Scenarex
Physical Goods	Supply Chain Management	Product, Code, Trace, Physical, Commodity	135	3.35%	Ripe.io, XATP
	IoT	Unit, Communication, Control, Network, Security	203	5.03%	modum, Teleinfo

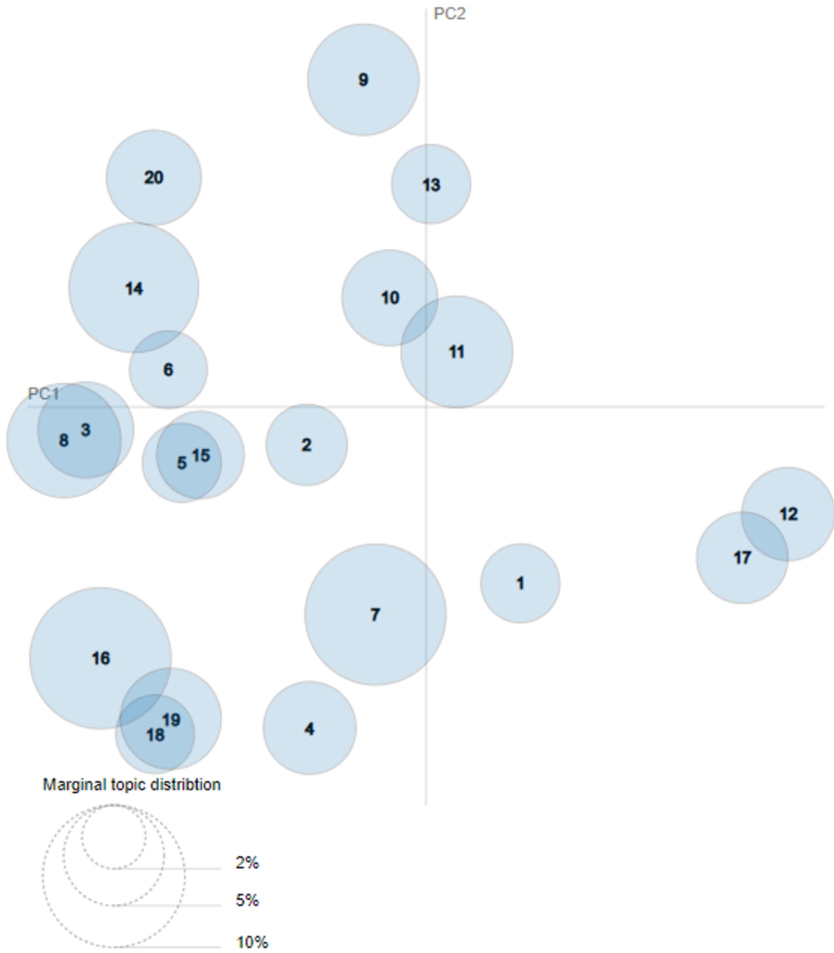


FIGURE 3.4: Two-dimensional visualization of the blockchain-related topics from LDA topic modeling.

Figure 3.5 depicts the number of patents for each category per year and helps to identify trends in the topics. The total number of patent fillings was low until 2015. Since then, the number of patent filings has started to increase rapidly. Before 2015, 344 patents were filed, whereas in the post-2015 period, the total number of patent filings was ten times higher at 3690. Due to the significant growth in 2015, I split my dataset into pre-2015 (excluding) and post-2015 (including) periods to further examine the distribution and changes of topics. Figure 3.6 and Figure 3.7 plot the distribution of topics and categories in the pre-2015 and post-2015 periods, respectively. In the following sections I will discuss the topics in detail.

Blockchain Design Behind the interface, developers build up the infrastructures on which all the blockchain applications can be implemented. Each component in

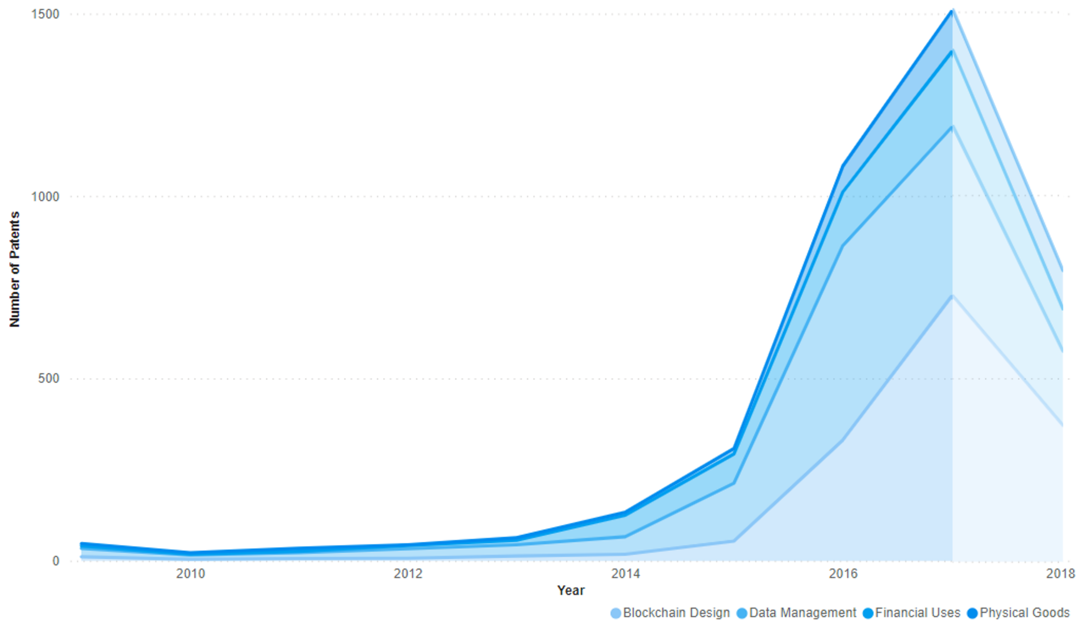


FIGURE 3.5: Trends of topics in years between 2009 and 2018 (Patent data for 2018 up to October 1, 2018).

blockchain (e.g., hashing functions, public-key cryptography, and consensus mechanism) can be fundamentally designed with such distinctive characteristics in various ways (Pilkington, 2016). The designs aim to reach particular performance or quality standards aligned with the desired functionalities (Xu et al., 2017). Blockchain design comprises 38% of the patent filings, the highest among the four categories. Before 2015, the share of blockchain design topics was comparatively low (17%), but it became increasingly crucial after 2015 (40%).

In the pre-2015 period, the main topics were the fundamental component design of the blockchain, while in the post-2015 period, the focus shifted to network infrastructure and data processing. The increased diversity of blockchain applications increased the demand for different infrastructures that can adapt to various use case needs, thus leading to a substantial increase in patent filings in this category.[3]

High data processing capabilities (e.g., fast and secure processing capacities, reliable data output) are also required due to the increasing usage. A large part of the blockchain discussion in data processing is related to the challenges of blockchain

[3]For instance, *ChromaWay* integrates the relational database concept with blockchain technology and develops its relational blockchain. It brings the benefits of relational databases (e.g., data independence, no redundancy) to blockchain infrastructure. Practical solutions are then built upon it to serve certain enterprises' needs.

adoption. The rapid development of blockchain requires higher scalability and interoperability, respectively, and therefore, many blockchains have been designed with scalability to handle large workloads quickly and smoothly, and the need for cross-blockchain communications has driven more projects in blockchain interoperability.

Also, smart contracts started to appear and grow. The introduction of Ethereum in 2015 created smart contract and brought up dramatic opportunities to build applications on the top of blockchain networks. A smart contract is essentially an automatable and enforceable digital agreement (Clack et al., 2016). The idea of smart contract in blockchain was brought up by Ethereum (2014), but many blockchains have built-in smart contracts that can implement their transaction logic (Dinh et al., 2018). With smart contracts, any legal agreements or trade deals on blockchains can be automatically executed based on the terms or contingencies written in smart contracts. It vastly increases the contractibility and enforceability of contracts in businesses without the involvement of trusted third parties (Cong and He, 2019). Since its relevance in business, many blockchain projects have begun to have smart contracts embedded or based on smart contracts.

Fundamental design elements of blockchain are not only on the technical layer, which is separated from applications built upon those blockchains. Instead, it represents the basis on which the actual direction of use can be determined. It comprises a large portion of blockchain-related patent filings and plays an increasingly vital role in developing blockchain applications. However, it is seldom discussed in existing business literature (Labazova et al., 2019).

Financial Uses Financial uses are the starting point of blockchain applications that instigate early applications. The patent distribution reflects the same pattern. In the pre-2015 period, financial uses comprised 27% of the total patent filings, with cryptocurrency and payment taking the highest percentage (14%) within the category. The very first usage of blockchain, namely Bitcoin, intended to build a peer-to-peer network with cryptocurrency Bitcoin that cut off the financial institutions in transactions. Though only a digital record of ownership, cryptocurrency is used the same way as cash or currency and can be traded through numerous exchanges (e.g., Coinbase and Bittrex).

The post-2015 period saw the proportion of financial uses decrease to 15%, and the concentration of patents in cryptocurrency declined. The focus was spread to many other areas. Namely, the number of filings focused on digital asset management and transaction processing exceeded cryptocurrency and payment. A more diverse application pattern in the financial sector started to appear.

Conventional financial instruments trading can be built via a blockchain-based trading platform. Such a platform matches the institutional traders in the loan market with trade requests without a broker and facilitates the loan settlement process. Hence, it reduces the traders' searching and transaction costs and provides traders with faster and more flexible settlement arrangements (Chiu and Koeppl, 2019).

Any asset, even physical goods, can be tokenized to digitally represent the asset ownership in the blockchain by a token. The token contains the owners' rights and legal responsibilities directly embedded into the token, along with an immutable record of ownership. It adds transparency to transactions and enables people to see the ownership history of tokens (Laurent et al., 2018). The tokens can also be traded for investment purposes. By shifting into this model, people can easily track their assets, release greater asset liquidity, and reduce the cost of capital (Conley et al., 2017). The increasing transaction volumes in many blockchains stimulate new transaction methods to streamline the processes, increase the capacity, and ensure the authenticity and security of transactions on the blockchain.

The financial sector embraced blockchain early thanks to its digital property, and it has attracted much attention from researchers and practitioners. Studies have found that financial- and banking-related applications in the current literature and industry articles are highly present (Labazova et al., 2019; Frizzo-Barker et al., 2020). However, the results of this paper show that the total amount of patent filings related to financial uses accounts for a relatively low ratio (16%) in the blockchain innovation landscape, particularly in the later period.

Data Management Data management comprises 38% of the patent filings in our sample, and especially in the pre-2015 period, it accounted for nearly half of the patents (46%). Within this category, inter-organizational data management is the most significant topic. The blockchain structure ensures security and information transparency among different parties, making it fit for inter-organizational data management. A blockchain network can digitize and store physical documents on the blockchain with a timestamp to become tamper-proof. Only authorized parties can access the data using appropriate cryptographic methods, ensuring privacy and security. Data management keeps its relevance (37%) in the post-2015 period. Moreover, the topics in this category have become more diversified over time, meaning the usages have mainly expanded.

Some applicants develop a blockchain-based cloud service platform to store and distribute data among multiple parties effectively. Blockchain offers decentralization and immutability features to combat the centralization and easy-to-tamper problems of the current cloud services, bringing the cloud service to another stage. In addition, more data management applications for specific purposes have been developed. Current healthcare can be primarily improved through blockchain. Although many healthcare systems have adopted Electronic Health Records, the data exchanges of patient medical records among medical organizations (e.g., hospitals and health centers) are minimal. They keep the records internally and rarely supplement them with other entities' data, which makes it hard for patients to obtain a complete medical record (Zhang et al., 2018). Blockchain shifts the focus from organizations to patients and leads to more patient-centric interoperable data exchanges driven and controlled by patient (Gordon and Catalini, 2018). Interoperability in data exchanges enables a patient's records from different sources to be seamlessly shared, facilitating the building up of a patient's complete health history. [4] All the data shared and disseminated around the system is ultimately owned by the patient, thus reducing the risk of data misuse by any entities (Agbo et al., 2019). Blockchain can also meliorate pharmaceutical issues by integrating a blockchain supply chain into the health system. More details will be discussed in Physical Goods.

Digital identity management has also become an increasingly important usage of blockchain. In integrating digital technologies into daily life, many services require users' basic information, such as name, address, and credit record, to be verified (Cao

[4]Prescription information, data from mobile health devices, and health insurance claim records, for instance, can all be collected and shared to complete the health history. For more detailed discussions, please read Gordon and Catalini (2018) and Agbo et al. (2019), and McGhin et al. (2019).

et al., 2019).[5] Governments in some countries have also developed systems to provide digitalized ID over blockchain to ease the redundant processes in government services (Hou, 2017).

Certificates can also be managed in blockchain networks. The physical certificates are prone to be forged or lost, and cross-verifying specific certificates among different systems/countries often requires a third-party notary (Maesa and Mori, 2020). For instance, for international students, the verification of transcripts and certificates could sometimes be very complex. Sony's Global Education division launched its blockchain-based education platform, where students can digitally manage their certificates and share them with other educational institutions. The authenticity of the certificates is guaranteed through blockchain.[6]

Data management has an extensive range of uses and can fit into many sectors' particular requirements. Although the focus has been given to data management in the financial sector in the literature (Zīle and Strazdiņa, 2018), this study finds that the application of blockchain data management is relatively diversified and has reached many sectors.

Physical Goods Patent filings related to physical goods on blockchain comprise 8% of the filings. The amount of filings has grown steadily over the years. The parties involved in today's supply chain can be geographically very dispersed. It enhances the companies' sourcing opportunities and extends the business scope, but such complexity also faces constraints. Firstly, information flow in the supply chain is usually carried by stand-alone information systems, which makes it problematic for other organizations to have an overview of all the procedures (Saberi et al., 2019). Such organizations can gain extra power for possessing the information, leading to a single point of failure (Abeyratne and Monfared, 2016). Secondly, the physical transfer of essential documents among parties makes the paper documents vulnerable to damage, loss, and fraud (Nærland et al., 2017). Additionally, the accurate and timely tracking of product

[5]Blockchain startup *Zamna*, for instance, specializes in passport and ID management for travelers over a blockchain. It securely shares identity data with travel organizations (e.g., airlines and hotels) in corporations to streamline travel difficulties.

[6]Another difficult-to-verify certificate is media integrity. Blockchain startup verifi.media sets up a network in the music industry to track down the origin of the work and prove the rights of each party (e.g., music creator, label, streaming service provider) in each song.

information (i.e., provenance, location, and conditions) is hard to achieve under current supply chain systems (Kshetri, 2018).

Blockchain provides transparency, traceability, and accountability into supply chain processes to solve the above problems (Saberi et al., 2019). Blockchain offers visibility into all the transaction and logistic details, allowing each participant to easily access the data of their interest and better understand the whole process. No single party will gain more power by controlling crucial information (Catalini and Gans, 2016), freeing the system from a single point of failure (Wang et al., 2019). The immutability of blockchain prevents malicious parties from manipulating the data, hence unleashing the benefits of visibility.

Supply chain management applications differ from many others because they require the verification and tracing of physical goods. It poses difficulties and generates extra costs to use a blockchain because one needs to bridge the distributed ledgers and physical goods with offline attributes (Catalini and Gans, 2016). In this context, IoT is often integrated into a blockchain system to capture all the relevant real-time information of physical goods at every stage in a physical industrial environment. Furthermore, in supply chain management, the information of goods, which includes the provenance, quality, and up-to-time status, needs to be traced and verified. To bridge the digital system and physical goods, such blockchain applications are often connected with the 5G network, IoT, and Radio-frequency identification (RFID). They are capable of identifying and tracking the products and enabling the data exchange among devices, therefore enhancing the functionality of blockchain-based blockchain systems (Bocek et al., 2017; Dewey and Plasencia, 2018; Tian, 2016).

The improvement of supply chain processes by blockchain not only impacts individual supply chains but facilitates the social and environmental supply chain sustainability (Saberi et al., 2019): Accountable and immutable information on blockchain prevents corrupted entities from illegitimately obtaining assets from people; the transparency through the whole process gives better assurance of human rights and work practices. Environmentally, the accurate tracking of goods reduces the rework and recall, decreasing resource consumption.

IoT is fundamentally a network of smart items (Dai et al., 2019): In the bottom, it contains a perception layer, in which devices (e.g., sensor, RFID tag) are embedded in physical items and collect the items' data. Upon the perception layer is the communication layer, in which communication devices (e.g., wireless/wired devices) are connected with IoT gateways to form a network and transmit the data in the network. With such a structure, IoT can capture many types of data, including activities and movement of the products. However, when incorporated with blockchain and smart contracts, it can also control the quality of products or the payment related to the products or services. For instance, in a pharmaceutical supply chain where the products are sensitive to temperature changes, the product sensor collects the temperature data and sends the data to the network. The data will be distributed to relevant parties in real-time through a blockchain network. The temperature information triggers the acceptance or rejection of the shipment based on the quality requirements written in a smart contract. Once accepted, the payment agreed on smart contract can be automatically executed (Rejeb et al., 2019).

The strength of IoT is that it allows communication among machines and devices. Combining it with blockchain, they can be used for various purposes, including the automatic payment of energy consumption based on smart contracts or wearable IoT devices for clinical trials and precision medicine (Fernández-Caramés and Fraga-Lamas, 2018).

Blockchain integration with supply chain and IoT has drawn massive attention from researchers and has found an essential topic in academia and industry (Abou Jaoude and Saade, 2019; Cao et al., 2019; Casino, Dasaklis, and Patsakis, 2019). Notwithstanding, as discussed above, bridging the digital and physical worlds is still a problem in many cases. The recent development of the 5G network would broadly release the potential of IoT and supply chain applications. It could vastly increase the capacity of real-time data transmission, thus promoting the IoT and supply chain applications rapidly (Dewey and Plasencia, 2018).

Topic Focus of Enterprise Applications

From the previous analysis it can be seen that the majority of the patent filings came from enterprises from China and the United States. Therefore, I further examine and compare the patterns of patent filings from these two countries. Figure 3.8 shows the patents filed from these two countries by year. Before 2016, the United States led the number of patent filings, but then China started to file more patents and overtook the United States in the number of patent filed. In the United States, more patents were filed by established firms than by startups in most years. In particular, large financial services and technology companies such as IBM, Mastercard and Bank of America account for the largest number of filings. A contrasting pattern can be seen in the case of China: in the early years, both types of enterprises filed few patents, but after 2015, significantly more patent filings came from startups than from non-startups.

Diving deeper into the topic distribution of different companies in these two major countries (see Figure 3.9), I find divergent focuses of different applicants. First, the focus of blockchain-related topics is more similar at the country level than the company type level. Chinese companies focus on consensus mechanisms, data processing, and network infrastructure, all of which are included in blockchain design category. On the other hand, companies in the United States have more distinct focuses. In general, they focus on cryptocurrency and payment as well as inter-organizational data management. Second, there is a difference in focus among startups and non-startups (mostly large established companies). In China, startups pay more attention to network infrastructure innovation, while non-startups focus on data processing. In the United States, the differences are more substantial. Inter-organizational data management is the dominant focus of non-startups. It is also an essential topic for startups, but they have a more diversified approach to innovation. The focuses include document digitalization, multimedia, cryptocurrency and payment, and identity verification, primarily specific domains for data management.

3.4.3 Summary and Discussions

The Growing Number of Patent Filings

The total number of patents filed has started to grow dramatically since 2015. This can be attributed to several factors around this time. Firstly, the launch of the Ethereum platform could be a primary factor influencing the increase in patents. It expanded the possibility of running smart contracts. Coming with a turing-complete programming language, it allows developers to create a variety of decentralized applications on top of the distributed platforms (Tasca and Tessone, 2019), thus greatly increasing not only the number of blockchain applications but also the diversity of them (Zhao et al., 2016). Second, there was a surge in corporate interest around this time. For example, IBM was involved in multiple blockchain projects, including supply chain solutions and the Hyperledger project, and invested heavily in the technology. Around the same time, the R3 consortium was formed, attracting major banks to develop blockchain-based solutions for the financial sector. The interest of large companies also catches the attention of the media and the public, further increasing the interest in blockchain innovation. Third, venture capital investment in blockchain startups increased significantly around 2015 (Redman, 2015), boosting innovation in this field.

The Changing Trend of Topics

From a broad perspective, the topics have changed over the years. In the early years, the focus was on financial uses and data management, and later, blockchain design innovations experienced a dramatic increase. The blockchain uses associated with supply chains and IoT have been growing steadily. Looking at each topic in more detail, more diversified topics emerge over time. Before 2015, more parents were exploring the fundamental component design of the blockchain, cryptocurrency applications, and inter-organizational data management platforms, while after 2015, more diverse and sophisticated applications were discovered. In particular, more industry-specific and niche usages, such as healthcare and multimedia, were explored.

The maturation of the technology has brought more opportunities beyond the initial financial uses and data storage, forming a more robust blockchain ecosystem that includes applications that reach numerous business sectors.

Differences Among Applicants

Given the dominance of Chinese and United States firms in patent filings, I specifically examine the topics of applicants from these two countries. The distinctions among the applicants reflect national priorities and market dynamics in blockchain innovation. At the country level, the United States led innovation in the early years, while China began to file more patents than the United States post-2016. China tends to focus more on blockchain design areas, while the United States exhibits a more diverse focus. Regarding startups and established firms, no notable differences are shown in Chinese companies. However, established firms in the United States are predominantly engaged in inter-organizational data management. At the same time, startups find specific market segmentations, such as multimedia and identity verification.

