*Article*

# Mapping the Landscape of Blockchain Technology Knowledge: A Patent Co-Citation and Semantic Similarity Approach

**Brian Tae-Seok Kim [1] and Eun-Jung Hyun [2,\*]**

[1] School of Commerce, Waseda University, Tokyo 169-8050, Japan
[2] College of Business Administration, Hongik University, Seoul 04066, Republic of Korea
\* Correspondence: ejhyun@hongik.ac.kr

**Abstract:** The potential applications of blockchain technology across various business functions and industries have generated significant interest. However, its underlying knowledge structure remains unclear. This study aimed to gain a deeper understanding of the technological domain and knowledge structure of blockchain technology by analyzing 4753 USPTO patent data from 2008 to 2019. We used multiple approaches, such as analyzing patent filing volumes, constructing co-citation networks, and examining text (patent abstract) data with a variant of bidirectional encoder representations from transformers (BERT). The results demonstrate the advantages of using an NLP-based BERT text analysis approach for examining technological knowledge and relationships within the blockchain technology field. Our findings reveal that the field of blockchain technology is expanding and diversifying, with increasing patent filings in both cryptocurrency and distributed ledger technologies and growing knowledge similarity between these two subdomains. We also found that patent assignees (companies) engage differently in innovative activities within the blockchain technology domain based on their prior experience in the field. These results hold potential for informing future research in emerging technology studies and guiding industry and policy decisions related to blockchain technology.

**Keywords:** blockchain technology; patent data; co-citation network; semantic analysis; natural language processing; bidirectional encoder representations from transformers (BERT)

## 1. Introduction

Blockchain technology has been the subject of much interest in recent years due to its transformative potential across a range of domains, including finance, supply chain management, and digital identity verification. At its core, blockchain technology is a system of linked blocks that securely record multiple transactions using cryptography. The secure and tamper-proof nature of blockchain makes it an ideal solution for applications that require transparency, security, and decentralization [1,2]. Distributed ledger technology extends these benefits by allowing for the secure and decentralized recording of information across a network of computers, creating a shared digital database where transactions and data can be stored and accessed simultaneously from multiple locations [3–5].

The decentralized and secure nature of blockchain technology can be used to create more efficient and transparent systems in various industries such as finance, healthcare, commerce, and supply chain management [3,6,7]. Many financial institutions and banks are exploring the use of blockchain technology for various applications such as cross-border payments, securities trading, and digital identity management. In the healthcare industry, blockchain technology is being explored as a way to securely store and share patient data, improve supply chain transparency, and combat counterfeit drugs [8–11]. In commerce and supply chain management, blockchain technology is being used to create more efficient and transparent systems for tracking goods as they move through the supply chain and improve the traceability of products [3,7,12].

However, despite this interest and the availability of large patent data, there is still a lack of studies conducted to gain a systematic, in-depth understanding of the underlying knowledge structure of blockchain technology [13]. In this study, we fill the gap in the literature using 4753 USPTO patent data from 2008 to 2019. We conducted multiple analyses to explore the landscape of the blockchain knowledge space. Although the technology continues to evolve, we think that it is still important to take stock of the current state of the field and to identify key players, incumbents, or new entrants that are participating in the field through an analysis of the patent data available to researchers.

The first approach was to analyze patent filing volumes by different subdomains and players, which can provide insight into how interest in this field has evolved. This can give us an understanding of the areas in which companies and institutions are focusing their research and development efforts. The second approach was to analyze the co-citation network, which can reveal the relationships and connections between different patents constituting the blockchain technology knowledge space. By identifying patterns of co-citation, we can infer the relationships between different technologies and subdomains [14–16]. This can help us understand how the different components of blockchain technology are related to each other and how they are evolving over time [17]. The third approach, semantic similarity analysis, was to analyze similarities between text data in abstracts. To this end, we employ a variant of bidirectional encoder representations from transformers (BERT), which is a natural language processing (NLP) algorithm specifically designed for understanding patents. This semantic analysis enables us to gain a more precise understanding of relationships between patents that comprise the semantic space of block technology by leveraging the technical details of the patents, which is difficult to achieve when researchers rely only on the co-citation network analysis.