The country differences can be explained by the approaches to blockchain in these two countries. In China, many innovations are primarily guided by policies, of which blockchain development is one (although cryptocurrency trading is prohibited). Supported by policy, the focus can be set to improve the current financial and data infrastructure, which lays the foundation for further innovation. In contrast, in the United States, established firms primarily lead blockchain innovation, focusing on developing comprehensive platforms with adequate resources and funding. Startups in the United States take a more pragmatic approach, identifying specific applications that lend themselves to blockchain capabilities and niche solutions. In the United States, blockchain innovation is relatively more entrepreneurial.

Research Gaps and Future Research

Comparing the prominent themes discussed in literature and patent applications, several discrepancies exist. Current literature highlights blockchain applications in the IoT, the financial sector, or general business-oriented applications. However, patent filings reveal a broader range of applications, including use cases for identity verification and multimedia. Future research can explore more specific and emerging niche areas.

Another limitation of the business-oriented blockchain literature is the interconnectedness of blockchain components. Researchers could explore how technical advances in blockchain can influence and drive the overall development and adoption of blockchain in industry-specific applications.

From an innovator perspective, the current literature also provides a limited view of blockchain innovation. For instance, innovators from different countries or types of entities tend to exhibit different approaches to blockchain-related businesses. The behaviors of blockchain startups, in particular, can be of interest, as it is intriguing to see how they interact with established firms and develop novel solutions using the technology.

3.5 Conclusion

This paper applies LDA model to analyze the textual information of blockchain-related patent filings empirically. It complements the literature by using a novel text analysis methodology to provide a comprehensive landscape of blockchain applications with insight into topic trends and innovator comparison.

Using patent filing text data in EPO worldwide database from 2009 to 2018 for LDA models, I identify and describe twenty topics in from the patents and capture not only the static topic distribution but also the topic trend and evolvement. In the early years, the topics focus on financial uses and data management, and later, more diverse and sophisticated applications are developed.

By combining topics and applicant information, I link the specific topics to different entities and reveal the distinctive innovation patterns of different types of applicants from different perspectives, namely 1) from different countries, especially two major countries, China and the United States; 2) from different types of enterprises, i.e., startups and established companies. The results indicate that the non-startups applicants initiated the innovation process, but startups got stronger and overtook non-startups in later years. The behaviors of startups in the blockchain area should be more exclusively studied to understand the role of startups in the blockchain ecosystem.

Subsequently, comparing the topic pattern in patent filings and the topics discussed in academic literature, this paper also underpins the topics with practice-relevant that are not discussed in current literature and proposes further research directions.

The patterns of blockchain applications I identify in this paper can be used as a starting point for researchers to explore the related potential research areas in the blockchain area and be used by practitioners to distinguish their potential business opportunities in blockchain-related sectors.

There is a limitation of this study that should be acknowledged. Not all inventions or innovation activities are patented. Some entities may prefer other methods to fulfill their innovations due to the high costs of filing patents (Nagaoka et al., 2010). Hence, patents potentially do not capture all the existing blockchain innovations through other channels.

The findings of this study highlight gaps in the current literature on blockchain applications and suggest several avenues for future research. Firstly, while blockchain design is a prominent theme in applications, its integration with business issues is often overlooked in business and management literature. Future studies could take a more integrative approach to exploring business applications. Secondly, the role of startups in driving blockchain innovation warrants deeper investigation, particularly their business models and impact on the progress of blockchain adoption in business contexts. Lastly, the study discusses the individual innovation patterns of entities. However, further research could explore the interactions, such as collaboration and competition, among innovators, including startups, established companies, and government agencies. For instance, how the competition and the merger and acquisition activities impact the blockchain market, and the influences of regulatory frameworks from government agencies. Understanding the dynamics and synergies could be essential for building a healthy and robust ecosystem for blockchain application development.

FIGURE 3.6: Distribution of topics in pre-2015 periods.

FIGURE 3.7: Distribution of topics in post-2015 periods.

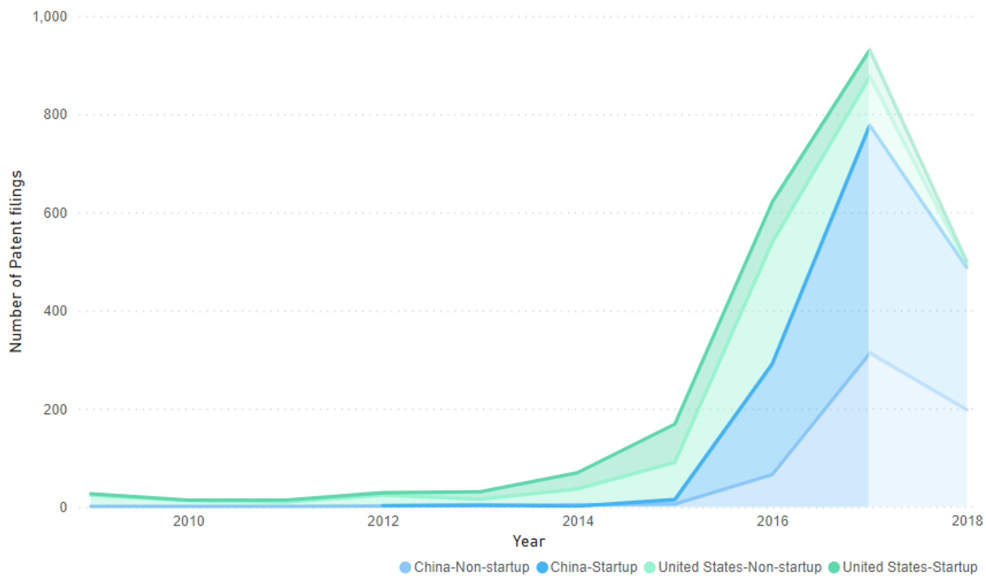


FIGURE 3.8: Trends in patent filed in years from companies in China and the United States between 2009 and 2018 (Patent data for 2018 up to October 1, 2018).

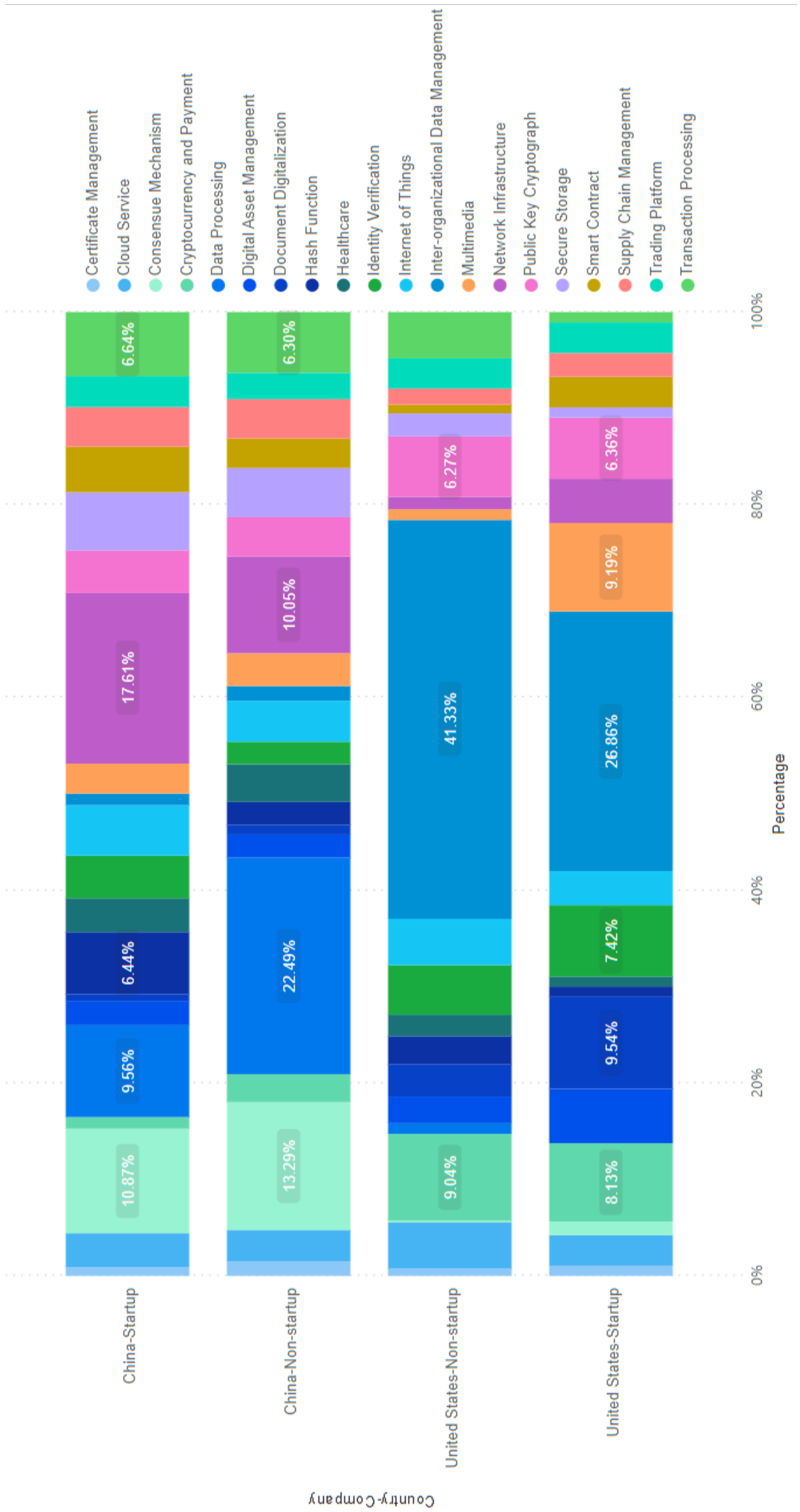


FIGURE 3.9: Topic distribution of startups and non-startups in China and the United States.

Chapter 4

Blockchain for Cross-border Payments and Financial Inclusion in Africa: The Case of Stellar Network

---

This chapter is a working paper. The manuscript is uploaded to SSRN to share the research idea and gather feedback. It can be found at: http://dx.doi.org/10.2139/ssrn.4550837

4.1 Blockchain as a Solution for Cross-border Payment

The global market for cross-border payment has experienced significant growth in the last few years. However, the conventional correspondent banking model has struggled to keep pace with market development due to its lengthy settlement periods and high costs. While countries with highly liquid currencies may perceive these issues as minor, for many developing countries with exotic currencies they pose significant challenges, affecting both businesses and individuals. In certain instances, cross-border money transfers can take up to ten days to settle, and the associated costs can exceed 10% (Cunliffe, 2020).

These shortcomings stem from various factors, i.e., the heterogeneous payment schemes in different jurisdictions, intransparent processes, limited access to financial services, and a lack of competition (Beck and Peria, 2009; Rice, Peter, and Boar, 2020). Blockchain technology is a prominent solution that promises to change the status quo. The decentralized nature of blockchain allows technology startups to enter the normally high entry barrier sector and provide more flexible and affordable financial services, especially with stablecoin applications. The emerging business models are particularly beneficial for individuals and businesses in developing countries excluded from the incumbent financial systems to securely store and transfer their wealth across borders (Kshetri, 2017; Thomason et al., 2018; Norta, Leiding, and Lane, 2019).

However, the academic literature in business focuses more on the holistic benefits and implications of blockchain for cross-border payments, and rarely delves into the specific mechanisms and business cases. Moreover, the discussion on blockchain and stablecoin tends to focus more on developed countries and strong currencies such as the US dollar, while developing countries, where the potential lies, are less often studied. This study aims to fill this gap by laying the focus on developing countries and exotic currencies in Africa.

We use the Stellar network[1], a blockchain network with numerous technology startups focused on cross-border payments, as a case study to illustrate a mechanism designed for blockchain solutions for cross-border payments. We then delve into the concept of stablecoins and their importance in blockchain-based cross-border payment

[1] https://stellar.org/

solutions, followed by the challenges of adoption in the context of developing countries. Furthermore, to examine the solutions that address the local needs of African countries, we present three Africa-focused blockchain startups on Stellar that are tackling local problems and offering distinct approaches to address relevant issues in the region. These cases provide real-world examples that demonstrate the potential impact and effectiveness of blockchain solutions in specific contexts.

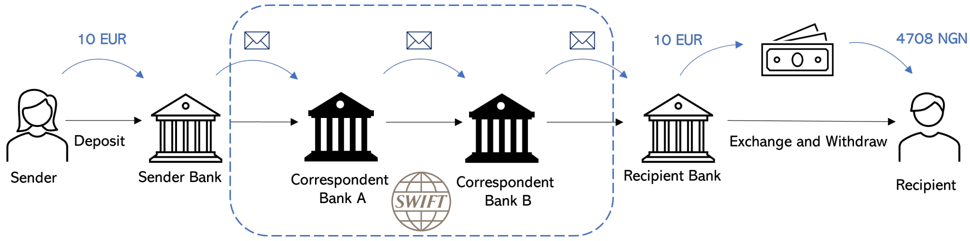
4.2 Correspondent Banking and Its Drawbacks

Correspondent banking is the most dominant model for conventional cross-border payment (World Bank, 2021). The initiation of cross-border payments by clients through their banks requires a physical presence of the bank in the destination country, which poses challenges for many banks with limited international reach. Banks can establish partnerships with correspondent banks that possess the requisite operational expertise on an international level to offer the service on behalf of the bank of the sender (Massacci, Ngo, and Williams, 2016). Correspondent banks impose fees for providing this service, typically borne by either the sender or the recipient. While frequently traded currencies often require only one correspondent bank, less common currencies may involve several correspondent banks (i.e., hops). Each correspondent bank is responsible for verifying the transfer against local regulations and updating the relevant accounting records. The protracted transfer chain leads to higher transaction fees and increased payment delays.

Figure 4.1 illustrates a simplified cross-border payment process using the correspondent banking model (via SWIFT). In this scenario, a sender in Germany intends to send 10 Euros to a recipient in Nigeria who wishes to withdraw Nigerian Naira (NGN). The sender deposits the amount into her/his bank account and initiates the payment. The payment message traverses the correspondent banks until it reaches the recipient's bank in Nigeria. The recipient receives a balance of 10 Euros in her/his account and subsequently exchanges the Euros for NGN at the prevailing exchange market rate.

Such a model exhibits several significant drawbacks, particularly impacting developing economies, notably in Africa. Firstly, the long chains of money transfers associated with the correspondent banking system result in elevated costs and delays.

FIGURE 4.1: An Example of simplified cross-border payment process through correspondent banks.



This issue is particularly prevalent in countries with less liquid currencies, encompassing many African nations. Secondly, the dominance of established financial institutions in the correspondent banking system limits competition and stifles innovation, impeding progress in the sector. Thirdly, one must possess a bank account to access the services the correspondent banking system provides. However, some African countries have more than 50% of their population unbanked (Merchant Machine, 2021). As a result, individuals are excluded from formal financial services, relying mainly on cash for their financial activities, depriving them of safe and efficient ways to manage their wealth. Lastly, the lack of access to formal financial services leads to a vibrant shadow economy (Medina and Schneider, 2017), further undermining the stability and health of financial markets.

4.3 Financial Inclusion and Blockchain in Africa

Financial inclusion refers to the accessibility of formal financial services to individuals and businesses, encompassing activities ranging from basic remittances to credit and insurance. By having access to financial services, people can securely store and transfer their wealth, gain opportunities for investments in education and healthcare, and start businesses. This improves their day-to-day living, enables them to plan for long-term goals, and provides a safety net for emergencies, thus a key to reducing poverty and boosting development (World Bank, 2022).

Nevertheless, many individuals and businesses in Africa still encounter friction in basic financial services, despite the importance of cross-border payments for the economies. For individuals, due to the limited job opportunities or political and economic instability, many move to foreign countries for higher salary jobs, and their families live on their remittances (Feyen et al., 2021). The money received is vital in lifting people out of poverty and economic development (OECD, 2022; UN, 2019). Small and medium-sized enterprises (SMEs) are pivotal in the economic development of Africa in contributing to employment creation and poverty alleviation (Daniel, Conor, and Janina, 2021). Still, individuals and SMEs face enormous barriers to smoothly accessing formal financial services.

Blockchain has emerged as a key technology enabler in the innovation process, offering potential solutions to the challenges of the current system and promoting greater financial inclusion. By reducing transaction fees and settlement times, introducing flexible financial products and services tailored to specific groups, and providing a transparent environment for cross-border payments (Natarajan, Krause, and Gradstein, 2017; Rühmann et al., 2020), blockchain solutions are well suited for small-value transfers, which are prevalent in Africa. Additionally, blockchain disrupts the traditional trust model (the trustworthiness of incumbents and the integrity of transactions ensured by high entry requirements) by shifting the basis of trust from a centralized authority to underlying consensus rules (Catalini, Gortari, and Shah, 2022). This reduces barriers to entry, allowing startups to enter the market and compete with established financial institutions. It therefore forces incumbents to develop more innovative and affordable solutions, expanding the availability of financial services to a broader user base and increasing financial inclusion.

4.4 Introduction to Stellar Network

Stellar is a public blockchain-based network designed to address the challenges inherent in the current correspondent banking systems (Lokhava et al., 2019). Many FinTech startups build up their applications upon the Stellar network, and therefore, studying the applications built on the Stellar network provides valuable insights into the multitude of possibilities within blockchain-based cross-border payment solutions.

4.4.1 Consensus Protocol

The disconnectedness of banking systems around the globe relies on international payment messaging systems to be connected. However, the complexity of correspondent bank processes and diverse regulations across jurisdictions pose challenges in eliminating frictions in cross-border payments (World Bank, 2021). A blockchain providing a peer-to-peer network and voting mechanism accessible to the public can be a viable alternative.

Consensus protocols are mechanisms that allow nodes (i.e., validators) to decide on the correct version of the transaction ledger without a central trusted party. It is vital to blockchain since it enables blockchain's decentralization and determines its security and performance (Xiao et al., 2020). The most commonly-used consensus protocols adopted by *permissionless* (public) blockchains include Proof of Work (PoW) and Proof of Stake (PoS) (Irresberger et al., 2023). In contrast, many *permissioned* blockchains use practical Byzantine Fault Tolerance (pBFT), which features lower energy utilization and faster transaction processing time. Yet, since the calculation complexity of pBFT dramatically increases when the number of joining nodes increases, it requires a *permissioned* environment in which only pre-selected participants can join, therefore demanding a centralized authority to manage the network (Bracciali, Grossi, and Haan, 2021).

Stellar employs a variation of pBFT called Federated Byzantine Agreement (FBA) and constructs its Stellar Consensus Protocol (SCP) based on it. It preserves the fast speed of BFT while making it feasible for public blockchains by introducing the notion of *quorum slides*. In SCP, each node autonomously selects which other nodes in the network to trust, and a group of nodes that trust each other form a *quorum slide*. By joining a *quorum slide*, nodes become validators who can vote on the validity of a ledger update. When a transaction occurs, the corresponding statement message is sent from one validator to others in the *quorum slide*. Once the agreement is reached within the *quorum slide*, the transaction is settled (Lokhava et al., 2019). Implementing such a trust model avoids the need to obtain consensus from the entire network for each transaction, allowing both fast speed and open membership.

4.4.2 Cross-border money transfer

Stellar's payment network is implemented on top of SCP. Users can issue tokens and trade them on Stellar decentralized exchange (SDEX) built into Stellar's ledger system.

The featuring *anchor* and *path payment* settings streamline the transfer and trade among assets.

Digital Assets Unlike many blockchain networks on which only native currencies can be used for transactions, Stellar allows users to create digital tokens that can be used for payments or traded against each other on the SDEX. The four most essential types of tokens on Stellar are:

- Fiat tokens (i.e., Stablecoin). Stablecoins are cryptocurrencies designed to maintain a stable value by pegged to a specific asset or a basket of assets (e.g., fiat currency and gold) (FSB, 2020). Their designs enable the ability to move stable-valued tokens faster than the actual currencies on legacy rails, making them significant in blockchain-based cross-border payments.

- Crypto tokens. These tokens represent their underlying cryptocurrencies, such as Bitcoin and Ether. Such tokens are generally created to enable faster and more flexible movements and exchanges of the underlying cryptocurrencies, which, in their own systems, can be slow and costly.

- Commodity tokens. They are backed and pegged to commodities featuring low volatility and high liquidity, such as gold or a basket of precious metals. Tokens make it easier to hold commodities without physically storing them. To keep its stability, it should be backed 100% by the commodity it is pegged to. They are less frequently mentioned due to the lack of diversified associated decentralized finance (DeFi) products.

- Native currency Lumen (XLM). It serves two roles: 1) A small amount of XLM is required to initiate the accounts and make transactions as a minimal threshold to prevent frivolous actors in the network; 2) it serves as a neutral currency to facilitate the exchange between assets.

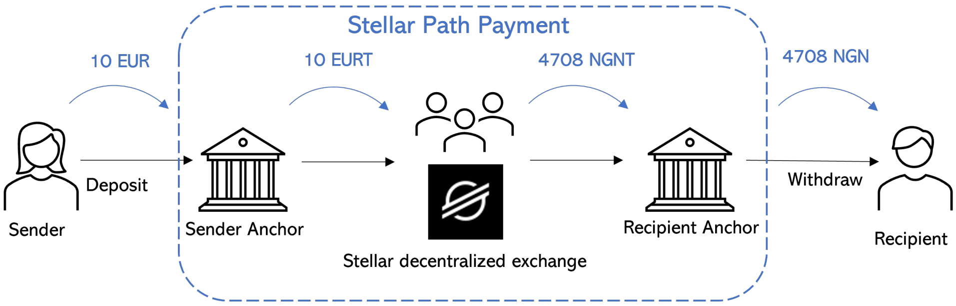
Anchor To facilitate on- and off-ramp between fiat currency and cryptocurrency, Stellar uses *anchors* as intermediaries to connect the blockchain-based and legacy infrastructure. Regulated financial institutions, money service businesses, or FinTech companies qualify to be *anchors* if they provide at least one following service:

- Issue stablecoins. When issuing stablecoins, they also need to maintain the equivalent value of reserves in insured accounts to guarantee that the users can redeem fiat currency at any time and provide timely audits for backed assets. The issuance and the reserves should comply with relevant regulations where the anchors operate their businesses.
- Offer a fiat on-/off-ramp. It is a service to exchange fiat currency for cryptocurrency and vice versa, and it should be provided through the local domestic conventional payment rails. Anchors are also responsible for fulfilling regulatory compliance and the AML/KYC requirements of the country (Khan, Ahmad, et al., 2019).

Path Payment An example of a path payment feature on Stellar is shown in Figure 4.2. The sender in Germany wishes to send 10 Euros to the recipient in Nigeria, whose account only supports NGN. The sender uses the service (e.g., e-wallet) of sender's anchor to specify the amount and the desired currency of the recipient. Sender's anchor converts the Euro into the corresponding EURT token on the Stellar network. The path payment feature determines the best exchange rate to convert EURT to NGNT (the stablecoin token representing the Nigerian Naira) on the SDEX. This process may involve the use of XLM as a neutral currency or additional hops. However, the entire process occurs in a single operation on the Stellar network, eliminating the need for the sender to compare quotes on exchanges. Once the trade for NGNT token is completed, the recipient's anchor in Nigeria facilitates the conversion of NGNT into NGN, which the recipient can withdraw as cash if desired.

Path payment differs from conventional cross-border payments in the following ways. Firstly, it eliminates the need for the sender to possess the destination currency and search for exchange offers. Secondly, path payment leverages blockchain technology to significantly reduce settlement times and lower fees. Lastly, all transaction data is visible on the blockchain, ensuring full transparency and traceability.

FIGURE 4.2: An Example of simplified cross-border payment process on Stellar.



4.5 Stablecoin: The Core of Blockchain-based Cross-border Payments

Stablecoin is a cryptocurrency designed to be price-stabilized by pegging its values to specific assets. It arose from the 2018 cryptocurrency bubble, after which it became clear that the high volatility of cryptocurrencies largely impaired their usability as a means of payment. It has proliferated in recent years, and according to CoinGecko[2], the market value has reached 131 million US dollars and accounts for 11% of the crypto market value. Thanks to its low fluctuation and close connection to the conventional financial system, stablecoins can facilitate efficient movement and exchange of currencies and open opportunities to reshape cross-border payments and promote financial inclusion (Deloitte, 2021).

4.5.1 Classification of Stablecoins Depending on Stability Mechanism

Algorithimic Stablecoin Algorithmic stablecoins operate without backed assets and instead rely on algorithms embedded in their design to regulate the token supply and control stablecoin prices. They can be implemented at either the protocol or application layers of blockchain. Stabilization at the protocol layer is complicated as it lies at the fundamental level of blockchain and requires most users' agreements. Stabilization at application is less complex and is achieved by dual coin, of which the best known is seigniorage shares. It artificially separates a certain stablecoin into two parts, *coins* and *shares*. *Shares* act like equity and are used as a tool to stabilize *coins*. When coin price goes high, and the supply needs to increase, new *coins* are minted and distributed to share holders in exchanges for *shares*, which will then be removed from circulation

[2]https://www.coingecko.com/en/categories/stablecoins, accessed on May 17, 2023.

(i.e., burned). When coin price drops low, and the supply needs to decrease, *shares* are auctioned for *coins* (Sams, 2015), which will be burned. Although the design seems sophisticated, such a stability mechanism requires the public's confidence in the coin. Especially in times of price drop, share holders should expect the coin price to increase.

Algorithmic stablecoins are fully decentralized and theoretically independent of the price fluctuations of cryptocurrencies and fiat currencies. However, such independence also brings high volatility and complication in the design of their stability mechanisms. Therefore, they are perceived as the most insecure stablecoins, which can devalue abruptly and lose the peg (Berentsen and Schär, 2019; Catalini and Gortari, 2021).

Crypto-backed Stablecoin Issuing these stablecoins involves sending the backing cryptocurrency to a governing smart contract, which then mints and distributes the corresponding amount of stablecoins to users (Bullmann, Klemm, and Pinna, 2019).

Crypto-backed stablecoins offer several advantages due to their reliance on on-chain cryptocurrencies and avoidance of centralized custody: faster and more cost-effective exchangeability of the backing currencies on the blockchain and the ability to quickly verify the status of collateral due to the high transparency. Nevertheless, since the prices of backing cryptocurrencies are volatile, crypto-backed stablecoins are, in principle, over-collateralized to absorb the price fluctuation. Many crypto-backed stablecoins also incorporate automatic liquidation mechanisms within the smart contract. If the value of the backing currencies falls below a certain threshold, the collateral is automatically utilized to purchase the stablecoin. Such over-collateralized, however, entails an inefficient and potentially risky use of capital (Moin, Sekniqi, and Sirer, 2020).

Off-chain Asset-backed Stablecoin Stablecoins backed by off-chain assets can be backed by any physical assets, although they are commonly backed by low-volatility assets such as strong fiat currencies or gold. Issuing these stablecoins requires a licensed third-party custodian to safeguard the backing assets and ensure their redemption when requested (Bullmann, Klemm, and Pinna, 2019). The reliance on off-chain assets and a single accountable issuer results in a centralized structure. Off-chain asset-backed stablecoins are typically 100% backed and exhibit high stability relative to their reference assets. However, they are also exposed to multiple risks, so capital buffers are needed to offset potential losses (Catalini and Gortari, 2021).

The disadvantage of these stablecoins is the slow redemption/liquidation process resulting from the exchanges between digital tokens and physical assets. Additionally, the costs associated with custodial storage of the collateral and regular external audits to verify its status can be significant, especially when the scale of the stablecoin increases (Mita et al., 2019).

Only highly trusted issuers, such as central banks, can ensure the integrity of off-chain asset-backed stablecoins. One extreme version of fiat-backed stablecoin is Central Bank Digital Currency (DBDC) issued by a central bank, which controls the issuance of both backing currency and the stablecoin. Therefore, it is the only stablecoin with zero volatility against reference assets. However, it comes with two drawbacks: it is issued by a conventional centralized authority, limiting competition and innovation; it may face challenges in integrating other international systems (Auer and Böhme, 2020; Catalini and Shah, 2021).

Hybrid Stablecoin To leverage the benefits and combat the drawbacks of certain stablecoins, issuers can have a basket of baking currencies or a portfolio of cryptocurrency and fiat currency. Issuers can also partner with each other. For instance, some central banks have discussed public-private corporations in issuing CBDCs (Bank of England, 2020; Catalini et al., 2021; Bolt, Lubbersen, and Wierts, 2022). In such arrangements, central banks provide the necessary payment infrastructure and reserve custody, while the private sector can contribute innovation, consumer interfaces, and product development. This collaborative approach allows efficient utilization of public and private resources, enabling fast, user-friendly, and inclusive financial services.

4.5.2 Stablecoins and cross-border payment

The high volatility of many cryptocurrencies has positioned them more as investment assets rather than reliable payment instruments (Arner, Auer, and Frost, 2020), leading individuals and businesses to seek the combination of cryptocurrency flexibility and fiat currency stability for cross-border payments. Stablecoins, with their balance between centralized and decentralized models, offer a more suitable medium of exchange compared to typical cryptocurrencies.

Among stablecoins, those pegged to the US dollar have the highest trading volumes. This can be attributed to the strength and dominance of the US dollar in the global financial system (Cohen, 2015). Additionally, in countries with unstable monetary policy or high inflation, individuals prefer to hold US dollars to preserve the value of their wealth. This global dominance of the US dollar extends to the world of cryptocurrencies.

Another group of stablecoins is the tokens pegged to fiat currencies in developing countries, especially exotic currencies that are relatively illiquid. Despite blockchain solutions' potential to increase financial inclusion in developing countries, such stablecoin projects on Stellar are at a small scale compared to US dollar stablecoins.

Stablecoins facilitate the implementation of blockchain-based solutions in cross-border payments by 1) enabling users to hold stablecoins they prefer without being bound by geographical restrictions, providing a means to hedge against inflation or exchange instability in countries with high inflation rates (e.g., Zimbabwe); 2) reducing the impact of currency volatility during cross-border payments, hence providing certainty and confidence to users.

4.5.3 Adoption Challenges

Design and Standardization

The design of reserve assets is one of the fundamentals in stablecoins, and it determines the stability mechanisms. Apart from primary stability mechanisms, stablecoins utilize secondary mechanisms, including fees/redemption limits, targeted rebates, and reactive mining rewards to regulate the token supply and stabilize prices (Bullmann, Klemm, and Pinna, 2019).

The distinct designs of stablecoins lead to a lack of standardization. The taxonomy of stablecoins is still under debate, and they are categorized as different instruments in different jurisdictions (Ostercamp, 2022). Off-chain assets-backed stablecoins are frequently seen as e-money, payment systems, or money market funds, while for crypto-backed stablecoins, the classification can be more complicated.

The lack of standardization sets difficulties for both stablecoin issuers and regulators to make progress in promoting stablecoins. Recognizing this, standard-setting bodies like the Committee on Payments and Market Infrastructures (CPMI), the International Organization of Securities Commissions (IOSCO), and the Financial Action Task Force (FATF) have engaged in examining the application of international standards to stablecoins. These bodies are working at both domestic and international levels to advance uniformity in stablecoin arrangements and encourage international cooperation.

Regulatory Framework

The increasing diversity and popularity of stablecoins have raised concerns regarding regulatory compliance for individual stablecoins and the potential impacts of global stablecoins on existing financial systems. These concerns are particularly relevant in many African countries with less robust financial systems.

Financial integrity is a crucial area of focus. Cryptocurrencies have been widely used for purchasing illegal products or services and money laundering due to their (pseudo-)anonymity nature and the lack of regulations. Thus, they are closely associated with unethical and illegal activities (Foley, Karlsen, and Putniņš, 2019), which could destroy the market integrity and decrease investor confidence in the market. To ensure financial integrity, the corresponding AML and Countering-terrorism Financing (CFT) measures covering the use of stablecoins should be implemented by the regulators.

Stablecoins have also raised the concern of consumer protection. Many existing stablecoin projects are not forced to regularly disclose the accurate details of their backing assets, which leaves consumers unclear about the risks they are exposed to and potentially blinded by false information. Safeguarding consumer data privacy in the context of stablecoins also presents challenges. For example, a lengthy discussion of whether the General Data Protection Regulation (GDPR) could cover blockchains was carried out due to the ambiguous nuances between data stored on public blockchain systems and other consumer private data (Finck, 2019).

Other risks of stablecoins include, among others, cyber security concerns and operational risks, which could all undermine the use of stablecoins in a financial system (FSB, 2020). To combat the problem, some authorities, such as the EU, have taken action to launch relevant regulations. Yet, the fast pace of technological development compared to financial law has resulted in most jurisdictions not yet enacting specific regulations for stablecoins.

More concerns are raised when the scale of a particular stablecoin is large enough, which makes it systematically important (CPMI and IOSCO, 2022) and poses systemic risks (also through the amplification of the above-mentioned risks) to the current financial system and affects the market stability (FSB, 2020). Although no stablecoin currently poses a significant threat to the financial system, regulators remain concerned about the long-term risks due to the rapid growth of stablecoins.

The first concern regards the fair competition policy. It sounds contradictory because stablecoin cross-border payments are intended to break the monopoly and encourage new players to enter the market. However, if the scales go up and the regulations fail to keep pace, fast-growing stablecoin projects could dominate the market and rule out other projects.

The absence of standardized stablecoin design, backing asset, and reserve requirements makes stablecoins susceptible to sudden value drops despite pegging to specific assets. Such value losses change the investment and saving decisions of many individuals. They could further cause liquidity problems and trigger panic among investors, exacerbating the situation and provoking broader market instability. Furthermore, the cross-border flow of stablecoins can facilitate rapid capital flows overseas (Catalini, Gortari, and Shah, 2022), potentially being used to circumvent capital controls in certain countries (Baydakova, 2021). This trend can accelerate the "dollarization" of weaker currencies, disrupting local financial policies and weakening the position of fiat currencies.

The challenges of global stablecoins come not only from their large scale but also the fact that they operate across multiple jurisdictions with different regulatory frameworks. This scales up the complexity of regulation and governance. As a result, individual jurisdictions may be unable to monitor and combat the risks adequately, and corporations among jurisdictions are strongly required. At the international level, multiple organizations are in constant discussions to set international standards and provide guidance for stablecoin-related arrangements (Arner, Auer, and Frost, 2020). In contrast, Africa has been slower in implementing stablecoin regulations due to factors including limited regulatory capacity and priorities focused on other financial stability concerns.

Social and Cultural Aspect

Although technological advancement and regulatory progress are the most discussed topics in blockchain adoption (Kshetri, 2018), these factors are insufficient for mainstream adoption. Blockchain should not be treated merely as a technological tool but as a socio-technical system interconnected with multiple socio-cultural aspects (Shahaab et al., 2020). Professional investors and institutions may take the first action in the stablecoin development, seeing benefits and potentials. However, for others, the decision to adopt involves impacts other than only financial considerations.

Hence, a vital task for blockchain adoption is understanding the specific contextual issues that arise from different societies and cultural backgrounds, as such issues are usually profoundly rooted in the uniqueness of social features. Only by doing so can solutions that tailor local features be developed and configured, therefore being accessible and accepted by the people who need them most (Shin and Ibahrine, 2020).

Knowledge of blockchain Hence, a vital task for blockchain adoption is understanding the specific contextual issues that arise from different societies and cultural backgrounds, as such issues are usually profoundly rooted in the uniqueness of social features. Only by doing so can solutions that tailor local features be developed and configured, therefore being accessible and accepted by the people who need them most (Donovan, 2012). Technical details are not necessary to be understood, but fundamental information, benefits, and risks of blockchain should be conveyed. In the process, social nudging can trigger individuals to associate blockchain models with positive perceptions and induce trust (Di Prisco and Strangio, 2021). For example, media articles can educate the public about the benefits of blockchain and increase awareness and exploration of the technology. Businesses can do it by showcasing established companies that have successfully implemented blockchain solutions or by offering rewards and incentives to users.

Financial and digital literacy Financial literacy refers to the ability to understand financial concepts, effectively access and use financial services, and employ the knowledge to make sound financial decisions. Unfortunately, many African countries have low financial literacy among adults (Klapper, Lusardi, and Van Oudheusden, 2015). Such ignorance stops people from searching for better ways to save and transfer money, leaving their financial conditions unimproved.

Another obstacle to blockchain adoption is digital literacy, which means one has the skills to operate digital devices (e.g., smartphones) and perform digital financial transactions. While Africa is experiencing rapid digitalization in the banking sector through FinTech, there is significant heterogeneity in digital literacy levels across African countries – Kenya, Nigeria, and South Africa have higher and more diversified digital skills than other countries (Choi, Dutz, and Usman, 2020).

Improving financial and digital literacy is a long-term endeavor that requires collaboration among various stakeholders, including governments, educational institutions, and non-profit organizations. As the report (Klapper, Lusardi, and Van Oudheusden, 2015) revealed, financial-included people generally have higher financial literacy. Companies should design user-friendly interfaces and focus on providing essential financial services, such as savings and remittances, so that individuals with low financial and digital literacy can incorporate these technologies into their daily lives and gradually gain more financial knowledge through their usage.

Fianncial habits Cash preferences, money transfers, and borrowing habits are all hurdles blockchain payment models face (Larios-Hernández, 2017). Despite the inherent risks, the unbanked and underbanked populations rely heavily on cash and alternative methods for their financial activities. These entrenched habits make it difficult for them to transition to digital solutions.

To fully unleash the benefits of blockchain solutions, government support is instrumental in driving the shift in financial habits by establishing the necessary infrastructure and providing policy support (Abu Daqar, Arqawi, and Abu Karsh, 2021). For companies, integrating new models into pre-existing habits is another vital aspect that could enhance acceptance. Rather than transform the payment system to completely cashless, the service providers could cooperate with local banking systems to offer cash-in and cash-out infrastructures (without opening a bank account) in remote areas.

4.6 Localized stablecoin projects in African countries

US dollar stablecoins have the potential to be used to mitigate the effects of inflation and economic instability in some countries. Paradoxically, the decreased barriers to holding the US dollar caused by the introduction of US dollar stablecoins have raised concerns that it could lead to further dollarization or even the displacement of fiat currency (Prasad, 2022; Blockworks, 2022). Such a tendency will essentially hinder the possibility for these countries to conduct their own monetary policy and could further deteriorate the financial situation.

The coexistence of high unbanked population rate and technological advances pushes African countries to overgrow in this area (Makina, 2019), and many African projects have developed solutions to smooth the cross-border payment problems with the local stablecoins. Thanks to the low administration costs and simple implementation, Stellar has been one of the primary choices for such startup projects.

We intend to determine how projects aimed at African countries operate considering local contexts. Therefore, we examined three blockchain startups on Stellar that focus on cross-border payment solutions with different market target groups in African countries. Cases were chosen based on three criteria: 1) focus on cross-border payments in the African market; 2) issue African country fiat-currency pegged stablecoin; 3) trading volume ranks among the top 100 on Stellar Network. Information is collected through companies' websites, whitepapers, social media accounts, industry articles, and news exposures.

4.6.1 Cowrie

Cowrie is a FinTech startup incorporated in the UK and Nigeria, focusing on blockchain-based cross-border payments in Nigeria. The high unbanked population and the support from the Central Bank of Nigeria give innovative startups opportunities to thrive. Cowrie issued Nigerian Naira Token (NGNT), a stablecoin pegged to the

Nigerian Naira, and offers an on-/off-ramp from stablecoin to fiat currency through the Nigerian banking system.

The primary target group of Cowrie is businesses in Nigeria that need to make international payments. This is especially helpful for SMEs seeking opportunities in foreign markets due to limited local markets but usually lack adequate funds to afford international business transactions through the conventional banking system. The development of these firms makes a massive difference in the economy because SMEs play a significant role: they account for 70% of industrial employment and around 50% of manufacturing output in Nigeria (Ogunmuyiwa and Okunleye, 2019).

In this case, Cowrie does not primarily deal with completely unbanked groups but with small businesses that need more affordable and efficient international business-to-business payment processes. Furthermore, to ensure easy redemption, Cowrie is connected to the Nigerian Interbank Settlement System (NIBSS) so that it has access to all the banks in Nigeria in the banking system.

To facilitate business expansion for Nigerian SMEs into the European market, Cowrie has partnered with Tempo, a blockchain-based payment institution based in France and the issuer of EURT, a stablecoin pegged to the Euro to open markets to new payment corridors between Europe and Nigeria, allowing users to make payments using their stablecoins and to redeem the local fiat currencies, cutting out the redundant procedures.

4.6.2 Afreum

Afreum is a decentralized ecosystem founded in the UK that focuses on building a financially inclusive economy in Africa. Regarding cross-border payments, AFR is the featuring stablecoin whose value at launch is pegged to the average of a basket of all African currencies against the US dollar (100% backed by USDC) and can be a hedge against currency fluctuations.

To enhance financial inclusion in African countries, Afreum developed two types of wallets. The first is the Afreum Wallet, a non-custodial wallet that can be accessed through website or smartphone applications and allows users to manage their funds. The second is the custodial Afreum SMS Wallet, designed for users without access to conventional banking services or who only have legacy phones. Users can start using the wallet via an intermediary that accepts cash and makes transactions on behalf of the user through SMS-based messages. The latter wallet compromises the users' full control over their money due to the existence of a custodian. Yet, it gives the unbanked access to basic financial services even when they do not have devices for digital services.

To change the perception of blockchain-based systems in the region, Afreum has developed a game engine with educational content about personal finance and blockchain. Users are rewarded with tokens that can be used in the ecosystem and exchanged for fiat currencies. The gamified approach aims to motivate users to increase their financial and digital literacy, gradually shifting their perceptions of blockchain and fostering acceptance of the transition to digital solutions.

4.6.3 Uhuru

Uhuru is a blockchain-based FinTech startup founded in South Africa in 2020 with the primary goal of offering Zimbabwean immigrants and workers in South Africa an affordable and convenient way to transfer money back to Zimbabwe. Many Zimbabweans seek job opportunities in South Africa due to the economic crisis and the collapse of the Zimbabwean fiat currency. Around 80% of the nearly one million Zimbabwean diasporas live in South Africa (Ndlovu, 2022), creating a significant demand for remittance services between the two countries. However, the average cost of remittance price lies at 13.88% of the money transferred and may take days to complete (World Bank, 2022).

One primary stablecoin of Uhuru is ZAR, which is pegged to the South African Rand. Zimbabweans in South Africa can transfer ZAR to their families in Zimbabwe, who can receive the money and withdraw it upon request. Considering the complications of technology, Uhuru has integrated its wallet into WhatsApp. The users can send messages to a bot to select services in the chatbox of the same app they use to communicate with families without being tangled with technology terms. Another function of Uhuru is the possibility of paying for utilities and services (e.g., electricity and cable TV). This feature allows users to cover essential expenses for their families directly, ensuring that funds are allocated for specific needs.