Each of these approaches can provide different insights into the technological knowledge embedded in patent data. For example, by analyzing patent filing volumes, we can see how interest in different subdomains of blockchain technology has evolved over time, and they vary by players (assignees). By analyzing the co-citation network, we can understand how different technologies and subdomains are related to each other. Additionally, by conducting semantic similarity analysis using texts in abstracts, we can develop similarity metrics that can be utilized to identify potentially promising knowledge areas in the blockchain technology domain. Results revealed that cryptocurrency-related and distributed ledger patents, the two main technological subdomains of blockchain technology, exhibit different patterns of evolution and structure. In particular, the technology of distributed ledgers seems to have a wide range of potential applications beyond just cryptocurrency, and the knowledge structure of blockchain technology is constantly evolving.

Overall, this study is expected to contribute to the related literature by providing a comprehensive analysis of the technological subdomains and the underlying knowledge structure of blockchain technology. Additionally, our work demonstrates the importance of understanding the underlying knowledge structure of emerging technologies. By allowing a detailed understanding of the potential of the technology, our study also has the potential to inform future research in emerging technology studies, as well as guide industry and policy decisions related to blockchain technology.

## 2. Literature Review

### 2.1. Prior Research on Blockchain Technology Applications

Blockchain technology has become a widely recognized tool for innovation and transformation in the business world, with potential to disrupt conventional processes and provide secure, transparent, and decentralized solutions across a variety of industries. This literature review aims to provide an overview of prior studies that have analyzed the opportunities and challenges of blockchain in various business functions and areas. We survey the findings and insights of these studies, which are organized into four parts, covering (1) blockchain applications in financial transactions, (2) supply and manufacturing, (3) the promotion of sustainability, and (4) the issues related to its decentralized, distributed

nature. The aim of this review is to offer a comprehensive understanding of the potential benefits and challenges of blockchain technology.

### 2.1.1. Blockchain Applications in Financial Transactions

The utilization of blockchain technology has been a significant area of interest in the financial sector, where it has revolutionized the way transactions are processed and recorded. Several studies have explored the benefits and challenges of adopting blockchain technology in financial services, with a focus on the banking industry, pension industry, and online auction platforms. Garg et al. [18] found that blockchain technology can offer increased efficiency, cost savings, improved security, and enhanced customer experience for banks through a combined analysis of the task-technology fit and technology acceptance theory. Chang et al. [19] identified trust, security, privacy, scalability, and cost as factors that affect the adoption of blockchain technology and noted that it can promote transactions and reduce costs in the banking and financial services sector. Liu et al. [20] found that the trend of research in the FinTech industry is moving towards blockchain and crowdfunding, which are likely to have a profound impact on the FinTech business model.

Moreover, Ali et al. [21] explored the growth of the nonfungible token (NFT) market-place and found that the trading of NFTs has influenced the growth of the decentralized application marketplace. Sarker and Datta [22] noted that blockchain technology has the potential to transform the pension industry by reducing turnaround time, lowering operating expenses, and facilitating pension reform agendas. Omar et al. [2] proposed a general framework for decentralized auctions using blockchain technology, which eliminates intermediaries, ensures transparency, and reduces transaction costs, and found it to be economically feasible and secure. In short, these studies suggest that blockchain technology holds significant potential for transformation in various financial services; however, further research is needed to fully understand its potential and challenges.

### 2.1.2. Blockchain Applications in Supply Chain Management and Manufacturing

The application of blockchain technology in supply chains has garnered interest from various researchers, who have explored the potential of blockchain in enhancing supply chain management through increased transparency, accountability, decentralization, and automation. One study by Wan et al. [23] found that blockchain technology has the potential to enhance collaborative innovation in the manufacturing industry in China by strengthening the positive impact of social trust. Rodríguez-Espíndola et al. [24] developed a model to understand the adoption of Industry 4.0 technologies, including blockchain, for risk management from an operations manager's perspective and found that perceived usefulness and adoption of these technologies are influenced by digital transformation maturity, market pressure, regulations, and resilience. Kamble et al. [25] proposed a decision support system for predicting the probability of blockchain adoption in organizations based on factors such as competitor pressure, partner readiness, perceived usefulness, and perceived ease of use. Chang et al. [3] proposed a blockchain-based framework for supply chain processes using smart contracts to demonstrate its benefits, while Kamble et al. [25] reviewed the current state of digital supply chain twins and the role of various technologies, including blockchain, in their development. Dal Mas et al. [26] analyzed the opportunities and potential of blockchain technologies for agrifood sustainability and found that it plays a crucial role in the digital transformation of agrifood supply chains. Pincheira et al. [27] also explored the potential of blockchain technology to improve agrifood traceability systems by examining the characteristics and costs associated with integrating blockchain and Internet of Things (IoT) technologies into agrifood traceability systems. They also discuss the advantages that such an integration could bring, such as increased transparency, improved data security, and reduced costs.