One significant problem of cross-border payments for many Zimbabweans is the lack of the identity document they need to open an account and use the services. Even for people with identity documents, individuals living in remote areas have difficulties proving their identities to service providers due to the poor digital infrastructure. Therefore, people are forced to use informal channels, such as Hawala operator's points, to transfer money (IMF, 2009).

Using blockchain-based solutions for essential cross-border payments may require a lower level of identity verification, but accessing broader services requires identity verification due to compliance requirements. To address the issue, Uhuru cooperates with FlexID, a Zimbabwean startup focusing on self-sovereign identity (SSI), to offer users digital identity credentials. Using blockchain, SSI creates tamper-proof digital identities with which users can control whom they share identity information. People with official identity documents can create a digital version of the documents. In contrast, various alternative verification methods (e.g., community-based verification) can be used to establish digital identity for individuals without government-issued documents.

4.6.4 Leveraging Blockchain for Cross-border Payments

The global problems of current cross-border payments are universal, yet each jurisdiction faces heterogeneous issues. To effectively unearth blockchain's potential for cross-border payments, particularly in African nations where the aim is to enhance financial inclusion, it is imperative for companies, especially startups with limited resources, to devise solutions that are tailored to the specific needs of their user base. Upon examining the cases, we have pinpointed four elements that could facilitate startups designing their products.

Firstly, it would be beneficial for companies to consider developing stablecoins pegged to local fiat currencies, allowing users to easily transfer in local currencies and reducing their dependency on foreign currencies. Secondly, users with limited technical expertise may be reluctant to transfer to new platforms, due to high learning effort and habit change. Integrating money transfers to pre-existing channels, such as

WhatsApp and SMS, can reduce the learning curve, encouraging adoption and usability. Thirdly, unbanked and underbanked in different regions face varying obstacles when attempting to access formal financial services, and such heterogeneity requires distinctive solutions. A specific target market allows companies to develop tailored solutions that address the unique needs of these groups, increasing adoption rate and user satisfaction. Moreover, catering to specific segments, companies can differentiate themselves on the market. Customizing features to align with local contexts, such as utility payment and games, can enhance the relevance and value of the products.

Lastly, for startups, forming collaborations with other companies within the ecosystem is a critical factor for success. Companies can consider both horizontal and vertical collaborations. Horizontal collaboration involves partnering with entities that offer similar services but operate in different markets (e.g., Cowrie and Tempo) to expand markets and foster interoperability. Vertical collaboration involves cooperation with complementary businesses (e.g., Cowrie and NIBSS, Uhuru and FlexID) to leverage existing networks, expertise, and infrastructure, enhancing the effectiveness of the products.

4.7 Concluding Remarks

This paper introduces the potential of blockchain solutions in cross-border payments to increase financial inclusion in African countries. It uses the case of the Stellar network to explain the payment mechanism which has advantages over the conventional systems. It further elaborates on the concept of stablecoin, followed by the challenges it faces. In particular, the social and cultural factors are of importance in shaping blockchain adoptions in African countries. This highlights the necessity for localized solutions tailored to the specific needs of particular regions. Three cases of Africa-focused blockchain startups were presented to provide insights into the practical blockchain applications in addressing cross-border payment challenges. Five points that can facilitate companies in designing their business models are given: local stablecoin, easy-to-use channel, specific target market, special product, and corporations.

In summary, despite the proliferation of projects, the blockchain case for cross-border payments in developing countries is still nascent. The establishment of standardization and regulatory frameworks is progressing rapidly to keep pace with technological advances. To realize the promise of blockchain to increase financial inclusion, innovative solutions are needed to address the unique challenges that exist in each region. Such outcomes can only be realized through the collaborative efforts and support of governments, financial institutions, educational institutions, NGOs, and other stakeholders. Overcoming the technical, regulatory, and social challenges will unlock the potential of blockchain in cross-border payments.

Chapter 5

Who cares about Decentralized Finance?

Evidence from DeFi Attacks

This chapter is to be included as a book chapter in *Handbook of Blockchain Analytics* (Springer).

5.1 Introduction

Decentralized Finance (DeFi) protocols have emerged as one of the major categories of applications in the blockchain ecosystem, gaining economic relevance and attracting substantial capital deployment (e.g., Total Valued Locked TVL in DeFi applications such as Decentralized Exchanges, borrowing & lending protocols, among others) (see, e.g., Irresberger et al., 2023). However, this success comes along with an economic interest in attacks on the security of blockchains and the applications they host. For example, the European Securities and Markets Authority (ESMA) highlights that DeFi comes with major risk of exploitation[1]. Meanwhile, centralized exchanges and borrowing & lending institutions, i.e., CeFi institutions, are competing fiercely with their decentralized counterparts on liquidity and user convenience. Many major centralized custodians, for example, Mt. Gox and Binance, have been hacked and siginificant user funds have been stolen directly from them. While DeFi's non-custodial access to complex financial intermediation protocols offers a clear advantage over centralized options such as the security against double-spending attacks (Pagnotta, 2022), it may still suffer from economic protocol exploits, among other shortcomings.

Many DeFi attacks exploit vulnerabilities specific to a particular blockchain or its smart contract platform. If these vulnerabilities are not promptly addressed and patched, it can undermine trust in the blockchain's security, affecting the value of its native cryptoasset. Any attack on the blockchain's infrastructure layer will therefore inevitably test the demand for the underlying native cryptoasset, which then will be reflected in the asset's exchange rates on cryptoasset exchanges. Likewise, users adopt blockchains because of the decentralized applications (dApps) that they host (see, e.g., John, Kogan, and Saleh, 2023) as they derive economic utility from being able to use digital marketplaces, speculate using decentralized finance platforms, or for token issuance (**IJMS**). Thus, if dApps' viability is undermined by economic or other attacks that undermines the security and functioning of the platform (cf. Daian et al., 2020; Qin et al., 2021), the demand for the underlying cryptoasset may be relatively reduced as users can migrate to other blockchains' dApps with similar functionality to avoid recurrence of the security issues.

[1] https://www.esma.europa.eu/sites/default/files/2023-10/ESMA50-2085271018-3349_TRV_Article_Decentralised_Finance_in_the_EU_Developments_and_Risks.pdf

Therefore, one hypothesis of this study is that DeFi attacks result in negative valuation effects on the underlying blockchain's native cryptoasset needed to operate on a given blockchain.

However, the relative impact of DeFi attacks on the underlying blockchain's native asset valuation may be insignificant or even positive. DeFi ecosystems are interconnected, and an incident in one DeFi project can have a spillover effect on others so that an attack on one platform will inevitably (negatively) affect others as well, resulting in negligible relative change in market valuations (Ji et al., 2019; Corbet et al., 2018).

When an adverse event occurs, such an event attracts media attention, which may even lead to more people learning about the associated DeFi application or infrastructure, and ultimately increase the demand for that blockchain. In addition, blockchain technology, including the DeFi application, heavily relies on the community for system development. In response to a DeFi attack or failure, the respective blockchain user communities and project developers can work together to solve the problem and enhance the overall resilience of the DeFi ecosystem through improved security measures. This will have a positive impact on the native cryptoasset, even if the attack appears to be a negative event in the very short term.

The second hypothesis of this study is that DeFi attacks may result in positive valuation effects on the underlying blockchain's native cryptoasset.

In this paper, we provide an overview of DeFi protocol attacks and test our hypotheses by exploring how investors in the underlying blockchain's native cryptoasset react to such exploits and attacks on DeFi protocols. Instead of analyzing individual DeFi tokens, we focus on the valuation of the native cryptoasset to understand the impact of vulnerabilities in decentralized financial applications (protocols) on the underlying smart contract blockchain. Such blockchains allow us to compare the effects of attacks on different blockchain ecosystems. We consider attacks on DeFi protocols as a potential source of Systemic risk to the entire blockchain-specific ecosystem, especially if its native asset prices are affected by a single DeFi protocol attack.

We collect a comprehensive sample of DeFi attacks from the REKT database and estimate (event study) difference-in-differences regressions to assess the impact of DeFi attacks on the market valuation of affected blockchains. In doing so, we observe whether there are differential effects on affected (treated) platforms compared to various control groups of blockchains. In our models, we consider multiple control groups (e.g., largest blockchains) and restrict our analysis to certain types of DeFi exploits (e.g., significant losses, Ethereum only) to enhance the robustness of our models and main finding.

Our key finding is that when major DeFi protocols on a blockchain encounter attacks that lead to financial losses for their users, the valuation of the blockchain's native cryptoassets (e.g., ETH for Ethereum) increases instead of decreasing after the events. This suggests that the blockchain infrastructure views these incidents as positive developments, as they may attract more new users, lead to more resilient smart contract code, thorough audits, and assurance that the developer community is able to swiftly cope with individual economic or technical exploits of blockchain-based applications.

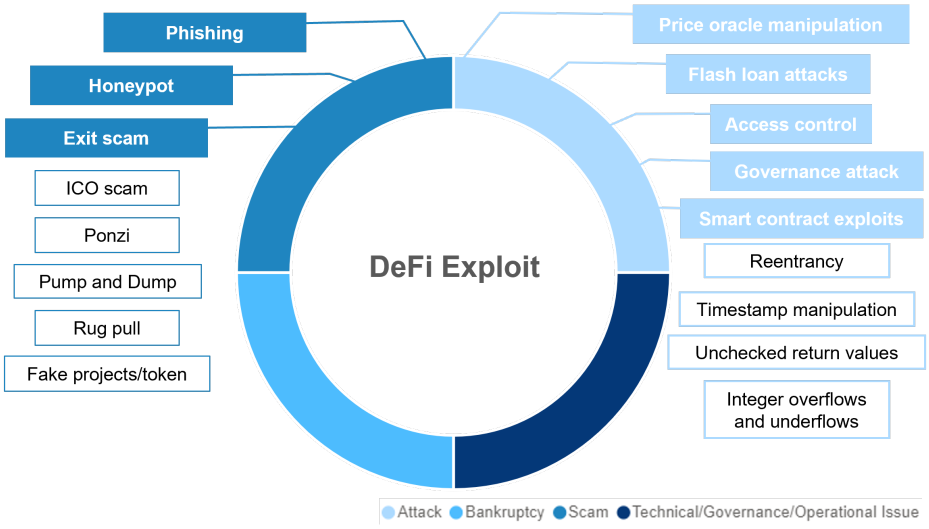
Section 5.2 provides an overview of DeFi exploit categories relevant to understanding the nature of the events we study. Section 5.3 introduces our data set and empirical methodologies we use to test hypotheses related to blockchain market valuation effects of DeFi attacks. Section 5.4 discusses our findings from the empirical analyses, and Section 5.5 concludes.

5.2 Categories of DeFi Exploit Events

While DeFi brings enormous technological advances to the financial markets, it is subject to many novel risks due to its intertwined nature and lack of oversight. Such risks may arise from, among others, external attackers, the founders or staff, or problems of governance, operation, or techniques. Given the relatively recent emergence of DeFi, the literature on the security and valuation effects of DeFi attacks is scarce. Zhou et al., 2023 provide a systematization of the knowledge, highlighting taxonomies and different types of DeFi attacks, including network, consensus, smart contract, or protocol attacks as major categories with potential security issues. Hornuf et al., 2023 provide an alternative taxonomy for Ethereum blockchain-related attacks. We consider the impact of attacks at the DeFi application level (i.e., not infrastructure or consensus issues) and evaluate whether attacks on protocols impair the valuation of the underlying blockchain's native cryptoasset, across multiple blockchains. Figure 5.1 illustrates the categories of some of the most common DeFi exploits (Li et al., 2022;

Zhou et al., 2023). Four types of exploits are listed, many of which are unique to the DeFi space. The categories are by no means exclusive and may overlap.

FIGURE 5.1: Categories of DeFi Exploits.



Attack DeFi applications are built on a smart contract-enabled blockchain to ensure the automated execution of transactions (Werner et al., 2022). Smart contract is, therefore, the backbone of DeFi applications. Once the code is deployed, it cannot be altered. If the code is flawed, the flaws can be exploited by attackers. DeFi attacks, therefore, majorly target the vulnerabilities of the code or the design of the smart contracts to disrupt the proper functioning of the smart contract and drain the funds.

Reentrancy is one of the most catastrophic types of attacks in the DeFi space and has been studied mainly in computer science research from a technical perspective (Sayeed, Marco-Gisbert, and Caira, 2020; Chen et al., 2022b). A reentrancy attack occurs when a smart contract interacts with an external contract before completing its state updates. An attacker creates a malicious contract to act as the "external contract". When the original contract interacts with it, the malicious contract callbacks or "reenters" a function (often a withdrawal) in the original contract, exploiting the delay in state updates. The importance of reentrancy attacks is illustrated by the 2016 DAO attack, in which the attackers exploited 3.6 million Ether (approximately $50 million). The event had a significant impact and led to the hard fork of Ethereum (Mehar et al., 2019).

Flash loan attack is another special attack that exists in the DeFi environment. Flash loan uses the unique structure of the blockchain to provide a way to borrow without collateral (Daian et al., 2020; Qin et al., 2021). It allows the borrower to take the loan and repay it within one blockchain transaction. If the borrower fails to repay the debt by the end of the transaction, the transaction fails and is reversed. Such a design allows the borrower to gain massive liquidity without providing upfront collateral. However, a malicious borrower could leverage the fund to exploit the vulnerability of other smart contracts or DeFi protocols (e.g., market price manipulation) to gain an unfair advantage before the borrower pays back the loan by the end of the blockchain transaction.

Scam Unlike attacks, scams are usually not purely technical. They involve more comprehensive structures and plans with the intent to defraud. Within this category, in comparison to phishing and honeypot, exit scams have a more greater impact and can result in significantly more financial loss.

Rug pull is one of the most impactful exit scams on investors. It is usually a complete project with a native token and some eye-catching features designed to attack investors. The scammer then tries to hype the project by fraudulently promoting the project by creating websites and social media accounts with misleading information, using wash trading to inflate the price of the token, etc. Once the scammer has raised enough funds, they abruptly abandon the project and leave with the money from investors. The impact of sophisticated scams is not limited to the financial losses, but can also weaken investor confidence in similar projects and the blockchain platform.

Technical/Governance/Operational Issue Governance or operational issues can occur during the operation of a project, mainly due to internal negligence or mismanagement without the intention of deception. Technical issues can be rooted in the flawed design of the products, the reduced reliability of the infrastructure and the smart contract, or problems with external systems (e.g., Oracle issues). This can result in financial losses but, more importantly, trigger external attacks if not correctly addressed in a timely manner.

Bankcruptcy Bankruptcy is the insolvency of a company, usually centralized, but is closely associated with the DeFi project. It is not a stand-alone risk, but a potential outcome of various risks and exploits in the DeFi ecosystem, including the exploits we discussed above. However, the unique nature of the DeFi environment makes the project particularly vulnerable to market volatility and the current ambiguous regulatory phase. For example, in November 2022, the collapse of one of the largest cryptocurrency exchanges, FTX, shook the market and led to the subsequent bankruptcies of related companies, including Genesis and BlockFi.

5.3 Data and Methodology

5.3.1 Data

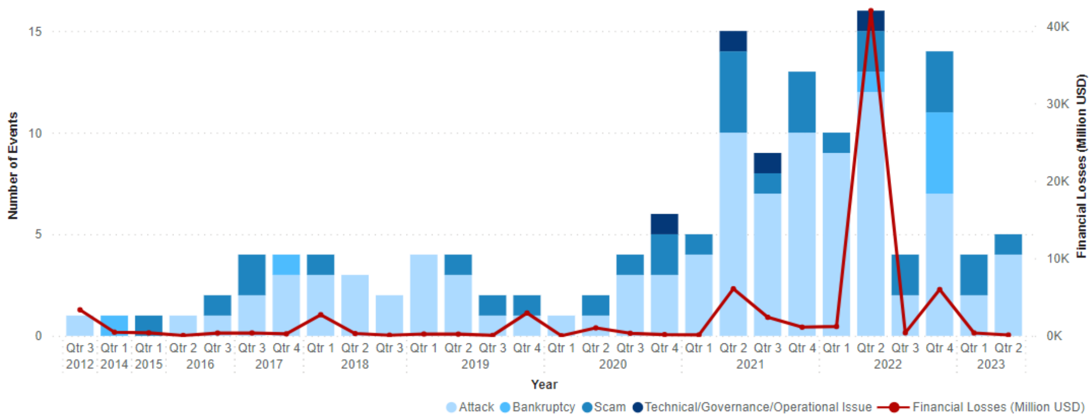
Our event set is from the REKT database, which covers relevant detected DeFi and CeFi attacks that have resulted in attackers extracting economic value from individual entities or DeFi protocols, causing protocol stakeholders to experience financial losses. It lists over 3000 events and the corresponding amount of US dollar funds lost as a result of an attack. Our sample includes all exploit events listed in the REKT database between September 1, 2012 and June 1, 2023 with more than $10 million in financial losses. This way, we obtain 143 exploit events, including 56 CeFi events and 87 DeFi events affecting 11 chains (Avalanche C-Chain, Binance Smart Chain, Cronos, MultiverseX (Elrond), Ethereum, FANTOM, Terra Classic/Terra, Polygon, Ronin, Solana, TRON)[2].

For each DeFi event, we collect the event date, the business model of entities, the type of the exploit event, financial losses in US Dollar, and affected blockchains (only for DeFi events) [3]. In addition, we include bitcoin (non-smart contract blockchain) and 49 smart contract blockchains that are never targeted by attack events as our control group sample. For each blockchain, including the treated and control group blockchains, we identify its native asset and collect daily market capitalization values from coinmarketcap.com.

Figure 5.2 gives an overview of the events on a quarterly basis. The number of events remained stable until 2020 when the number of event started to increase substantially. The majority of events were attacks, mostly exploiting the technical vulnerabilities in the systems for financial gain. These were followed by scams, which

[2]Note that we treat Layer 2 solutions as individual chains if they have their own consensus process.
[3]The March 29 2022 attack event on Ronin chain is excluded from our dataset for further modeling due to the lack of market capitalization data for RONIN.

FIGURE 5.2: Timeline of Crypto-related Exploit Events between 2012 and 2023.



were mostly perpetrated by the founders who created the projects with the intention of committing fraud.

The financial loss curve is largely influenced by single high-loss events. The most significant event was the Terra-LUNA crash which happened in May 2022 and caused approximately \$ 40 billion financial loss to investors. TerraUSD (UST), a US dollar-pegged algorithmic stablecoin launched by Terraform Labs, maintained its peg through an underlying mechanism that regulated the supply of LUNA, the native token of the Terra network. Such algorithmic stablecoins rely heavily on public confidence in the coin and can have high volatility, despite their sophisticated designs (Zhuo, Irresberger, and Bostandzic, 2023). On May 7th, the UST began to lose its peg to the US dollar, triggered by a series of large withdrawals of UST, causing a bank run. The selling pressure led to a “death spiral" that eventually drove the LUNA price to zero.

The second largest event is the rugpull scam perpetrated by the founders of a South African centralized crypto platform, Africarypt. In April 2021, shortly after the platform collected the funds from investors, investors were notified that their accounts had been frozen due to a hack. Shortly after the claim, the founders disappeared with 69,000 BTC (\$3.6 billion) collected from the investors. In April 2023, the founders were arrested in Zurich, but denied the fraud and claimed the funds were missing due to a hack. And the third largest event is a major hack on the darknet marketplace Silk Road. In 2012, the hacker pulled off the theft by exploiting a flaw in Silk Road’s bitcoin withdrawal mechanism for sellers and was able to steal 50,000 BTC (\$3.36 billion) from the system

(United States Attorney's Office, 2022). Nine years after the event, the hacker was arrested, and most of the lost funds were recovered.

5.3.2 Methodology

To evaluate the impact of DeFi hacks on respective blockchains' market values, we perform two-way fixed effects (TWFE) DID regressions in which we regress daily market capitalization of the blockchain's native cryptoasset on treatment dummies as well as chain- and time-fixed effects for each cohort (i.e., each event has its own fixed effects).

$$\text{MCAP}_{k(i),t(i)} = \alpha_{k(i)} + \lambda_{t(i)} + \beta\left(\text{POST}_{t(i)} \times \text{TREATED}_{k(i)}\right) + \rho_{k(i),t(i)}$$

$\text{MCAP}_{k(i),t(i)}$ is either the total market capitalization of blockchain k in event i or the ratio of its market capitalization and Bitcoin's market capitalization, which captures the valuation effect on the treated chain relative to the non-smart contract blockchain Bitcoin. The coefficient β of the DID term, $\text{POST}_{t(i)} \times \text{TREATED}_{k(i)}$, captures the change in native asset market capitalization (or ratio) following the treatment of a blockchain by a major DeFi attack.

We also run (TWFE) DID event study regressions (cf. Baker, Larcker, and Wang, 2022) in which we estimate treatment coefficients for each time period before and after the event in time $t = 0$. That is, for each event i, we estimate

$$\text{MCAP}_{k,t} = \alpha_{k(i)} + \lambda_{t(i)} + \sum_{l=-L, l\neq 0}^{L} \mu_l D^l_{k(i),t(i)} + \varepsilon_{i,t}$$

where $D^l_{k(i),t}$ is the treatment dummy for blockchain $k(i)$ in cohort i in DeFi hack (cohort) being $L = \{10, 15\}$ days from the start of treatment.