Together, these studies provide a comprehensive understanding of the potential of blockchain technology in supply chain management and the benefits it can bring, such as increased transparency, accountability, decentralization, and automation, as well as

contribution to sustainability goals. These findings offer insights for firms, policymakers, and researchers to effectively implement blockchain in supply chain management.

2.1.3. Blockchain Applications in Promoting Sustainability

Blockchain technology has shown potential for promoting sustainability in various domains. In social businesses, Devine et al. [28] explored the use of blockchain and smart contracts to build trust and support the coexistence of social and economic logics of social ventures. The authors present a social business blockchain model codifying the principles of social business as smart contract functions, offering insights into how blockchain can be used to improve the sustainability of social ventures and improve trust relationships between stakeholders. Marsal-Llacuna [29] proposed using blockchain to reimagine the delivery of smart city agendas and their performance measurement, as a means to make them more empowering and collaborative. The author suggests using blockchain to create a people-centric approach while also creating new measurement tools.

Chin et al. [30] suggested that blockchain technology can drive green innovation in ecosystem-based business models by creating a secure, transparent, and efficient system for value exchange between stakeholders. Friedman and Ormiston [31] explored the potential of blockchain technology to contribute to sustainable transformations within food supply chains through expert interviews and found that blockchain can help build trust, increase transparency, and promote sustainability in food supply chains. Köhler et al. [7] assessed the relationship between blockchain-based technologies and voluntary sustainability standards, finding that most cases are coexisting, with a few cases having a synergistic relationship and one case having an antagonistic relationship.

Pazaitis et al. [32] discussed the potential of blockchain technology in creating a new system of value that better supports the dynamics of social sharing. The authors envision a blockchain-based decentralized cooperation that can enable the creation of commons-oriented ecosystems in the sharing economy. Pólvora et al. [33] argued for a transdisciplinary approach to address the uncertainties and challenges associated with the development and uptake of blockchain technology in Europe, based on a research project that focused on multistakeholder engagement and cocreation. Overall, the studies suggest that blockchain technology has the potential to promote sustainability through increased trust and transparency in social businesses, improved smart city agendas, green innovation in business models, improved sustainability governance in agrifood supply chains, and the creation of a new system of value in the sharing economy.

2.1.4. Decentralized and Distributed Feature of Blockchain Technology

Blockchain technology has been a topic of extensive research for its decentralized and distributed nature and its impact on various business applications. In recent years, various studies have shed light on different aspects of blockchain technology. Grida et al. [34] examined factors that need to be considered when adopting and implementing blockchain technology, in particular, distributed ledger technology. These factors include security, scalability, privacy, interoperability, cost-effectiveness, and user experience. The paper also discusses the potential benefits of blockchain technology such as increased efficiency, transparency, and trustworthiness. Liu et al. [35] discussed the use of a distributed storage scheme based on blockchain technology to store images securely and efficiently, which combines the advantages of distributed pooling and blockchain technology, allowing for secure and efficient image storage.

Hou et al. [4] explored the potential of blockchain to improve the competitiveness of distributed energy resources in China by enabling the integration of advanced technologies such as AI and IoT. Nguyen and Nguyen [36] studied the relationship between platform decentralization and market value and found that centralization enhances market value through voluntary disclosures and developers' activities, while decentralization promotes developers' activities to enhance market value. Nam et al. [37] proposed an IP decentralized ledger based on blockchain to reduce IP administration costs and improve IP use. Liu

et al. [38] propose an improved method for change address identification, which uses a combination of heuristics and machine learning techniques to process raw datasets, allowing for faster and more accurate clustering of Bitcoin addresses.

Santana and Albareda [39] reviewed the literature on decentralized autonomous organizations and proposed an integrative model based on decentralization, automation, and autonomy. Pereira et al. [40] compared blockchain-based platforms with centralized platforms and evaluated the benefits and costs of each in terms of transaction costs, technology costs, and community involvement. Jovanovic et al. [41] studied the impact of decentralization on information security in blockchain networks and found that while decentralization can improve security, it also creates new security risks. These studies collectively demonstrate the benefits, challenges, and implications of the decentralized and distributed nature of blockchain technology in various domains.

Overall, the literature reviewed in this section highlights the widespread recognition of blockchain technology as a catalyst for transformation and innovation in the business world. These studies suggest that blockchain is being applied to various business functions, such as the service sector and supply chain logistics, to enhance security, transparency, and efficiency. This is because many organizations and decision-makers hold that this technology can offer significant opportunities to reduce operational costs, improve business operations, and drive positive change in multiple industries. Together, a growing body of scholarly work on business applications of blockchain technology provides insights into the benefits and challenges associated with the implementation of blockchain technology and emphasizes the importance of considering its unique characteristics, such as its decentralized and distributed nature, in the adoption process.

### 2.2. Patent Analysis in Prior Research on Blockchain Technology

Patent data have proven to be a valuable resource for studying the development of blockchain technology and its applications. Various patent analysis techniques have been used to gain insights into the blockchain landscape and make informed decisions about its future use. Co-citation analysis, the latent Dirichlet allocation topic model, main path analysis, and patent network analysis are some of the methods used in these studies. The studies conducted using patent data explore various aspects of blockchain such as technological evolution, IP challenges, geographical distribution, and innovation trends.

For example, Daim et al. [42] used patent citations to understand the relative position of a company in the technological network of IoT, cybersecurity, and blockchain. Zhang et al. [43] developed a new analysis method, the latent Dirichlet allocation topic model, to assess technology development and predict future trends in the blockchain field. Yu and Pan [44] used main path analysis to understand the innovation path of blockchain technology. Zanella et al. [45] suggested a framework for analyzing blockchain-related patents using cosine- and density-based outlier analysis. Someda et al. [46] proposed a method for detecting industry sectors impacted by spillover effects of emerging technologies, such as blockchain technology, by combining patent analysis with input–output analysis to model knowledge spillover.

These studies show that patent analysis provides valuable insights into the development of blockchain technology and its applications. By analyzing patent data, researchers, policymakers, and industry actors can better understand the blockchain landscape and make informed decisions about its future development and implementation.

### 2.3. An Overview of This Study

As we reviewed above, the field of blockchain technology has seen rapid growth and development in recent years, with many organizations and industries exploring its potential applications. Despite this interest, there is limited understanding of the underlying knowledge structure of blockchain technology. To address this gap in the literature, this study aims to provide a comprehensive analysis of the technological landscape and the underlying knowledge structure of blockchain technology.

To achieve this goal, we conducted multiple analyses using 4753 USPTO patent data from 2008 to 2019. Our analysis included trend analysis to understand the evolution of the technological landscape and identify key players in the blockchain arena. Trend analysis involved analyzing the volume of patent filings over time to identify patterns and trends in the field. We also used co-citation network analysis to trace the overall patterns of technological relationships among blockchain patents over time. This approach allowed us to analyze the attributes of the entire network, such as network density, transitivity, and assortativity, and enabled us to identify the positional attributes of each patent in the network. In addition to these analyses, we also used semantic data analysis to provide a deeper understanding of the technological knowledge embedded in patent data. This approach involved using a variant of BERT, an NLP algorithm, to analyze the text data within patents. This helped us to better understand the technological landscape and knowledge structure of blockchain technology, including the subdomains and research areas that are being explored.

While previous studies have utilized patent data to gain insights into various aspects of blockchain technology, such as technological evolution, IP challenges, geographical distribution, and innovation trends, our study takes a unique approach by conducting a comprehensive analysis of the technological domain and underlying knowledge structure of blockchain technology using a combination of trend analysis, co-citation network analysis, and semantic analysis. This multiple analysis approach provides a more comprehensive understanding of the landscape and knowledge structure of blockchain technology, offering a deeper understanding of the technological knowledge embedded in patent data, the relationships among patents and key players in the field, and the evolution of the technology over time. Accordingly, our approach is a more robust and nuanced examination of the blockchain landscape, filling the gap in the literature by providing a more in-depth understanding of the field compared to prior studies that have focused on a single analysis method.

## 3. Data

*Blockchain Patent Data*

We constructed our sample of blockchain patents with the following steps. First, we retrieved from the Patentsview platform the entire universe of utility patents granted by the United States Patent and Trademark Office (USPTO) as of June 2022. Second, we focused on a subset of granted patents whose application dates were between 2008 and 2019. Since we are interested in the emergence and evolution of blockchain technology over time, we used application dates rather than grant dates to capture the earliest observable time recorded for the patents. Following other studies on blockchain patents and publications [46,47], we set 2008 as the starting year for our analysis since it marks the year Bitcoin, the most successful application of blockchain technology, was first disclosed in a whitepaper [48]. We chose 2019 as the end year to alleviate the right-censoring problem of working with granted patents.