To build control groups of unaffected blockchains to compare to treated ones within each cohort (event), we follow three strategies. First, we sample all blockchains that are not affected by event i and do not experience overlapping events in the respective time window of length L (i.e., blockchains not treated in event i). Second, for each event, we construct two control groups. The first one includes all the not effected blockchains mentioned in first strategy. In the second control group, we restrict our sample of control chains to the largest four (as of market caps one day before the event window starts). This alleviates concerns that mainly smaller chains are used as control observations to compare against mainly larger chains that are affected by DeFi security issues. Third, in order to identify the factors of event scale and platform characteristics, we run all of the analyses above in three types of event settings. 1) full event sample, 2) the events with more than 100 million US Dollar in financial losses ("large events"), and 3) all events on the Ethereum blockchain.

5.4 Results and Discussion

5.4.1 TWFE DID regressions

Our analysis begin with TWFE DID regressions. The results of these regressions are detailed in Table 5.1, where 24 models are presented, each varying in their estimation windows, event settings, and choice of control groups.

For the all-event sample, we find a positive effect of DeFi exploits events within a ten-day window on the blockchain market capitalization when benchmarked against all control chains. In comparison, when estimating the market capitalization ratio of the blockchain against bitcoin, a significant positive effect is observed across both ten- and fifteen-day windows. Narrowing down the control group to the four largest blockchains, the positive effect only remains within the ten-day window for the market capitalization ratio.

Interestingly, when we solely consider large events that result in financial losses exceeding $100 million (Set 2), regardless of the composition of the control group, the events exhibit neither a significant impact on the market capitalization of the blockchain platform nor on the market capitalization ratio of the blockchain to bitcoin. Notably, all eight models of ETH events present statistically significant positive impacts on native asset market values.

5.4.2 DID event study regressions

The TWFE models provide insights into the average effects of DeFi exploits on blockchain platforms. To further explore the day-to-day effects within the estimation windows, we

TABLE 5.1: TWFE DID regressions with estimation windows of 10 days and 15 days.

	10 Days				15 Days			
	Mcap blockchain		Mcap ratio		Mcap blockchain		Mcap ratio	
	Set 1 - All Events							
	A All chains				**A All chains**			
Post X treatment	5,244,491,842	*	0.0069	**	4,703,499,605		0.0072	**
Prob.	0.0613		0.0227		0.1472		0.0400	
R^2	0.0076		0.0200		0.0044		0.0150	
	B Four largest chains				**B Four largest chains**			
Post X treatment	4,723,051,490		0.0061	*	3,881,056,266		0.0056	
Prob.	0.1440		0.0563		0.2922		0.1317	
R^2	0.0138		0.0258		0.0028		0.0085	
	Set 2 - Large Events							
	A All chains				**A All chains**			
Post X treatment	2,547,524,131		0.0025		2,653,221,133		0.0029	
Prob.	0.6241		0.6587		0.7131		0.6750	
R^2	0.0029		0.0200		0.0200		0.0037	
	B Four largest chains				**B Four largest chains**			
Post X treatment	(2,181,311,825)		0.0042		57,002,336		0.0050	
Prob.	0.7672		0.4414		0.9951		0.4710	
R^2	0.0020		0.0116		0.0000		0.0108	
	Set 3 - ETH Events							
	A All chains				**A All chains**			
Post X treatment	11,398,000,000	**	0.0137	***	10,837,000,000	**	0.0140	**
Prob.	0.0118		0.0064		0.0426		0.0186	
R^2	0.0507		0.0724		0.0278		0.0535	
	B Four largest chains				**B Four largest chains**			
Post X treatment	11,398,000,000	**	0.0137	***	10,837,000,000	**	0.0140	**
Prob.	0.0137		0.0076		0.0470		0.0211	
R^2	0.0511		0.0730		0.0280		0.0541	

Significance levels are denoted *** p<1%, ** p<5%, * p<10%.

employ DID event study regressions, giving us a detailed view of valuation dynamics. Figures 3-5 display the results of 24 models with the same settings as the TWFE DID regressions. Prior to the treatment dates, all models consistently show insignificant results, supporting the parallel trend assumption, as there are no observable significant differences in market value changes prior to the attack event.

In the all-event sample regressions (see Figure 5.3) with all control chains, positive effects are revealed in both ten- and fifteen-day windows on the market capitalization value and ratio. The positive effects diminish after eight days of the events toward the end of the estimation windows. Next, when using the four largest chains as control group, more significant impacts are revealed in comparison to TWFE models. The positive effects on the market capitalization value are found in the ten-day window between the third and eighth days post-event. For the market capitalization ratio, positive effects are found, but the effects periods are earlier than for the all-chain control group.

The results for the large-event sample (see Figure 5.4) mirror those from the corresponding TWFE models. All eight models indicate no significant effects on market capitalization value and ratio, confirming that the substantial financial losses from events do not impact the market capitalization value of the blockchain. Finally, Figure 5.5 depicts the results from the ETH event specifications, of which the patterns align closely with the all-event sample regressions. However, while the event study regressions reveal positive effects following ETH events, these effects are more subtle than those observed in the TWFE models with ETH events.

5.4.3 Discussion

The results of the regressions are consistent with the hypothesis that DeFi attacks positively impact the underlying native cryptoasset. An exploit event, although seemingly negative, can have a positive impact on the market capitalization of the native tokens of the affected blockchains. There are several reasons for this. First, the spillover effect of an event may cause the unaffected blockchain to be affected, resulting in negligible relative changes. Second, the news of such incidents may generate interest in researching the associated DeFi applications, which could stimulate demand for the blockchain. Third, in response to a DeFi attack or failure, the respective blockchain user communities and project developers can collaborate to address the issue and enhance the DeFi ecosystem's overall resilience by strengthening security measures.

In particular, the positive impact of ETH events is more pronounced. Ethererum is the most established and dominant blockchain in the DeFi space, accounting for over 56% of the total TVL (Binance Research, 2023). It hosts a wide range of applications and protocols, providing diversity and connectivity in the ecosystem. This makes it a preferred choice for many users and investors, with a large network of developers. For instance, multiple online platforms have been gathering points for Ethereum developers and network participants to discuss the trends and technical developments. Ethereum channel on Reddit [4] and Ethereum Magicians [5] are two of the largest channels for such discussions. Therefore, the network effect can play a critical role in the event of an attack, as any improvements made can have a more profound impact than those made on other blockchains.

5.5 Conclusion

Our study analyzes a large set of DeFi attacks and subsequent valuation effects on the underlying blockchains' native asset market valuation. To determine the impact of such events on market values, we conduct both TWFE and event study difference-in-difference regressions, which allow us to observe the average and day-by-day treatment effects, showing the differences between market values changes of treated and control blockchains in response to attacks on the DeFi protocols they host. We establish various control groups and event types to examine the degree of the impacts of the exploit events in different settings. Contrary to the intuition that exploits and protocol attacks would have negative consequences for the underlying blockchain, we consistently find that market values of affected blockchains increase rather than decrease relative to control group blockchains. Especially in the case of events on the Ethereum blockchain, we find amplified positive impacts resulting from DeFi exploit events. The findings suggest that the DeFi exploit events, particularly those that occur on established platforms

[4] https://www.reddit.com/r/ethereum/
[5] https://ethereum-magicians.org/

like Ethereum, signal improved security measures and platform audits. This leads to increased investor and participant confidence, resulting in an increase in market value.

This paper presents evidence in favor of treating DeFi protocol exploits as a temporary loss to the individual platform, but overall strengthening the resilience and suitability of the underlying blockchain to host such applications. It serves as a valuable foundation for future research in the area of DeFi attack and security, enabling researchers to delve deeper into the relationship between exploit incidents and the underlying blockchain infrastructure, as well as the attribution of their impact. Furthermore, the paper offers insight into how investors and the wider DeFi community perceive these exploits, highlighting implications for practitioners and regulators in terms of a nuanced understanding of how such incidents can influence market perceptions and the actions of various stakeholders.

FIGURE 5.3: DID event study regressions for all-event sample with estimation windows of 10 days and 15 days.

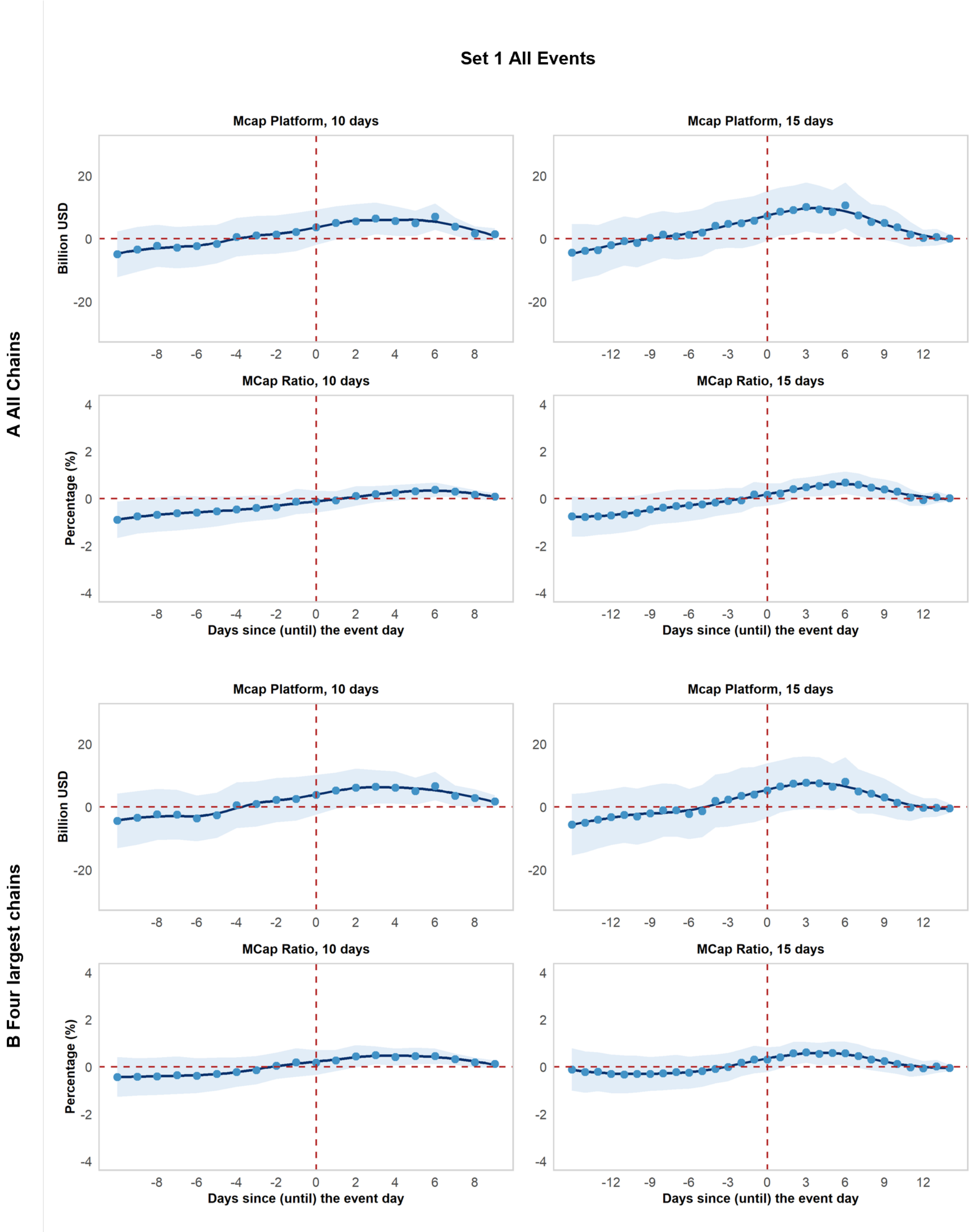


FIGURE 5.4: DID event study regressions for all-event sample with estimation windows of 10 days and 15 days.

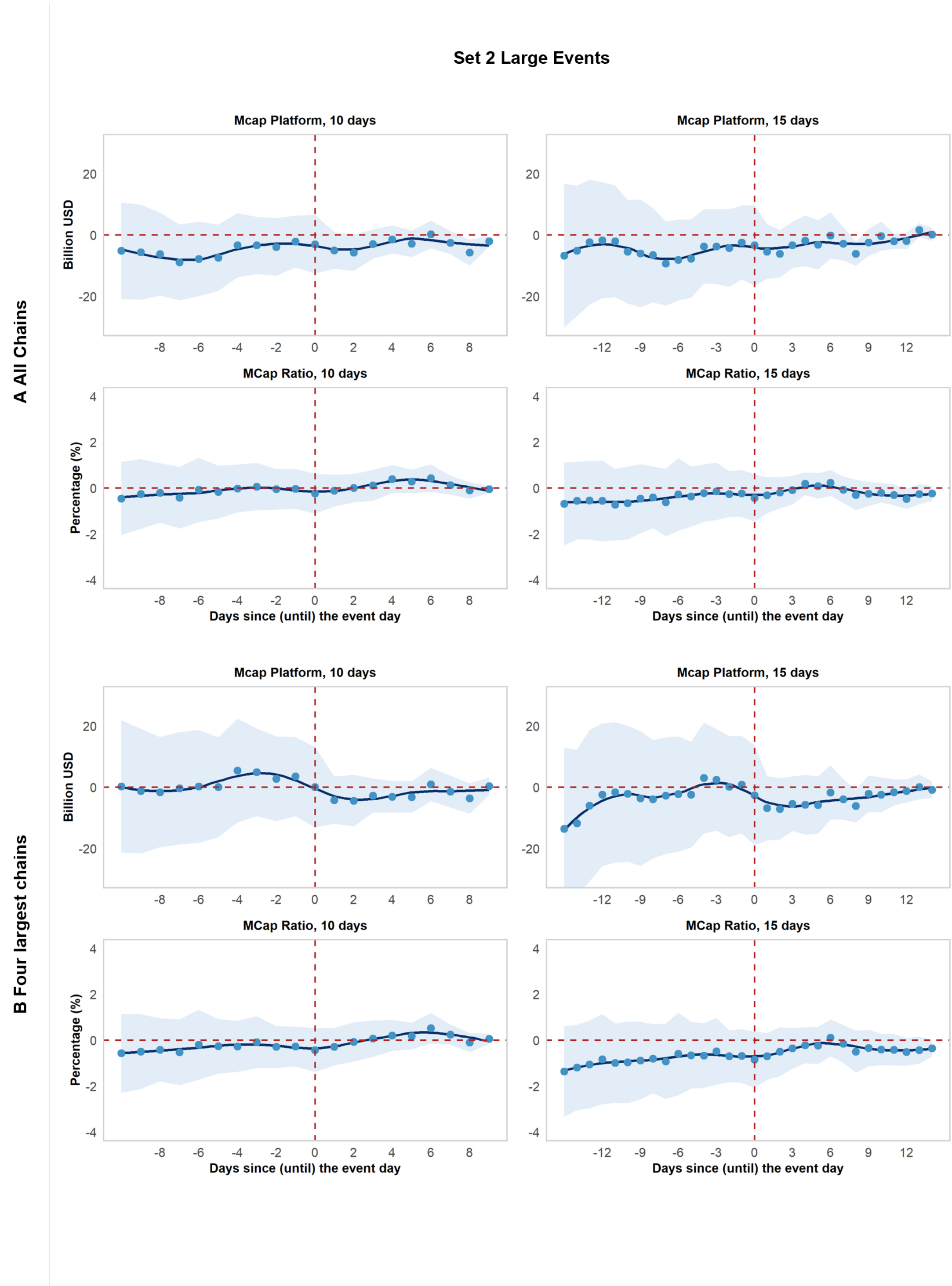
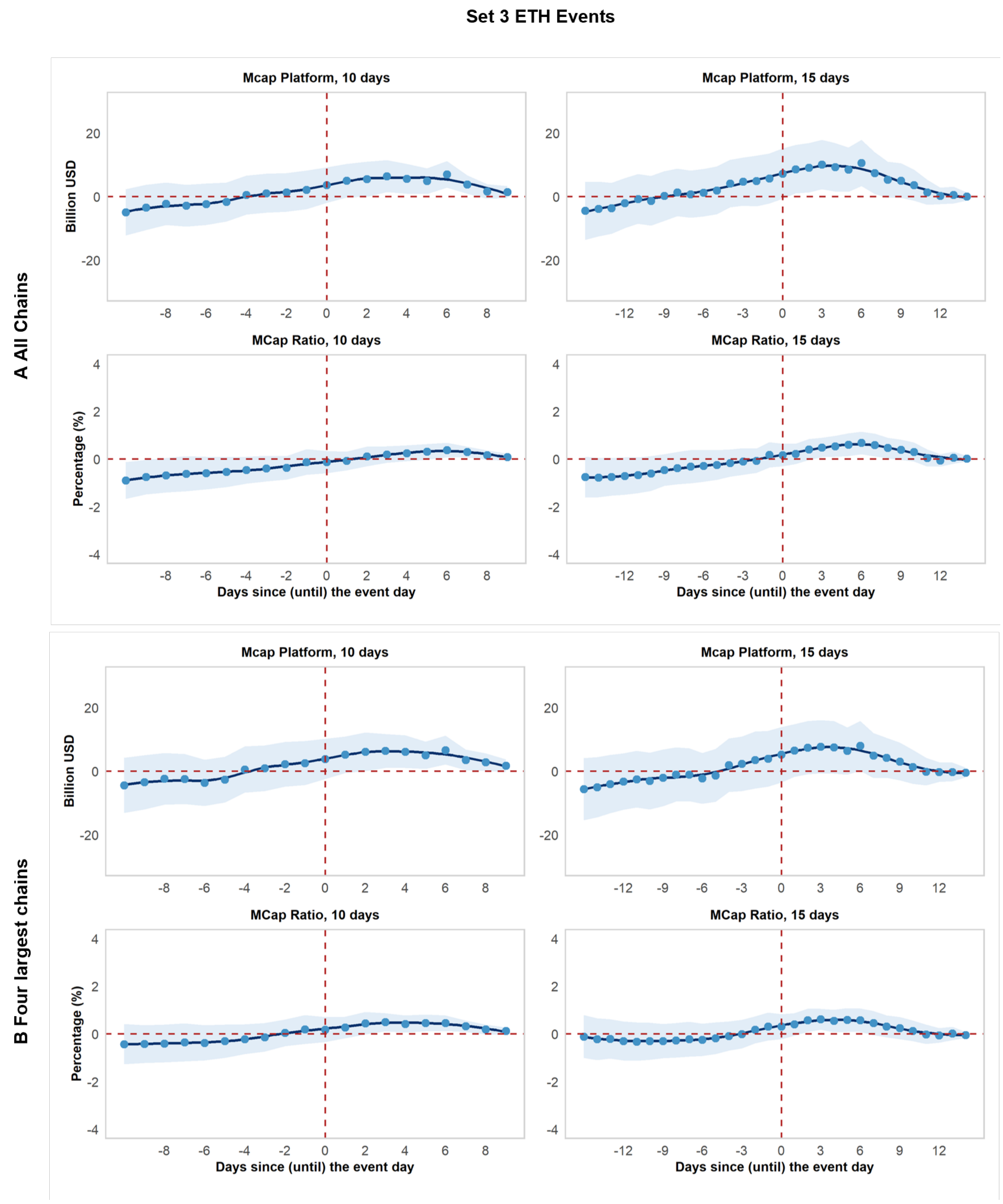


FIGURE 5.5: DID event study regressions for all-event sample with estimation windows of 10 days and 15 days.



Chapter 6

Concluding Remarks

6.1 Main Findings and Contributions

This dissertation comprises of four interconnected studies that examine blockchain business innovation from different perspectives. The first study is a literature review that explores and evaluates the computer-based text analysis techniques that could be used in blockchain related studies. It goes beyond presenting individual elements such as research scope, text data type, and methodology, and explores the connections between them. It highlights the importance of selecting the optimal combinations, depending on data characteristics and research questions. By integrating blockchain topics into text analysis, this study points out five key research areas that exist in the extant literature. Furthermore, it identifies three emerging research directions where more attention need to be drawn.

It builds up the foundation for the second study, which utilizes one of the unsupervised machine learning algorithms, LDA discovered in the first study to be ideal for studying emerging areas, to explore the blockchain innovation landscape through patent filings from EPO database. It gives an overview and evolution of the business landscape of blockchain innovation using business-relevant data, identifying the sub-topics in blockchain area and differentiating the innovators in their innovation approaches. Twenty topics, which are categorized into four groups, namely blockchain design, financial uses, data management, and physical goods, are identified. Furthermore, the study explores the discrepancies between academic papers and business applications and proposes future research areas. The insights drawn from this study would be of help for both researchers and practitioners to identify their research areas or establish businesses.

Utilizing the overview dynamics from the second study, two promising blockchain applications in the frame of finance area were selected for further examination. The third study introduces the application of blockchain in cross-border payment, with a focus on stablecoin. It introduces the mechanisms, identifying the adoption challenges, and provides solutions and business insights through case studies of three Africa-focused blockchain startups. The fourth study delve into the emerging DeFi world and examine the impacts of DeFi exploit events to the underlying blockchain platforms. It deploys a TWFE DiD model and event study DiD model to capture the pattern of the underlying cryptoassets of blockchain platforms pre- and post-events. The key finding is that when DeFi protocols on a blockchain experience attacks, the valuation of the native cryptoassets increases after the events, suggesting that the blockchain infrastructure views these incidents as positive developments.