Third, we used the search query method outlined by Clarke et al. [49] to identify blockchain patents. Their method combines keywords and patent classifications to pinpoint patents related to cryptocurrency, distributed ledger, and smart contract technology. Their search query was validated by subject matter experts and has been shown to effectively reduce false positive results. Our study applied this query introduced by Clarke et al. [49] to the USPTO database, resulting in the identification of a significant number of blockchain patents. It is important to note that our categorization of these patents is not meant to compare the concepts but rather to provide a comprehensive view of the types of blockchain patents during the study period. This approach allows us to give a comprehensive overview of the emergence and evolution of blockchain technology in the patent space.

Table 1 lists the combinations of keywords and CPC classes included in this query. Guided by this search query, we were able to screen a total of 4753 blockchain patents during the study period.

**Table 1.** Search query for blockchain patents.

| Concept | Keyword | CPC Class |
|---|---|---|
| Cryptocurrency | BITCOIN* OR BIT-COIN* OR COLORED-COIN* OR COLORED-COIN* OR CRYPTOCURRENC* OR CRYPTO-CURRENC* OR DOGECOIN OR DOGE-COIN OR ETHEREUM OR LITECOIN OR LITE-COIN OR ZEROCASH OR ZCASH | in combination with: G06 OR H04 |
| | COUNTERPARTY OR XCP OR DIGITALCURRENC* OR DIGITAL-CURRENC* OR ETHER OR LISK OR META-COIN* OR METACOIN* OR NAME-COIN* OR NAMECOIN* OR NXT OR STELLAR OR TYPE-COIN* OR TYPECOIN* OR ZEROCOIN OR ZERO-COIN | in combination with: H04L9/3247 OR H04L9/3249 OR H04L9/3252 OR H04L9/3255 OR H04L9/3257 OR H04L9/3236 OR H04L9/3239 OR H04L9/3242 OR H04L9/0637 OR H04L9/0643 OR H04L2209/38 OR H04L2209/56 OR H04L2209/30 OR H04L2209/46 OR H04L2209/463 OR H04L2209/466 OR G06Q20/065 OR G06Q20/0652 OR G06Q20/0655 OR G06Q20/0658 OR G06Q20/02 OR G06Q20/023 OR G06Q20/027 OR G06Q20/401 OR G06Q20/4012 OR G06Q20/4014 OR G06Q20/40145 OR G06Q20/4016 OR G06Q20/4018 |
| Distributed ledger | BLOCKCHAIN* OR BLOCK-CHAIN* OR BLOCKSIGN OR DISTRIBUTED-LEDGER OR FACTOM OR PAY-TO-SCRIPT-HASH OR P2SH OR PROOF-OF-STAKE OR SIDECHAIN* | in combination with: G06 OR H04 |
| | FORKING OR FORKS OR LEDGER* OR MERKLE-TREE OR MERKLETREE OR HASHTREE OR HASHTREE OR MERKLE-ROOT OR MERKLEROOT OR PROOF-OF-WORK OR HASH-CASH OR HASHCASH OR RIPPLE OR ZEROKNOWLEDGE OR ZERO-KNOWLEDGE | in combination with: H04L9/3247 OR H04L9/3249 OR H04L9/3252 OR H04L9/3255 OR H04L9/3257 OR H04L9/3236 OR H04L9/3239 OR H04L9/3242 OR H04L9/0637 OR H04L9/0643 OR H04L2209/38 OR H04L2209/56 OR H04L2209/30 OR H04L2209/46 OR H04L2209/463 OR H04L2209/466 OR G06Q20/065 OR G06Q20/0652 OR G06Q20/0655 OR G06Q20/0658 OR G06Q20/02 OR G06Q20/023 OR G06Q20/027 OR G06Q20/401 OR G06Q20/4012 OR G06Q20/4014 OR G06Q20/40145 OR G06Q20/4016 OR G06Q20/4018 |

**Table 1.** *Cont.*

| Concept | Keyword | CPC Class |
|---|---|---|
| | CODIUS OR SMART-CONTRACT* OR SMARTCONTRACT* | in combination with: G06 OR H04 |
| Smart contracts | CHAINCOD* OR HAWK OR ROOTSTOCK OR RSK OR SYMBIONT | in combination with: H04L9/3247 OR H04L9/3249 OR H04L9/3252 OR H04L9/3255 OR H04L9/3257 OR H04L9/3236 OR H04L9/3239 OR H04L9/3242 OR H04L9/0637 OR H04L9/0643 OR H04L2209/38 OR H04L2209/56 OR H04L2209/30 OR H04L2209/46 OR H04L2209/463 OR H04L2209/466 OR G06Q20/065 OR G06Q20/0652 OR G06Q20/0655 OR G06Q20/0658 OR G06Q20/02 OR G06Q20/023 OR G06Q20/027 OR G06Q20/401 OR G06Q20/4012 OR G06Q20/4014 OR G06Q20/40145 OR G06Q20/4016 OR G06Q20/4018 |

The table is excerpted from the paper by Clarke et al. ([5], Table 2).

## 4. Patent Trend Analysis

### 4.1. Trend Analysis by Subtechnology

The search query also allowed us to categorize the patents into two different types of blockchain technologies using this query. Among these patents, 2391 were classified as cryptocurrency technology, and 2707 patents were categorized as distributed ledger technology, with a few patents (345) being identified as both distributed ledger and cryptocurrency technology. For smart contract technology, there were very few patents (19), and all of them were simultaneously categorized as either cryptocurrency or distributed ledger technology. We observe two main categories of blockchain patents that are nearly evenly split: patents related to cryptocurrency and patents related to distributed ledgers. We use this categorization throughout the analysis.

Figure 1a illustrates the quarterly distribution of our sample blockchain patents by type. The distribution of the volume of blockchain patents applied across time reveals the growth of blockchain technology in the past few decades. The number of blockchain patents has been gradually increasing since 2008, particularly with explosive growth after the third quarter of 2017, which coincides with the cryptocurrency bubble period [50]. Figure 1b further displays the same distribution by the two types of blockchain patents. The volume of cryptocurrency patents has increased throughout the years at a gradual pace. In contrast, distributed ledger patents have experienced exponential growth in a relatively short period with the start of the bubble.

(**a**)



(**b**)
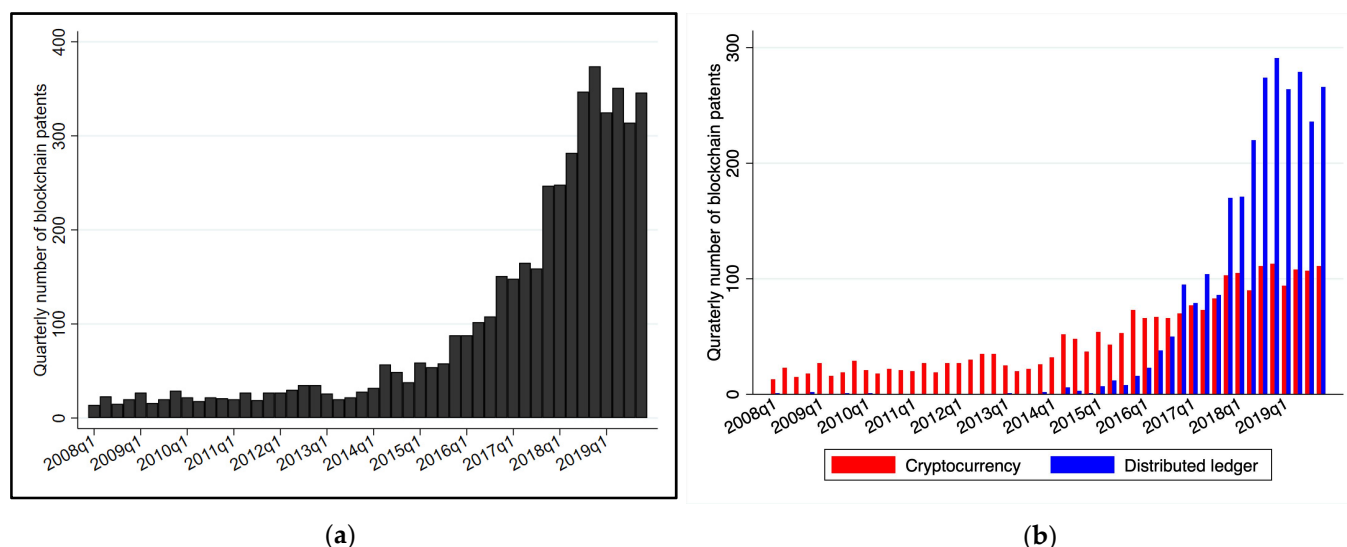
**Figure 1.** (**a**) Quarterly number of blockchain patents, (**b**) Quarterly number of blockchain patents by type.

*4.2. Trend Analysis by Patent Assignees*

We also coded three different types of assignees associated with sample blockchain patents. We differentiated assignees based on their prior activities in patenting blockchain technology. First, we discerned whether the assignee of each blockchain patent had invented a blockchain patent before (We searched for the entire patenting history of assignees predating the sample period to ensure the accurate identification of new entrants.). If a given patent was the assignee's first blockchain patent on the date of its application, we defined its assignee as a new entrant.