This dissertation aims to deepen the existing understanding of fast-developing blockchain applications in business context, and explore the new arenas for future research. The contributions of this dissertation are threefold.

First, this dissertation provides insights into using novel text analysis methodologies for blockchain-related research topics. One of the primary challenges in emerging blockchain research is data availability. While numerical data is easily accessible on public blockchains, it remains challenging to obtain for private blockchains. By using text data for quantitative research, researchers can bridge the gap in data availability and approach research questions from various perspectives. This approach enables researchers to draw inferences that are not possible with numeric data alone. This dissertation provides researchers with a comprehensive overview of the possibilities and necessary considerations in implementing text analysis techniques in blockchain-related research and offers a novel and valuable perspective for researchers to investigate blockchain-related research questions.

Second, this dissertation examines blockchain innovation from a business perspective by illustrating the trends of blockchain innovation landscapes, identifying the different innovation focus of different countries as well as different types of innovators (i.e., startups and established companies), highlighting the discrepancy between literature and business usages. Due to the interdisciplinary nature of blockchain, research in blockchain-related areas tend to have a focus on one specific area, sometimes overlooking the significance of other perspectives, which limits the further development of this research area. This dissertation connects blockchain business-related studies and technical perspective (i.e., patent filings), uncovers the current innovation landscape of blockchain, identifies the key innovation areas and innovators, and underpins the discrepancies between the literature and real-world usage of blockchain technology. It facilitates to shape a more comprehensive and holistic review of the blockchain business ecosystem, enabling researchers and practitioners to better understand the current state of the field, and identify the paths to bring forward the future studies in this rapidly growing area.

Third, this dissertation dives into two specific finance areas in blockchain, namely cross-border payments and DeFi, to extend the understanding of the business usages. These areas are selected based on the research gaps identified in this dissertation - the blockchain research in the financial sectors have spread widely, yet, the literature has been focused on cryptocurrency. Given the existing applications and potential of blockchain usages in financial sectors, this dissertation provides insights of 1) using blockchain for cross-border payments, as a enabler to increase financial inclusion in developing countries, especially in African countries, exploring the social perspective of the technology application; 2) the impact of DeFi exploits on blockchain networks and identifying the factors that influence the value of underlying tokens, building foundation for DeFi security topics.

Overall, given the ever-evolving blockchain research areas, this dissertation closes the research gaps and contributes to the literature by providing novel text analysis techniques to mitigate the data availability problem, providing a dynamic blockchain innovation landscape, and exploring the emerging yet understudied blockchain usages in financial sector.

6.2 Limitations and Future Research

Despite the novelty and contributions of this dissertation, it is important to acknowledge that the research herein is not without its limitations. The limitations are mainly due to the nascence of the blockchain topic. This emerging field provides numerous research opportunities, but also poses some limitations and drawbacks for researchers. First, although this dissertation discusses and provides text analysis techniques to overcome the data availability issues, there are still data that we cannot access due to the immaturity of the market and the lack of disclosure mandate. Every effort has been made to incorporate as much available data as possible. Second, given the nascent nature of blockchain technology, it is expected that the technology will undergo rapid development in the future, so the insights provided by this dissertation will require regular updates.

Nevertheless, the limitations of this dissertation also provide avenues for future research directions. This dissertation provides a solid foundation for researchers to explore such areas of potential. First and foremost, the text analysis techniques outlined in this dissertation can be applied to a wide range of textual data to provide a more comprehensive understanding of economic and financial issues. With the increasing sophistication of machine learning algorithms, complicated and lengthy texts can be examined and analyzed to provide valuable insights. Additionally, as the dissertation concludes, young firms are the powerhouse of blockchain innovation, adding diversity to the ecosystem. More studies that explicitly focus on the impact of young firms should be conducted. Furthermore, given the development of blockchain, its security and related regulations have been a critical issue. The gaps between technology development and the introduction of relevant regulatory frameworks, as well as their aftermath, need to be studied. It could also provide vital insights for practitioners and policymakers in their approaches to building business and governing the market. Finally, the research frameworks are still under development, and the research field is constantly evolving. Like the attempt of this dissertation to advance the understanding of the blockchain ecosystem in business, future research has ample opportunities to explore the emerging application areas based on blockchain.

Appendix A

List of keywords for the query

The initial list of keywords with fundamental blockchain concepts based on our knowledge of the blockchain ecosystem (i.e., blockchain, cryptocurrency, smart contract, and ICO) and expand our list by sampling academic papers that include additional keywords. In this way, we build up a wider set of keywords by adding non-redundant keywords after observing keywords used in the academic literature. Our list of keywords is an intersection of keywords used in many blockchain-related papers. The complete list of keywords for the query is as follows:

Category	Search term	Keyword
Blockchain	blockchain*	blockchain
		blockchains
	cryptocurrenc*	cryptocurrency
		cryptocurrencies
	stablecoin*	stablecoin
		stablecoins
	“crypto token*"	crypto token
		crypto tokens
		crypto-token
		crypto-tokens
	“smart contract*"	smart contract
		smart contracts
	“initial coin offering*"	initial coin offering
		initial coin offerings
	“security token offering*"	security token offering
		security token offerings
	“initial exchange offering*"	initial exchange offering
		initial exchange offerings
	“non fungible token*"	non fungible token
		non fungible tokens
		non-fungible token
		non-fungible tokens
Text analysis	“text* analysis"	text analysis
		textual analysis
	“text analytics"	text analytics
	“topic model*"	topic model
		topic models
		topic modeling
		topic modelings
		topic modellings
	“natural language processing*"	natural language processing
		natural language processings
	“word embedding*"	word embedding
		word embeddings
	“sentence embedding*"	sentence embedding
		sentence embeddings
	“bag of words"	bag of words
		bag-of-words
	“sentiment analysis"	sentiment analysis

Appendix B

Original outputs of the LDA model and author-generated topic labels

The outputs of the model include a number of the topic order (indicates no importance level) and the relevant keywords that are associated with the topic. The label (i.e., name of the topic) is generated by the author. The original outputs of the LDA model, which includes ten most significant keywords for each topic, and the author-generated topic labels are presented as follows:

Topic Number	Author-generated Topic Labels	Outputs from LDA Models (Keywords)
1	Trading Platform	Trading, Member, Trade, Bitcoin, Fund, Purchase, Order, Receipt, Exchange, Buyer
2	Document Digitalization	Document, Output, Task, Input, Proof, Machine, Ticket, Model, Script, Mining
3	Secure Storage	Storage, File, Medium, Content, Store, Database, Computer, Present, Readable, Memory
4	Cloud Service	Service, Server, Client, Application, Network, Cloud, Compute, Access, Request, Configuration
5	Supply Chain Management	Product, Code, Trace, Physical, Commodity, Anticounterfeiting, Tag, Delivery, Circulation, Verification
6	Healthcare	Electronic, Record, Medical, Audit, Time, Evidence, Patient, Generate, Store, Event
7	Inter-organizational Data Management	Distribute, Associate, Ledger, Network, Resource, Plurality, Entity, Receive, Computer, Store
8	Consensue Mechanism	Consensus, Network, Time, Verification, Mechanism, Step, Vote, Operation, Main, Improve
9	Public Key Cryptograph	Key, Public, Private, Signature, Encryption, Encrypt, Generate, Sign, Message, Secret
10	Identity Verification	Authentication, Identity, Verification, Terminal, Authenticate, Server, Request, Personal, Registration, Trust
11	Transaction Processing	Transaction, Address, Network, Account, Request, Generate, Record, Verification, Party, Correspond
12	Cryptocurrency and Payment	Payment, Account, Token, Cryptocurrency, Financial, Request, Card, Transaction, Amount, Wallet
13	Certificate Management	Certificate, Server, Register, Issue, Public, Hash, Key, Step, Support, Specific
14	Data Processing	Request, Target, Processing, Process, Send, Correspond, Apparatus, Receive, Obtain, Present
15	Smart Contract	Contract, Smart, Intelligent, Execution, Execute, Insurance, Code, Call, Party, State
16	Network Infrastructure	Management, Platform, Technology, Share, Security, Credit, Present, Layer, Disclose, Application
17	Digital Asset Management	Digital, Asset, Currency, Virtual, Transfer, Object, Exchange, Money, Wallet, Address
18	Multimedia	Vehicle, Power, Source, Energy, Plurality, Video, Charge, Monitor, Stream, Distribute
19	IoT	Unit, Communication, Control, Network, Security, Connect, Terminal, Thing, Internet, Mobile
20	Hash Function	Hash, Number, Generate, Random, Domain, Root, Tree, Current, Correspond, Time

Bibliography

Abeyratne, S. A. and R. P. Monfared (2016). “Blockchain ready manufacturing supply chain using distributed ledger”. In: *International Journal of Research in Engineering and Technology* 5.9, pp. 1–10.

Abou Jaoude, J. and R. G. Saade (2019). “Blockchain applications–usage in different domains”. In: *IEEE Access* 7, pp. 45360–45381.

Abraham, J. et al. (2018). “Cryptocurrency price prediction using tweet volumes and sentiment analysis”. In: *SMU Data Science Review* 1.3.

Abu Daqar, M. A., S. Arqawi, and S. Abu Karsh (2021). “Fintech in the eyes of Millennials and Generation Z (the financial behavior and Fintech perception)”. In: *Banks and Bank Systems* 15, pp. 20–28.

Agbo, C. C. et al. (2019). “Blockchain technology in healthcare: a systematic review”. In: *Healthcare*. Vol. 7. 2. Multidisciplinary Digital Publishing Institute, p. 56.

Akba, F. et al. (2021). “Manipulator detection in cryptocurrency markets based on forecasting anomalies”. In: *IEEE ACCESS* 9, pp. 108819–108831.

Alahi, I. et al. (2019). “Identifying the challenges of the blockchain community from Stackexchange topics and trends”. In: *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*. Vol. 1, pp. 123–128. ISBN: 978-1-7281-2607-4.

Anamika, A. and S. Subramaniam (2022). “Do news headlines matter in the cryptocurrency market?” In: *Applied Economics*, pp. 1–17.

Antonopoulos, A. M. (2017). *Mastering Bitcoin: Programming the open blockchain*. " O’Reilly Media, Inc."

Arner, D., R. Auer, and J. Frost (2020). “Stablecoins: risks, potential and regulation”. In: *BIS Working Papers*.

Aslam, N. et al. (2022). “Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model”. In: *IEEE Access* 10, pp. 39313–39324.

Auer, R. and R. Böhme (2020). “The technology of retail central bank digital currency”. In: *BIS Quarterly Review, March*.

Ayvaz, S. and M. O. Shiha (2018). “A scalable streaming big data architecture for real-time sentiment analysis”. In: *Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing*, pp. 47–51.

Azqueta-Gavaldón, A. (2020). “Causal inference between cryptocurrency narratives and prices: Evidence from a complex dynamic ecosystem”. In: *Physica A: Statistical Mechanics and its Applications* 537, p. 122574.

Baccianella, S., A. Esuli, and F. Sebastiani (2010). “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining”. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Bahamazava, K. and R. Nanda (2022). “The shift of DarkNet illegal drug trade preferences in cryptocurrency: The question of traceability and deterrence”. In: *Forensic Science International: Digital Investigation* 40, p. 301377. ISSN: 2666-2817.

Bahamazava, K. and S. Reznik (2022). “The comparative analysis of regulations in the Italian Republic and the Russian Federation against cryptolaundering techniques”. In: *Journal of Money Laundering Control* ahead-of-print.ahead-of-print. ISSN: 1368-5201.

Baker, A. C., D. F. Larcker, and C. C. Wang (2022). “How much should we trust staggered difference-in-differences estimates?” In: *Journal of Financial Economics* 144(2), pp. 370–395.

Baldwin, C. (2016). “Bitcoin worth $72 million stolen from Bitfinex exchange in Hong Kong”. In: *Reuters Media*. URL: https://www.reuters.com/article/us-bitfinex-hacked-hongkong-idUSKCN10E0KP.

Balfagih, A. M. and V. Keselj (2019). “Evaluating Sentiment C1assifiers for Bitcoin Tweets in Price Prediction Task”. In: *2019 IEEE International Conference on Big Data (Big Data)*, pp. 5499–5506. ISBN: 978-1-7281-0858-2.

Bank of England (2020). “Central Bank Digital Currency—Opportunities, Challenges and Design”. In: *Discussion Paper, March*.

Bao, Y. and A. Datta (2014). “Simultaneously discovering and quantifying risk types from textual risk disclosures”. In: *Management Science* 60.6, pp. 1371–1391.

Barraza, B. (2019). “The Worth of Words: How Technical White Papers Influence ICO Blockchain Funding.” In: *MIS Quarterly Executive* 18.4.

Barth, J. R. et al. (2020). “Cryptocurrency valuation and ethics: A text analytic approach”. In: *Journal of Management Analytics* 7.3, pp. 367–388.

Basberg, B. L. (1987). “Patents and the measurement of technological change: a survey of the literature”. In: *Research policy* 16.2-4, pp. 131–141.

Bashchenko, O. (2022). “Bitcoin Price Factors: Natural Language Processing Approach”. In: *Available at SSRN 4079091*.

Baydakova, A. (2021). “Millions in Crypto Is Crossing the Russia-China Border Daily. There, Tether Is King”. In: *CoinDesk*. URL: https://www.coindesk.com/business/2019/07/30/millions-in-crypto-is-crossing-the-russia-china-border-daily-there-tether-is-king/.

Beck, T. and M Peria (2009). “What explains the cost of remittances?” In: *An examination across* 119.

Beinke, J. H., D. Nguyen, and F. Teuteberg (2018). “Towards a Business Model Taxonomy of Startups in the Finance Sector using Blockchain”. In: *International Conference on Interaction Sciences*.

Berentsen, A. and F. Schär (2019). “Stablecoins: The quest for a low-volatility cryptocurrency”. In: *The economics of Fintech and digital currencies*, pp. 65–75.

Binance Research (2023). *Are We Entering a Bull Market? Top 10 Narratives to Follow*. URL: https://public.bnbstatic.com/static/files/research/top-10-narratives.pdf.

Bird, S. et al. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O’Reilly Media, Inc."

Birim, Ş. Ö. and F. E. Sönmez (2022). “Social Sentiment Analysis for Prediction of Cryptocurrency Prices Using Neuro-Fuzzy Techniques”. In: *International Conference on Intelligent and Fuzzy Systems*, pp. 606–616.

Blei, D. M. and J. D. Lafferty (2006). “Dynamic topic models”. In: *Proceedings of the 23rd international conference on Machine learning*, pp. 113–120.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). “Latent dirichlet allocation”. In: *Journal of machine Learning research* 3.Jan, pp. 993–1022.

Blockworks (2022). *Global Stablecoin Adoption Points to Increased Dollarization*. URL: https://blockworks.co/news/global-stablecoin-adoption-points-to-increased-dollarization.

Bloomberg (2023). *Ethereum Developers Push Ahead With Shanghai Upgrade to Enable Withdrawals*. URL: https://www.bloomberg.com/news/articles/2023-01-05/ethereum-developers-push-ahead-with-update-enabling-withdrawals.

Bocek, T. et al. (2017). "Blockchains everywhere-a use-case of blockchains in the pharma supply-chain". In: *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, pp. 772–777.

Bojanowski, P. et al. (2017). "Enriching word vectors with subword information". In: *Transactions of the association for computational linguistics* 5, pp. 135–146.

Bolt, W., V. Lubbersen, and P. Wierts (2022). "Getting the balance right: Crypto, stablecoin and central bank digital currency". In: *Journal of Payments Strategy & Systems* 16.1, pp. 39–50.

Bondi, A. B. (2000). "Characteristics of scalability and their impact on performance". In: *Proceedings of the 2nd international workshop on Software and performance*, pp. 195–203.

Boser, B. E., I. M. Guyon, and V. N. Vapnik (1992). "A training algorithm for optimal margin classifiers". In: *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152.

Bracciali, A., D. Grossi, and R. de Haan (2021). "Decentralization in open quorum systems: Limitative results for ripple and stellar". In: *2nd International Conference on Blockchain Economics, Security and Protocols (Tokenomics 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Bullmann, D., J. Klemm, and A. Pinna (2019). "In search for stability in crypto-assets: are stablecoins the solution?" In: *ECB Occasional Paper* 230.

Burnie, A. and E. Yilmaz (2019). "Social media and Bitcoin metrics: Which words matter". In: *Royal Society open science* 6.10, p. 191068.

Buterin, V. (2016). "Chain interoperability". In: *R3 Research Paper*.

Buterin, V. et al. (2014). "A next-generation smart contract and decentralized application platform". In: *white paper* 3.37.

Caliskan, K. (2020). "Platform works as stack economization: Cryptocurrency markets and exchanges in perspective". In: *Sociologica* 14.3, pp. 115–142.

Cao, S. et al. (2019). “Financial Reporting and Blockchains: Audit Pricing, Misstatements, and Regulation”. In: *Misstatements, and Regulation (June 2019)*.

Cary, M. (2021). “Down with the #dogefather: Evidence of a cryptocurrency responding in real time to a crypto-tastemaker”. In: *Journal of Theoretical and Applied Electronic Commerce Research* 16.6, pp. 2230–2240.

Casino, F., T. K. Dasaklis, and C. Patsakis (2019). “A systematic literature review of blockchain-based applications: Current status, classification and open issues”. In: *Telematics and Informatics* 36, pp. 55–81.

Catalini, C. and J. S. Gans (2016). *Some simple economics of the blockchain*. Tech. rep. National Bureau of Economic Research.

Catalini, C. and A. de Gortari (2021). “On the Economic Design of Stablecoins”. In: *Available at SSRN 3899499*.

Catalini, C., A. de Gortari, and N. Shah (2022). “Some Simple Economics of Stablecoins”. In: *Annual Review of Financial Economics* 14, pp. 117–135.

Catalini, C. and N. Shah (2021). “Setting Standards for Stablecoin Reserves”. In: *Available at SSRN 3970885*.

Catalini, C. et al. (2021). “From Stablecoins to CBDCs: The Public Benefits of a Public-Private Partnership”. In: *Available at SSRN 3986192*.

Caviggioli, F. et al. (2020). “Technology adoption news and corporate reputation: sentiment analysis about the introduction of Bitcoin”. In: *Journal of Product & Brand Management* 29.7, pp. 877–897.

Cavnar, W. B., J. M. Trenkle, et al. (1994). “N-gram-based text categorization”. In: *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval*. Vol. 161175.

Chakkarwar, V. and S. C. Tamane (2019). “Quick Insight of Research Literature Using Topic Modeling”. In: *Smart Trends in Computing and Communications 2019*. Ed. by Y.-D. Zhang et al. Vol. 165, pp. 189–197.

Chang, V. et al. (2020). “How Blockchain can impact financial services–The overview, challenges and recommendations from expert interviewees”. In: *Technological Forecasting and Social Change* 158, p. 120166.

Chen, C. Y.-H. et al. (2019a). *What makes cryptocurrencies special? Investor sentiment and return predictability during the bubble*. Tech. rep. IRTG 1792 Discussion Paper.

Chen, M. A. et al. (2019b). "How valuable is FinTech innovation?" In: *The Review of Financial Studies* 32.5, pp. 2062–2106.

Chen, M. A. et al. (2022a). "Can Blockchain Technology Help Overcome Contractual Incompleteness? Evidence from State Laws". In: *Available at SSRN 3915895*.

Chen, T. and C. Guestrin (2016). "XGBoost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Chen, Y. et al. (2022b). "A survey on blockchain systems: Attacks, defenses, and privacy preservation". In: *High-Confidence Computing* 2.2, p. 100048.

Cheuque Cerda, G. and J. L. Reutter (2019). "Bitcoin price prediction through opinion mining". In: *Companion Proceedings of The 2019 World Wide Web Conference*, pp. 755–762.

Chiarello, F. et al. (2021). "Value creation in emerging technologies through text mining: The case of blockchain". In: *Technology Analysis & Strategic Management* 33.12, pp. 1404–1420.

Chiu, J. and T. V. Koeppl (2019). "Blockchain-based settlement for asset trading". In: *The Review of Financial Studies* 32.5, pp. 1716–1753.

Choi, J., M. A. Dutz, and Z. Usman (2020). *The Future of Work in Africa: Harnessing the Potential of Digital Technologies for All*. Ed. by J. Choi, M. A. Dutz, and Z. Usman. The World Bank.

Choi, J. et al. (2022). "Discovering Message Templates on Large Scale Bitcoin Abuse Reports Using a Two-Fold NLP-based Clustering Method". In: *IEICE TRANSACTIONS on Information and Systems* 105.4, pp. 824–827.

Chokor, A. and E. Alfieri (2021). "Long and short-term impacts of regulation in the cryptocurrency market". In: *The Quarterly Review of Economics and Finance* 81, pp. 157–173.

Chousein, Z. et al. (2020). "Tension between GDPR and Public Blockchains: A Data-Driven Analysis of Online Discussions". In: *13th International Conference on Security of Information and Networks*, pp. 1–8.

Christie, W. G. and R. D. Huang (1995). "Following the pied piper: Do individual returns herd around the market?" In: *Financial Analysts Journal* 51.4, pp. 31–37.

Chursook, A. et al. (2022). “Twitter Sentiment Analysis and Expert Ratings of Initial Coin Offering Fundraising: Evidence from Australia and Singapore Markets”. In: *TEM Journal* 11.1, pp. 44–55.

Clack, C. D. et al. (2016). “Smart contract templates: foundations, design landscape and research directions”. In: *arXiv preprint arXiv:1608.00771*.