Second, we further differentiated remaining assignees based on whether the assignee had relatively small or large shares in the blockchain patents. Specifically, from the application date of each patent, we traced each assignee's share among all blockchain patents filed within a three-year window in the past. An assignee of a patent was coded as a small assignee if its share was below 1%, the median value in the distribution, and a large assignee if its share was above this threshold.

Table 2 illustrates the breakdown of the sample blockchain patents by assignee type. The breakdown by assignee type reveals a high concentration level in the blockchain technology domain. Blockchain patents of 48 large assignees account for 37% (1763) of the total, while those of 432 small assignees and 939 new entrants make up the rest.

**Table 2.** Breakdown of blockchain patents by assignee type.

| Assignee Type | Number of Assignees | Number of Patents |
|---|---|---|
| New entrant | 939 | 1022 |
| Small assignee | 432 | 1968 |
| Large assignee | 48 | 1763 |

Table 3 provides the list of the 15 largest assignees appearing in our sample. The list is consistent with prior studies on blockchain technology [49,51]. Big financial firms, such as Bank of America and Mastercard, and information and communication technology (ICT) giants, including International Business Machine, Intel, and Amazon, represent larger players in the blockchain technology domain.

**Table 3.** List of top 15 largest assignees.

| Assignee Name | Blockchain Patent | Crypto-Currency Patent | Distributed Ledger Patent |
|---|---|---|---|
| International Business Machines Corp. | 497 | 141 | 407 |
| Advanced New Technologies Co., Ltd. | 131 | 39 | 120 |
| Alibaba Group Holding Limited | 122 | 23 | 112 |
| Bank of America Corporation | 120 | 55 | 76 |
| Mastercard International Incorporated | 113 | 48 | 65 |
| Capital One Services, LLC | 108 | 67 | 49 |
| Amazon Technologies, Inc. | 108 | 95 | 13 |
| Accenture Global Solutions Limited | 75 | 14 | 69 |
| Intel Corporation | 66 | 46 | 23 |
| Visa International Service Association | 62 | 51 | 14 |
| Microsoft Technology Licensing, LLC | 56 | 27 | 33 |
| Wells Fargo Bank, N.A. | 54 | 28 | 29 |
| Paypal, Inc. | 54 | 40 | 16 |
| Cisco Technology, Inc. | 48 | 18 | 32 |
| EMC IP Holding Company LLC | 48 | 16 | 35 |

Figure 2 visualizes the quarterly number of different types of assignees actively patenting blockchain technology. The number of active large-share assignees has been relatively constant over time. However, in recent years, we observe an influx of a large number of new assignee organizations and an increase in the number of active small-share assignees in the blockchain technology domain.
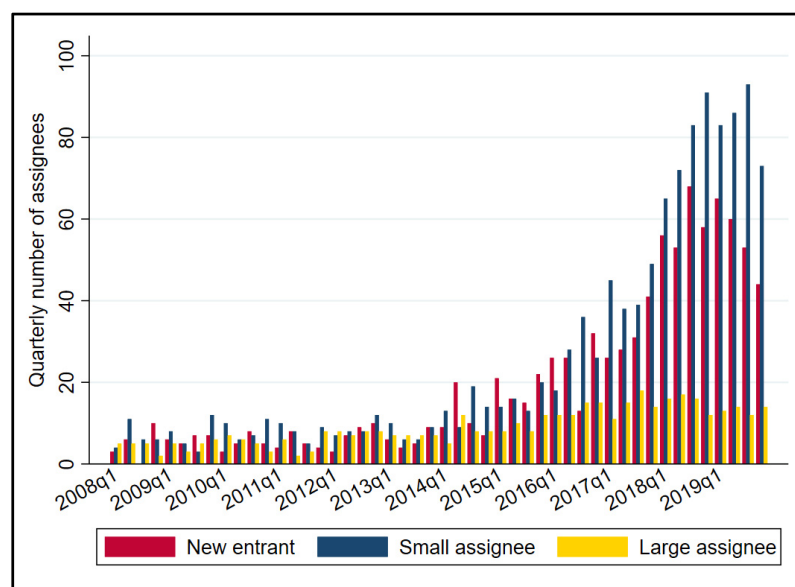


**Figure 2.** Quarterly number of blockchain patent assignees by type.

## 5. Co-Citation Network Analysis

In this section, we show how we use co-citation network analysis to examine the connections among patents in the blockchain technology field. A patent's backward citations to other previously published patent documents represent an inventory of prior knowledge

it draws on. Sharing common backward citations or having co-citations can be a proxy for how a pair of patents are technologically related to one another [42,52].