Clarke, N. S. et al. (2020). “Blockchain patent landscaping: An expert based methodology and search query”. In: *World Patent Information* 61, p. 101964.

Cohen, B. J. (2015). *Currency Power*. Princeton University Press.

Cohney, S. et al. (2019). “Coin-operated capitalism”. In: *Columbia Law Review* 119.3, pp. 591–676.

Coleman, M. and T. L. Liau (1975). “A computer readability formula designed for machine scoring.” In: *Journal of Applied Psychology* 60.2, p. 283.

Cong, L. W. and Z. He (2019). “Blockchain disruption and smart contracts”. In: *The Review of Financial Studies* 32.5, pp. 1754–1797.

Cong, L. W. et al. (2021). “Analyzing textual information at scale”. In: *Information for Efficient Decision Making: Big Data, Blockchain and Relevance*. Singapore: World Scientific, pp. 239–271.

Conley, J. P. et al. (2017). “Blockchain and the economics of crypto-tokens and initial coin offerings”. In: *Vanderbilt University Department of economics working papers* 17-00008.

Corbet, S. et al. (2018). “Exploring the dynamic relationships between cryptocurrencies and other financial assets”. In: *Economics Letters* 165, pp. 28–34.

CPMI and IOSCO (2022). *Application of the Principles for Financial Market Infrastructures to stablecoin arrangements*. URL: https://www.bis.org/cpmi/publ/d206.htm.

Critien, J. V., A. Gatt, and J. Ellul (2022). “Bitcoin price change and trend prediction through Twitter sentiment and data volume”. In: *Financial Innovation* 8.1, pp. 1–20.

Cunliffe, J. (2020). “Cross-border payment systems have been neglected for too long”. In: *Financial Times*. URL: https://www.ft.com/content/a241d7e0-e1de-4812-b214-b350cbb7d046.

Dai, H.-N. et al. (2019). “Blockchain for Internet of Things: A survey”. In: *IEEE Internet of Things Journal* 6.5, pp. 8076–8094.

Daian, P. et al. (2020). "Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability". In: *IEEE Symposium on Security and Privacy (SP)*, pp. 910–927.

Daim, T. U. et al. (2006). "Forecasting emerging technologies: Use of bibliometrics and patent analysis". In: *Technological Forecasting and Social Change* 73.8, pp. 981–1012.

Dale, E. and J. S. Chall (1948). "A formula for predicting readability: Instructions". In: *Educational Research Bulletin*, pp. 37–54.

Daluwathumullagamage, D. J. and A. Sims (2021). "Fantastic Beasts: Blockchain Based Banking". In: *Journal of Risk and Financial Management* 14.4.

Daluwathumullagamage, D. J. and A. Sims (2020). "Blockchain-enabled corporate governance and regulation". In: *International Journal of Financial Studies* 8.2.

Daniel, F. R., M. S. Conor, and S. Janina (2021). *Supporting Small and Medium Enterprises in Sub-Saharan Africa through Blended Finance*. URL: https://www.csis.org/analysis/supporting-small-and-medium-enterprises-sub-saharan-africa-through-blended-finance.

Deloitte (2021). *So, You Want to Be a Stablecoin Issuer?* URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/regulatory/so-you-want-to-be-a-stablecoin-issuer-2021.pdf.

Devlin, J. et al. (2018). "BERT: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805*.

Dewey, J. N. and R Plasencia (2018). "Blockchain and 5G-enabled Internet of Things (IoT) will redefine supply chains and trade finance". In: *Proc. Secured Lender*, pp. 43–45.

Di Prisco, D. and D. Strangio (2021). "Technology and financial inclusion: a case study to evaluate potential and limitations of Blockchain in emerging countries". In: *Technology Analysis & Strategic Management*, pp. 1–14.

Dinh, T. T. A. et al. (2018). "Untangling blockchain: A data processing view of blockchain systems". In: *IEEE Transactions on Knowledge and Data Engineering* 30.7, pp. 1366–1385.

Dittmar, R. and D. A. Wu (2019). "Initial Coin Offerings Hyped and Dehyped: An Empirical Examination". In: *Available at SSRN 3259182*.

Donovan, K. (2012). "Mobile money for financial inclusion". In: *Information and Communications for development* 61.1, pp. 61–73.

Dumay, J. and L. Cai (2014). "A review and critique of content analysis as a methodology for inquiring into IC disclosure". In: *Journal of Intellectual Capital* 15.2, pp. 264–290.

Dyer, T. et al. (2017). "The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation". In: *Journal of Accounting and Economics* 64.2-3, pp. 221–245.

El-Masri, M. and E. M. A. Hussain (2021). "Blockchain as a mean to secure Internet of Things ecosystems – A systematic literature review". In: *Journal of Enterprise Information Management* 34.5, pp. 1371–1405.

Ester, M. et al. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise." In: *KDD–96*. Vol. 96. 34, pp. 226–231.

Fang, F. et al. (2022). "Cryptocurrency trading: a comprehensive survey". In: *Financial Innovation* 8.1, pp. 1–59.

Farimani, S. A. et al. (2022). "Investigating the informativeness of technical indicators and news sentiment in financial market price prediction". In: *Knowledge-Based Systems* 247, p. 108742.

Farre-Mensa, J. et al. (2020). "What is a patent worth? Evidence from the US patent "lottery"". In: *The Journal of Finance* 75.2, pp. 639–682.

Fernández-Caramés, T. M. and P. Fraga-Lamas (2018). "A Review on the Use of Blockchain for the Internet of Things". In: *Ieee Access* 6, pp. 32979–33001.

Feyen, E. et al. (2021). "What does digital money mean for emerging market and developing economies?" In: *The Palgrave Handbook of Technological Finance*. Springer, pp. 217–241.

Fiedler, M. and P. Sandner (2017). "Identifying leading blockchain startups on a worldwide level". In: *Frankfurt School Blockchain Center*.

Financial Times (2023). "Venture capital not done with crypto yet". In: *Financial Times*. URL: https://www.ft.com/content/14428c6e-63f8-4868-a85f-2d20f2c7973e.

Finck, M. (2019). *Blockchain and the general data protection regulation: Can distributed ledgers be squared with European data protection law?* European Union.

Flesch, R. (1979). *How to write plain English: A book for lawyers and consumers (1st ed.)* New York: Harper & Row.

Florysiak, D. and A. Schandlbauer (2022). "Experts or charlatans? ICO analysts and white paper informativeness". In: *Journal of Banking and Finance* 139.

Foley, S., J. R. Karlsen, and T. J. Putniņš (2019). "Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies?" In: *The Review of Financial Studies* 32.5, pp. 1798–1853.

Friedlmaier, M. et al. (2018). "Disrupting industries with blockchain: The industry, venture capital funding, and regional distribution of blockchain ventures". In: *Venture Capital Funding, and Regional Distribution of Blockchain Ventures (September 22, 2017). Proceedings of the 51st Annual Hawaii International Conference on System Sciences (HICSS)*.

Friedman, J. H. (2001). "Greedy function approximation: A gradient boosting machine". In: *Annals of statistics* 29.5, pp. 1189–1232.

Frizzo-Barker, J. et al. (2020). "Blockchain as a disruptive technology for business: A systematic review". In: *International Journal of Information Management* 51, p. 102029.

FSB (2020). *Addressing the regulatory, supervisory and oversight challenges raised by "global stablecoin" arrangements*.

Fu, C., A. Koh, and P. Griffin (2019). "Automated Theme Search in ICO Whitepapers". In: *Journal of Financial Data Science* 1.4, pp. 140–158.

FXStreet (2023). *Ethereum Shanghai Upgrade: Guide to the ETH hard fork, unstaking and liquid staking projects*.

Galeshchuk, S., O. Vasylchyshyn, and A. Krysovatyy (2018). "Bitcoin response to Twitter sentiments". In: *CEUR Workshop Proceedings*, pp. 160–168.

Gallagher, R. J. et al. (2017). "Anchored correlation explanation: Topic modeling with minimal domain knowledge". In: *Transactions of the Association for Computational Linguistics* 5, pp. 529–542.

Garanina, T., M. Ranta, and J. Dumay (2021). "Blockchain in accounting research: Current trends and emerging topics". In: *Accounting, Auditing and Accountability Journal* 35.7, pp. 1507–1533.

Ge, C. et al. (2021). "Investigating the Demand for Blockchain Talents in the Recruitment Market: Evidence from Topic Modeling Analysis on Job Postings". In: *Information and Management*.

Gentzkow, M., B. Kelly, and M. Taddy (2019). "Text as data". In: *Journal of Economic Literature* 57.3, pp. 535–74.

Georgoula, I. et al. (2015). "Using time-series and sentiment analysis to detect the determinants of Bitcoin prices". In: *Available at SSRN 2607167*.

Goldberg, Y. (2017). "Neural network methods for natural language processing". In: *Synthesis lectures on human language technologies* 10.1, pp. 1–309.

Goldstein, R. A. M. and H. Tabuchi (2014). "Erosion of Faith Was Death Knell for Mt. Gox". In: *NY Times*. URL: https://dealbook.nytimes.com/2014/02/28/mt-gox-files-for-bankruptcy/.

Gordon, W. J. and C. Catalini (2018). "Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability". In: *Computational and structural biotechnology journal* 16, pp. 224–230.

Grassman, R. et al. (2021). "Attitudes to Cryptocurrencies: A Comparative Study Between Sweden and Japan". In: *The Review of Socionetwork Strategies* 15.1, pp. 169–194.

Grimmer, J. and B. M. Stewart (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". In: *Political Analysis* 21.3, pp. 267–297.

Grover, P. et al. (2019). "Perceived usefulness, ease of use and user acceptance of blockchain technology for digital transactions – insights from user-generated content on Twitter". In: *Enterprise Information Systems* 13.6, pp. 771–800.

Gunning, R. (1952). *Technique of clear writing*. McGraw-Hill.

Günther, E. and T. Quandt (2016). "Word Counts and Topic Models". In: *Digital Journalism* 4.1, pp. 75–88.

Gurdgiev, C. and D. O'Loughlin (2020). "Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty". In: *Journal of Behavioral and Experimental Finance* 25, p. 100271.

Han, S., S. Ye, and H. Zhang (2020). "Visual exploration of Internet news via sentiment score and topic models". In: *Computational Visual Media* 6.3, pp. 333–347.

Hanley, K. W. and G. Hoberg (2019). "Dynamic interpretation of emerging risks in the financial sector". In: *The Review of Financial Studies* 32.12, pp. 4543–4603.

Hansen, S. et al. (2018). "Transparency and deliberation within the FOMC: a computational linguistics approach". In: *The Quarterly Journal of Economics* 133.2, pp. 801–870.

Hashimoto, K. et al. (2016). "A joint many-task model: Growing a neural network for multiple NLP tasks". In: *arXiv preprint arXiv:1611.01587*.

Hassan, M. K., F. A. Hudaefi, and R. E. Caraka (2021). "Mining netizen's opinion on cryptocurrency: Sentiment analysis of Twitter data". In: *Studies in Economics and Finance* 39.3, pp. 365–385.

Hastie, T. et al. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Henry, E. (2008). "Are investors influenced by how earnings press releases are written?" In: *The Journal of Business Communication (1973)* 45.4, pp. 363–407.

Hinds-Charles, C. et al. (2019). "A Longitude Analysis on Bitcoin Issue Repository". In: *2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN)*, pp. 212–217. ISBN: 978-1-5386-4870-4.

Hirata, E., M. Lambrou, and D. Watanabe (2021). "Blockchain technology in supply chain management: Insights from machine learning algorithms". In: *Maritime Business Review* 6.2, pp. 114–128.

Ho, T. K. (1995). "Random decision forests". In: *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1, pp. 278–282.

Hoberg, G. and G. Phillips (2016). "Text-based network industries and endogenous product differentiation". In: *Journal of Political Economy* 124.5, pp. 1423–1465.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Hornuf, L. et al. (2023). "Cybercrime on the Ethereum Blockchain". In: *Available at SSRN 4527415*.

Hou, H. (2017). "The application of blockchain technology in E-government in China". In: *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, pp. 1–4.

Huang, A. H. et al. (2018). "Analyst information discovery and interpretation roles: A topic modeling approach". In: *Management Science* 64.6, pp. 2833–2855.

Huang, X. et al. (2021). "LSTM based sentiment analysis for cryptocurrency prediction". In: *International Conference on Database Systems for Advanced Applications*, pp. 617–621.

Huberman, G. et al. (2017). "Monopoly without a monopolist: An economic analysis of the bitcoin payment system". In: *Bank of Finland Research Discussion Paper* 27.

Hutto, C. and E. Gilbert (2014). "VADER: A parsimonious rule-based model for sentiment analysis of social media text". In: *Proceedings of the international AAAI conference on web and social media*. Vol. 8. 1, pp. 216–225.

Huynh, T. L. D. (2021). "Does Bitcoin react to Trump's tweets?" In: *Journal of Behavioral and Experimental Finance* 31, p. 100546.

— (2022). "When Elon Musk Changes his Tone, Does Bitcoin Adjust Its Tune?" In: *Computational Economics* 62, 639–661.

Ibba, G., M. Ortu, and R. Tonelli (2021). "Smart contracts categorization with topic modeling techniques". In: *CEUR Workshop Proceedings*. Ed. by B. Marin et al. Vol. 3031, pp. 64–73.

IMF (2009). *International Transactions in Remittances: Guide for Compilers and Users (RCG).* International Monetary Fund.

Inamdar, A. et al. (2019). "Predicting cryptocurrency value using sentiment analysis". In: *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pp. 932–934.

Irresberger, F. et al. (2023). "The public blockchain ecosystem: An empirical analysis". In: *NYU Stern School of Business*.

Jaccard, P. (1912). "The distribution of the flora in the Alpine zone." In: *New phytologist* 11.2, pp. 37–50.

Jain, A. et al. (2018). "Forecasting price of cryptocurrencies using tweets sentiment analysis". In: *2018 eleventh international conference on contemporary computing (IC3)*, pp. 1–7.

Ji, Q. et al. (2019). "Dynamic connectedness and integration in cryptocurrency markets". In: *International Review of Financial Analysis* 63, pp. 257–272.

John, K., L. Kogan, and F. Saleh (2023). "Smart Contracts and Decentralized Finance". In: *Annual Review of Financial Economics* 15, pp. 523–542.

Jurafsky, D. and J Martin (2017). "Naive bayes and sentiment classification". In: *Speech and language processing*, pp. 74–91.

Kang, K., J. Choo, and Y. Kim (2020). "Whose Opinion Matters? Analyzing Relationships Between Bitcoin Prices and User Groups in Online Community". In: *Social Science Computer Review* 38.6, pp. 686–702.

Karalevicius, V., N. Degrande, and J. de Weerdt (2018). “Using sentiment analysis to predict interday Bitcoin price movements”. In: *The Journal of Risk Finance* 19.1, pp. 56–75.

Khan, N., T. Ahmad, et al. (2019). “Feasibility of Stellar as a Blockchain-Based Micropayment System”. In: *International Conference on Smart Blockchain*, pp. 53–65.

Kilimci, Z. H. (2020). “Sentiment analysis based direction prediction in Bitcoin using deep learning algorithms and word embedding models”. In: *International Journal of Intelligent Systems and Applications in Engineering* 8.2, pp. 60–65.

Kim, S. et al. (2018). “A survey of scalability solutions on blockchain”. In: *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, pp. 1204–1207.

Kim, S., H. Park, and J. Lee (2020). “Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis”. In: *Expert Systems with Applications* 152, p. 113401.

Kim, Y. B. et al. (2016). “Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies”. In: *PloS one* 11.8, e0161197.

Klapper, L., A. Lusardi, and P. Van Oudheusden (2015). “Financial literacy around the world”. In: *World Bank. Washington DC: World Bank* 2, pp. 218–237.

Klare, G. R. (1974). “Assessing readability”. In: *Reading research quarterly*, pp. 62–102.

Kou, G. et al. (2020). “Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods”. In: *Applied Soft Computing* 86, p. 105836.

Kraaijeveld, O. and J. de Smedt (2020). “The predictive power of public Twitter sentiment for forecasting cryptocurrency prices”. In: *Journal of International Financial Markets, Institutions and Money* 65.

Kshetri, N. (2017). “Potential roles of blockchain in fighting poverty and reducing financial exclusion in the global south”. In: *Journal of Global Information Technology Management* 20.4, pp. 201–204.

— (2018). “1 Blockchain’s roles in meeting key supply chain management objectives”. In: *International Journal of Information Management* 39, pp. 80–89.

Kwon, O.-W. and J.-H. Lee (2003). "Text categorization based on K-nearest neighbor approach for web site classification". In: *Information Processing & Management* 39.1, pp. 25–44.

Labazova, O. et al. (2019). "From hype to reality: A taxonomy of blockchain applications". In: *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS 2019)*.

Lambert, T., D. Liebau, and P. Roosenboom (2021). "Security token offerings". In: *Small Business Economics*, pp. 1–27.

Larios-Hernández, G. J. (2017). "Blockchain entrepreneurship opportunity in the practices of the unbanked". In: *Business Horizons* 60.6, pp. 865–874.

Laturnus, V. (2023). "Optimal Disclosure Strategies of Soft Information: Evidence from Token Offerings". In: *Available at SSRN 4087795*.

Laurent, P. et al. (2018). "The tokenization of assets is disrupting the financial industry. Are you ready?" In: *Inside. Triannual insights from Deloitte* 19, pp. 62–67.

Le, Q. and T. Mikolov (2014). "Distributed representations of sentences and documents". In: *International conference on machine learning*, pp. 1188–1196.

Lee, D. and H. S. Seung (2000). "Algorithms for non-negative matrix factorization". In: *Advances in neural information processing systems* 13.

Lee, D. D. and H. S. Seung (1999). "Learning the parts of objects by non-negative matrix factorization". In: *Nature* 401.6755, pp. 788–791.

Lee, J., H. J. Zo, and T. Steinberger (2022). "Exploring Trends in Blockchain Publications with Topic Modeling: Implications for Forecasting the Emergence of Industry Applications". In: *Available at SSRN 4079332*.

Li, T. R. et al. (2019). "Sentiment-Based Prediction of Alternative Cryptocurrency Price Fluctuations Using Gradient Boosting Tree Model". In: *Frontiers in Physics* 7.

Li, W. et al. (2022). "Security Analysis of DeFi: Vulnerabilities, Attacks and Advances". In: *2022 IEEE International Conference on Blockchain (Blockchain)*, pp. 488–493.

Liberati, A. et al. (2009). "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration". In: *Annals of internal medicine* 151.4, W–65.

Lin, C. and Y. He (2009). "Joint sentiment/topic model for sentiment analysis". In: *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 375–384.

Linton, M. et al. (2017). "Dynamic Topic Modelling for Cryptocurrency Community Forums". In: *Applied Quantitative Finance*. Ed. by W. K. Härdle, C. Y.-H. Chen, and L. Overbeck. Berlin: Springer, pp. 355–372. ISBN: 978-3-662-54485-3.

Liu, Y., J. Sheng, and W. Wang (2021). "Technology and cryptocurrency valuation: Evidence from machine learning". In: *Available at SSRN 3577208*.

Loginova, E. et al. (2021). "Forecasting directional Bitcoin price returns using aspect-based sentiment analysis on online text data". In: *Machine Learning*, pp. 1–24.

Lokhava, M. et al. (2019). "Fast and secure global payments with Stellar". In: *Proceedings of the 27th ACM Symposium on Operating Systems Principles*, pp. 80–96.

Loughran, T. I. and B. McDonald (2016). "Textual Analysis in Accounting and Finance: A Survey". In: *Journal of Accounting Research* 54.4, pp. 1187–1230.

Loughran, T. and B. McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". In: *The Journal of finance* 66.1, pp. 35–65.

Lu, H.-K. et al. (2017). "A study on adoption of Bitcoin in Taiwan: Using big data analysis of social media". In: *Proceedings of the 3rd International Conference on Communication and Information Processing*, pp. 32–38.

MacQueen, J (1967). "Classification and analysis of multivariate observations". In: *5th Berkeley Symp. Math. Statist. Probability*, pp. 281–297.

Maesa, D. D. F. and P. Mori (2020). "Blockchain 3.0 applications survey". In: *Journal of Parallel and Distributed Computing* 138, pp. 99–114.

Mai, F. et al. (2018). "How does social media impact Bitcoin value? A test of the silent majority hypothesis". In: *Journal of management information systems* 35.1, pp. 19–52.

Makina, D. (2019). "The Potential of FinTech in Enabling Financial Inclusion". In: *Extending Financial Inclusion in Africa*. Ed. by M. Daniel. Elsevier, pp. 299–318. ISBN: 9780128141649.

Mao, H., S. Counts, and J. Bollen (2011). "Predicting financial markets: Comparing survey, news, Twitter and search engine data". In: *arXiv preprint arXiv:1112.1051*.

Mareddy, S. and D. Gupta (2022). "Analysis of Twitter Data for Identifying Trending Domains in Blockchain Technology". In: *Computer Networks and Inventive Communication Technologies*. Ed. by S. Smys et al. Vol. 75, pp. 651–672. ISBN: 978-981-16-3728-5.

Massacci, F., C.-N. Ngo, and J. M. Williams (2016). "Decentralized Financial Intermediation Beyond Blockchains". In: *Available at SSRN 2794913*.

McGhin, T. et al. (2019). "Blockchain in healthcare applications: Research challenges and opportunities". In: *Journal of Network and Computer Applications* 135, pp. 62–75.

McLaughlin, G. H. (1969). "SMOG grading - A new readability formula". In: *Journal of reading* 12.8, pp. 639–646.

Medhi, P. K. (2020). "Blockchain-enabled supply chain transparency, supply chain structural dynamics, and sustainability of complex global supply chains - a text mining analysis". In: *Information For Efficient Decision Making: Big Data, Blockchain And Relevance*. Singapore: World Scientific Publishing Co, pp. 273–312.

Medina, L. and F. Schneider (2017). *Shadow economies around the world: New results for 158 countries over 1991-2015*. Tech. rep. CESifo Working Paper Series.

Mehar, M. I. et al. (2019). "Understanding a revolutionary and flawed grand experiment in blockchain: the DAO attack". In: *Journal of Cases on Information Technology (JCIT)* 21.1, pp. 19–32.

Mendoza-Tello, J. C. et al. (2018). "Social commerce as a driver to enhance trust and intention to use cryptocurrencies for electronic payments". In: *IEEE Access* 6, pp. 50737–50751.

Merchant Machine (2021). *Share of unbanked population worldwide 2021*. Tech. rep. Merchant Machine. URL: https://www.statista.com/statistics/1246963/unbanked-population-in-selected-countries/.

Mikolov, T. et al. (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Mimno, D. et al. (2011). "Optimizing semantic coherence in topic models". In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 262–272.

Mita, M. et al. (2019). "What is stablecoin?: A survey on price stabilization mechanisms for decentralized payment systems". In: *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*. IEEE, pp. 60–66.

Mnif, E., I. Lacombe, and A. Jarboui (2021). "Users' perception toward Bitcoin Green with big data analytics". In: *Society and Business Review* 16.4, pp. 592–615.

Mohammad, S. M. (2018). "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words". In: *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*.

Moher, D. et al. (2009). "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement". In: *Annals of internal medicine* 151.4, pp. 264–269.

Moin, A., K. Sekniqi, and E. G. Sirer (2020). "SoK: A classification framework for stablecoin designs". In: *International Conference on Financial Cryptography and Data Security*. Springer, pp. 174–197.

Mougayar, W. (2016). *The business blockchain: promise, practice, and application of the next Internet technology*. John Wiley & Sons.

Mousa, A. and B. Schuller (2017). "Contextual Bidirectional Long Short-Term Memory Recurrent Neural Network Language Models: A Generative Approach to Sentiment Analysis". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Valencia, Spain: Association for Computational Linguistics, pp. 1023–1032.

Moustafa, H., M. Malli, and H. Hazimeh (2022). "Real-time Bitcoin price tendency awareness via social media content tracking". In: *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*, pp. 1–6.

Muzammal, M. et al. (2019). "Renovating blockchain with distributed databases: An open source system". In: *Future generation computer systems* 90, pp. 105–117.

Nærland, K. et al. (2017). "Blockchain to Rule the Waves-Nascent Design Principles for Reducing Risk and Uncertainty in Decentralized Environments." In: *Proceedings/International Conference on Information Systems (ICIS)*.

Nagaoka, S. et al. (2010). "Patent statistics as an innovation indicator". In: *Handbook of the Economics of Innovation*. Vol. 2. Elsevier, pp. 1083–1127.

Nakamoto, S. (2008). *A peer-to-peer electronic cash system*. Accessed on Oct. 10 2019. URL: https://bitcoin.org/bitcoin.pdf.

Narman, H. S., A. D. Uulu, and J. Liu (2018). "Profile analysis for cryptocurrency in social media". In: *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 229–234.

Natarajan, H., S. Krause, and H. Gradstein (2017). “Distributed ledger technology and blockchain”. In: URL: http://documents.worldbank.org/curated/en/134831513333483951/Distributed-Ledger-Technology-DLT-and-blockchain-Fintech-note-no-1.

Ndlovu, R. (2022). “Four Out of Five Diasporan Zimbabweans Live in South Africa”. In: *Bloomberg*. URL: https://www.bloomberg.com/news/articles/2022-09-06/four-out-of-five-diasporan-zimbabweans-live-in-south-africa#xj4y7vzkg.

Newman, D. et al. (2010). “Automatic evaluation of topic coherence”. In: *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, pp. 100–108.

Nguyen, T. H. and K. Shirai (2015). “Topic modeling based sentiment analysis on social media for stock market prediction”. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1354–1364.

Nielsen, F. Å. (2011). “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs”. In: *arXiv preprint arXiv:1103.2903*.

Nizzoli, L. et al. (2020). “Charting the Landscape of Online Cryptocurrency Manipulation”. In: *IEEE ACCESS* 8, pp. 113230–113245.

Norta, A., B. Leiding, and A. Lane (2019). “Lowering Financial Inclusion Barriers with a Blockchain-Based Capital Transfer System”. In: *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, pp. 319–324. ISBN: 978-1-7281-1878-9.

OECD (2022). *Development Co-operation Report 2021*. Tech. rep. URL: https://www.oecd-ilibrary.org/content/publication/ce08832f-en.

Ogunmuyiwa, M. S. and B. Okunleye (2019). “Small and medium enterprises and sustainable economic development in Nigeria”. In: *Izvestiya: Journal of Varna University of Economics* 3.

Ortu, M. et al. (2022). “On technical trading and social media indicators for cryptocurrency price classification through deep learning”. In: *Expert Systems with Applications* 198, p. 116804.

Ostercamp, P. (2022). “Stablecoin Regulation: EU, UK and US Perspectives”. In: *Available at SSRN 4038843*.

Page, M. J. et al. (2021a). “PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews”. In: *BMJ* 372.n160.

Page, M. J. et al. (2021b). “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews”. In: *BMJ (Clinical research ed.)* 372.n71.

Pagnotta, E. (2022). “Bitcoin as Decentralized Money: Prices, Mining Rewards, and Network Security”. In: *Review of Financial Studies* 35(2), pp. 866–907.

Pan, L., L. Feng, and Q. Jiayin (2020). “Adaptive Evolution Mechanism of Blockchain Community Based on Token-based Halving Event”. In: *2020 Chinese Automation Congress (CAC)*, pp. 6140–6144.

Pant, D. R. et al. (2018). “Recurrent neural network based Bitcoin price prediction by Twitter sentiment analysis”. In: *2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*, pp. 128–132.

Park, J. Y. and C. S. Sung (2020). “A business model analysis of blockchain technology-based startup”. In: *Entrepreneurship and Sustainability Issues* 7.4, pp. 3048–3060.

Patel, R., M. Migliavacca, and M. E. Oriani (2022). “Blockchain in banking and finance: A bibliometric review”. In: *Research in International Business and Finance* 62, p. 101718.

Patil, A. P., T. S. Akarsh, and A. Parkavi (2018). “A Study of Opinion Mining and Data Mining Techniques to Analyse the Cryptocurrency Market”. In: *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pp. 198–203.

Pennington, J., R. Socher, and C. D. Manning (2014). “GloVe: Global vectors for word representation”. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Perdana, A. et al. (2021). “Distributed ledger technology: Its evolutionary path and the road ahead”. In: *Information and Management* 58.3.

Peters, G. W. and E. Panayi (2016). “Understanding modern banking ledgers through blockchain technologies: Future of transaction processing and smart contracts on the internet of money”. In: *Banking beyond banks and money*. Springer, pp. 239–278.

Phillips, R. C. and D. Gorse (2018). “Mutual-excitation of cryptocurrency market returns and social media topics”. In: *4th International Conference on Frontiers of Educational Technologies*, pp. 80–86.

Pilkington, M. (2016). “Blockchain technology: principles and applications”. In: *Research handbook on digital transformations*. Edward Elgar Publishing.

PitchBook (2023). *Crypto Report: VC trends and emerging opportunities*.

Polasik, M. et al. (2015). “Price Fluctuations and the Use of Bitcoin: An Empirical Inquiry”. In: *International Journal of Electronic Commerce* 20.1, pp. 9–49.

Prasad, E. (2022). “Enduring Preeminence: The US dollar might slip, but it will continue to rule”. In: *Finance & Development* 59 (002), p. 70.

Prokhorenkova, L. et al. (2018). “CatBoost: unbiased boosting with categorical features”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio et al. Vol. 31. Curran Associates, Inc.

Pustišek, M., N. Živic, and A. Kos (2022). *Blockchain Technology and applications for Industry 4.0, smart energy, and smart cities*. Berlin and Boston: De Gruyter. DOI: 10.1515/9783110681130.

Qin, K. et al. (2021). “Attacking the defi ecosystem with flash loans for fun and profit”. In: *International conference on financial cryptography and data security*. Springer, pp. 3–32.

Rahman, S. et al. (2018). “Sentiment analysis using R: An approach to correlate cryptocurrency price fluctuations with change in user sentiment using machine learning”. In: *2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pp. 492–497.

Ramos, J. (2003). “Using TF-IDF to determine word relevance in document queries”. In: *Proceedings of the first instructional conference on machine learning*. Vol. 242. 1, pp. 29–48.

Redman, J. (2015). “$1 Billion Invested So Far in Bitcoin & Blockchain Infrastructure”. In: *Cointelegraph*. URL: https://cointelegraph.com/news/1-billion-invested-so-far-in-bitcoin-blockchain-infrastructure.

Reimers, N. and I. Gurevych (2020). “Making monolingual sentence embeddings multilingual using knowledge distillation”. In: *arXiv preprint arXiv:2004.09813*.

Rejeb, A. et al. (2019). “Leveraging the internet of things and blockchain technology in supply chain management”. In: *Future Internet* 11.7, p. 161.

Rice, T., G. von Peter, and C. Boar (2020). "On the global retreat of correspondent banks". In: *BIS Quarterly Review, March*.

Röder, M. et al. (2015). "Exploring the space of topic coherence measures". In: *Proceedings of the eighth ACM international conference on Web search and data mining*, pp. 399–408.

Rognone, L., S. Hyde, and S. S. Zhang (2020). "News sentiment in the cryptocurrency market: An empirical comparison with Forex". In: *International Review of Financial Analysis* 69, p. 101462.

Rosenthal, R. (1979). "The file drawer problem and tolerance for null results." In: *Psychological bulletin* 86.3, p. 638.

Ruoti, S. et al. (2019). "Blockchain technology: what is it good for?" In: *Communications of the ACM* 63.1, pp. 46–53.

Rühmann, F. et al. (2020). "Can blockchain technology reduce the cost of remittances?" In: *OECD Development Co-operation Working Papers* 73.

Saberi, S. et al. (2019). "Blockchain technology and its relationships to sustainable supply chain management". In: *International Journal of Production Research* 57.7, pp. 2117–2135.

Salton, G., C.-S. Yang, and C. T. Yu (1975). "A theory of term importance in automatic text analysis". In: *Journal of the American society for Information Science* 26.1, pp. 33–44.

Sams, R. (2015). "A note on cryptocurrency stabilisation: Seigniorage shares". In: *Brave New Coin*, pp. 1–8.

Sapkota, N. and K. Grobys (2021). "Fear Sells: Determinants of Fund-Raising Success in the cross-section of Initial Coin Offerings". In: *Available at SSRN 3843138*.

Şaşmaz, E. and F. B. Tek (2021). "Tweet sentiment analysis for cryptocurrencies". In: *2021 6th International Conference on Computer Science and Engineering (UBMK)*, pp. 613–618.

Sattarov, O. et al. (2020). "Forecasting Bitcoin price fluctuation by Twitter sentiment analysis". In: *2020 International Conference on Information Science and Communications Technologies (ICISCT)*. ISBN: 978-1-7281-9969-6.

Sayeed, S., H. Marco-Gisbert, and T. Caira (2020). "Smart contract: Attacks and protections". In: *IEEE Access* 8, pp. 24416–24427.

Schwenkler, G. and H. Zheng (2021). "News-Driven Peer Co-Movement in Crypto Markets". In: *Available at SSRN 3572471*.

Sebastião, H. and P. Godinho (2021). "Forecasting and trading cryptocurrencies with machine learning under changing market conditions". In: *Financial Innovation* 7.1, pp. 1–30.

SEC (2022). *Staff Accounting Bulletin No. 121*. URL: https://www.sec.gov/oca/staff-accounting-bulletin-121.

Senter, R. and E. A. Smith (1967). "Automated readability index". In: *AMRL-TR. Aerospace Medical Research Laboratories (U.S.)*, pp. 1–14.

Shahaab, A. et al. (2020). "Blockchain-a panacea for trust challenges in public services? A socio-technical perspective". In: *The Journal of the British Blockchain Association*.

Shahid, M. N. and H. Jungpil (2020). "A cross-disciplinary review of blockchain research trends and methodologies: Topic modeling approach". In: *53rd Annual Hawaii International Conference on System Sciences, HICSS 2020*. Vol. 2020-January, pp. 4053–4059.

Shin, D. and M. Ibahrine (2020). "The socio-technical assemblages of blockchain system: How blockchains are framed and how the framing reflects societal contexts". In: *Digital Policy, Regulation and Governance* 22.3, pp. 245–263.

Siddaway, A. P., A. M. Wood, and L. V. Hedges (2019). "How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses". In: *Annual review of psychology* 70, pp. 747–770.

Sikorski, J. J. et al. (2017). "Blockchain technology in the chemical industry: Machine-to-machine electricity market". In: *Applied energy* 195, pp. 234–246.

Silva, C. F. da and S. Moro (2021). "Blockchain technology as an enabler of consumer trust: A text mining literature analysis". In: *Telematics and Informatics* 60, p. 101593.

Siu, G. A., B. Collier, and A. Hutchings (2021). "Follow the money: The relationship between currency exchange and illicit behaviour in an underground forum". In: *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 191–201.

Stanley, M. (2019). "The application of behavioural heuristics to Initial Coin Offerings valuation and investment". In: *The Journal of The British Blockchain Association* 2.1, p. 7776.

Steinert, L. and C. Herff (2018). "Predicting altcoin returns using social media". In: *PloS one* 13.12, e0208119.

Stone, P. J., D. C. Dunphy, and M. S. Smith (1966). *The general inquirer: A computer approach to content analysis*. The general inquirer: A computer approach to content analysis. Oxford, England: M.I.T. Press.

Stratopoulos, T. C., V. X. Wang, and H. Ye (2022). "Use of Corporate Disclosures to Identify the Stage of Blockchain Adoption". In: *Accounting Horizons* 36.1, pp. 197–220.

Swan, M. (2015). *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc."

Tasca, P. and C. J. Tessone (2019). "Taxonomy of blockchain technologies. Principles of identification and classification". In: *Ledger* 4, pp. 1–39.

Thewissen, J. et al. (2022). "Unpacking the black box of ICO white papers: A topic modeling approach". In: *Journal of Corporate Finance*, p. 102225.

Thomason, J. et al. (2018). "Blockchain—Powering and Empowering the Poor in Developing Countries". In: *Transforming climate finance and green investment with blockchains*. Ed. by A. Marke. Amsterdam: Academic Press, pp. 137–152.

Tian, F. (2016). "An agri-food supply chain traceability system for China based on RFID & blockchain technology". In: *2016 13th international conference on service systems and service management (ICSSSM)*. IEEE, pp. 1–6.

Toma, A. M. and P. Cerchiello (2020). "Initial coin offerings: Risk or opportunity?" In: *Frontiers in artificial intelligence* 3, p. 18.

UN (2019). *Remittances matter: 8 facts you don't know about the money migrants send back home*. URL: https://www.un.org/development/desa/en/news/population/remittances-matter.html.

United States Attorney's Office (2022). *U.S. Attorney Announces Historic $3.36 Billion Cryptocurrency Seizure And Conviction In Connection With Silk Road Dark Web Fraud*. URL: https://www.justice.gov/usao-sdny/pr/us-attorney-announces-historic-336-billion-cryptocurrency-seizure-and-conviction#:~:text=ZHONG%20pled%20guilty%20on%20Friday,valued%20at%20over%20%243.36%20billion..

U.Today (2023). *Ethereum (ETH): Shanghai Hard Fork Causes Concern Among Developers, Here Are Reasons*.

Vacca, S. et al. (2021). "Investigation of Coronavirus impact on Blockchain and cryptocurrencies markets". In: *2021 IEEE/ACM 4th International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*, pp. 56–60.

Valencia, F., A. Gómez-Espinosa, and B. Valdés-Aguirre (2019). "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning". In: *Entropy* 21.6, p. 589.

Vohra, S. and J. Teraiya (2013). "A comparative study of sentiment analysis techniques". In: *Journal Jikrce* 2.2, pp. 313–317.

Voskobojnikov, A. et al. (2021). "The U in crypto stands for usable: An empirical study of user experience with mobile cryptocurrency wallets". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Wang, J. et al. (2021a). "Technology hotspot tracking: Topic discovery and evolution of china's blockchain patents based on a dynamic LDA model". In: *Symmetry* 13.3.

Wang, Q. et al. (2021b). "Non-fungible token (NFT): Overview, evaluation, opportunities and challenges". In: *arXiv preprint arXiv:2105.07447*.

Wang, Y. et al. (2019). "Understanding blockchain technology for future supply chains: a systematic literature review and research agenda". In: *Supply Chain Management: An International Journal* 24.1, pp. 62–84.

Werner, S. et al. (2022). "Sok: Decentralized finance (DeFi)". In: *Proceedings of the 4th ACM Conference on Advances in Financial Technologies*, pp. 30–46.

World Bank (2021). *Central Bank Digital Currencies for Cross-border Payments: A Review of Current Experiments and Ideas*. URL: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/369001638871862939/central-bank-digital-currencies-for-cross-border-payments-a-review-of-current-experiments-and-ideas.

— (2022). *Financial Inclusion Overview*. URL: https://www.worldbank.org/en/topic/financialinclusion/overview.

Xiao, Y. et al. (2020). "A survey of distributed consensus protocols for blockchain networks". In: *IEEE Communications Surveys & Tutorials* 22.2, pp. 1432–1465.

Xie, P. (2021). "The Interplay Between Investor Activity on Virtual Investment Community and the Trading Dynamics: Evidence From the Bitcoin Market". In: *Information Systems Frontiers*.

Xu, M., X. Chen, and G. Kou (2019). "A systematic review of blockchain". In: *Financial Innovation* 5.1, pp. 1–14.

Xu, W. et al. (2021). "Prediction of initial coin offering success based on team knowledge and expert evaluation". In: *Decision Support Systems* 147, p. 113574.

Xu, X. F. and Y. Y. He (2022). "Blockchain application in modern logistics information sharing: A review and case study analysis". In: *Production Planning & Control*, pp. 1–15.

Xu, X. et al. (2017). "A taxonomy of blockchain-based systems for architecture design". In: *2017 IEEE international conference on software architecture (ICSA)*. IEEE, pp. 243–252.

Yadav, J. et al. (2021). "Netizens' behavior towards a blockchain-based esports framework: A TPB and machine learning integrated approach". In: *International Journal of Sports Marketing and Sponsorship* 23.4, pp. 665–683.

Yao, W., K. Xu, and Q. Li (2019). "Exploring the influence of news articles on Bitcoin price with machine learning". In: *2019 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–6.

Yen, J.-C. and T. Wang (2021). "Stock price relevance of voluntary disclosures about blockchain technology and cryptocurrencies". In: *International Journal of Accounting Information Systems* 40, p. 100499.

Yen, J.-C., T. Wang, and Y.-H. Chen (2021). "Different is better: How unique initial coin offering language in white papers enhances success". In: *Accounting & Finance* 61.4, pp. 5309–5340.

Youtie, J. et al. (2008). "Assessing the nature of nanotechnology: can we uncover an emerging general purpose technology?" In: *The Journal of Technology Transfer* 33.3, pp. 315–329.

Zeng, S. et al. (2018). "A Bibliometric Analysis of Blockchain Research". In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. Vol. 2018-June, pp. 102–107.

Zhang, H., T. Daim, and Y. P. Zhang (2021). "Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain". In: *Technological Forecasting and Social Change* 167, p. 120729.

Zhang, P. et al. (2018). "Blockchain technology use cases in healthcare". In: *Advances in computers*. Vol. 111. Elsevier, pp. 1–41.

Zhang, Y. et al. (2021). "Code Will Speak: Early detection of Ponzi Smart Contracts on Ethereum". In: *2021 IEEE International Conference on Services Computing (SCC)*, pp. 301–308.

Zhang, Y., R. Jin, and Z.-H. Zhou (2010). "Understanding bag-of-words model: A statistical framework". In: *International journal of machine learning and cybernetics* 1.1, pp. 43–52.

Zhao, J. L. et al. (2016). "Overview of business innovations and research opportunities in blockchain and introduction to the special issue". In: *Financial Innovation* 2.28.

Zhou, L. et al. (2023). "SoK: Decentralized Finance (DeFi) Attacks". In: *IEEE Symposium on Security and Privacy (SP)*, pp. 2444–2461.

Zhou, Q. et al. (2020). "Solutions to scalability of blockchain: A survey". In: *IEEE Access* 8, pp. 16440–16455.

Zhuo, X., F. Irresberger, and D. Bostandzic (2023). "Blockchain for Cross-border Payments and Financial Inclusion: The Case of Stellar Network". In: *Available at SSRN 4550837*.

Zīle, K. and R. Strazdiņa (2018). "Blockchain use cases and their feasibility". In: *Applied Computer Systems* 23.1, pp. 12–20.

Eidesstattliche Erklärung

Ich, Xian Zhuo, versichere an Eides statt, dass die vorliegende Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf erstellt worden ist.

Düsseldorf, June 27, 2024