In particular, co-citation network analysis offers two benefits to examining the evolution of blockchain patents. First, we can trace how the overall patterns of technological relationships among blockchain patents have evolved over time by tracing the attributes of the entire networks (e.g., network density, transitivity, and assortativity) across time. Second, we can also specifically identify the positional attributes (e.g., degree centrality, eigenvector centrality, and coreness) of each patent in the network, further providing insights into understanding variation among patents and how this variation relates to other patent-level attributes such as assignee types. We describe below the construction of co-citation networks included in our analysis.

### 5.1. Citation Data

We constructed co-citation networks of the sample blockchain patents in the following steps. We first traced a total of 117,930 backward citations (about 37 citations per each patent on average) our sample patents had made to 69,536 patent publications. Figure 3 illustrates the distribution of the volume of backward citations across the quarterly dates of the sample blockchain patents and their backward-cited patents. The heatmap not only reveals that the volume of backward citations of blockchain patents has increased in recent years but also shows that more diverse knowledge and new ideas underlie recent blockchain patents. While further analysis may be required to confirm this, we expect to see fewer connections (as determined by co-citation) among newer blockchain patents.
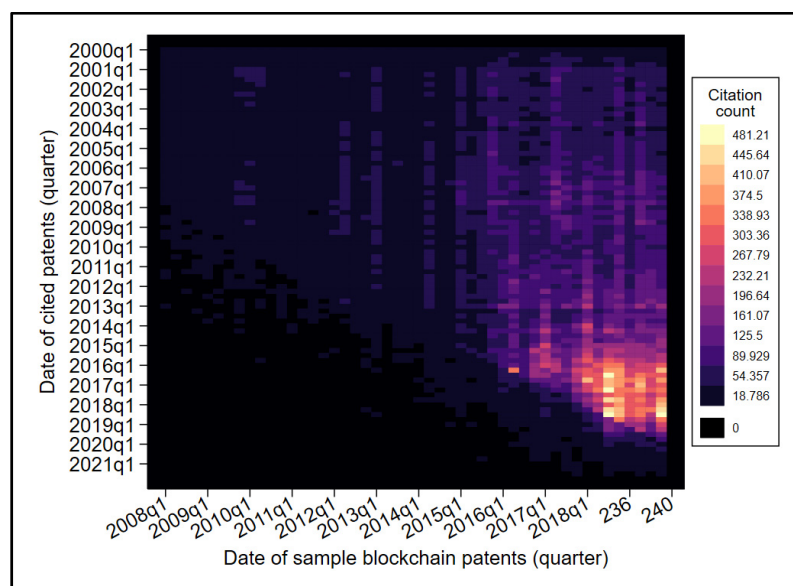


**Figure 3.** Heatmap of citation counts across the dates of sample and backward-cited patents (Data for backward-cited patents predating the year 2000 was not shown in the figure for the purpose of illustration. Citations to patents filed before 2000 comprise about 9% of the total citations.).

We used this list of backward citations to construct quarterly updated co-citation networks. Specifically, the nodes included in each quarterly network were the blockchain patents applied in a given quarter and the blockchain patents filed within a one-year window in the past. We set a time window in constructing our networks to avoid sparsity in the networks. Then, we defined an edge in each quarterly network if a pair of patent nodes cited at least one common patent publication. Hence, the resulting co-citation networks represent the connections among patents in terms of their reliance on common underlying knowledge. We created two separate networks comprising each type of blockchain patent.

In Figure 4, we visualize the entire blockchain patent co-citation network observed during our sample period. The network was visualized using Gephi 0.10.0 with ForceAtlas2

layout, which simulates a balanced state of a physical system assuming repulsion and attraction between nodes based on the presence of edges between them [53]. Hence, the relatively well-connected nodes have a close distance from each other on a graph. Red-colored nodes denote cryptocurrency patents, while the blue-colored ones represent distributed ledger patents.

The co-citation network plot reveals a distinct separation between various types of patents and a high concentration of patents within the same type. The graph illustrates that two different types of blockchain patents can be distinguished only by analyzing the structural patterns of their co-citation network. This is noteworthy because the network representation of patents closely mirrors the classification of blockchain patents determined by in-depth, expert-curated keywords and patent categories.
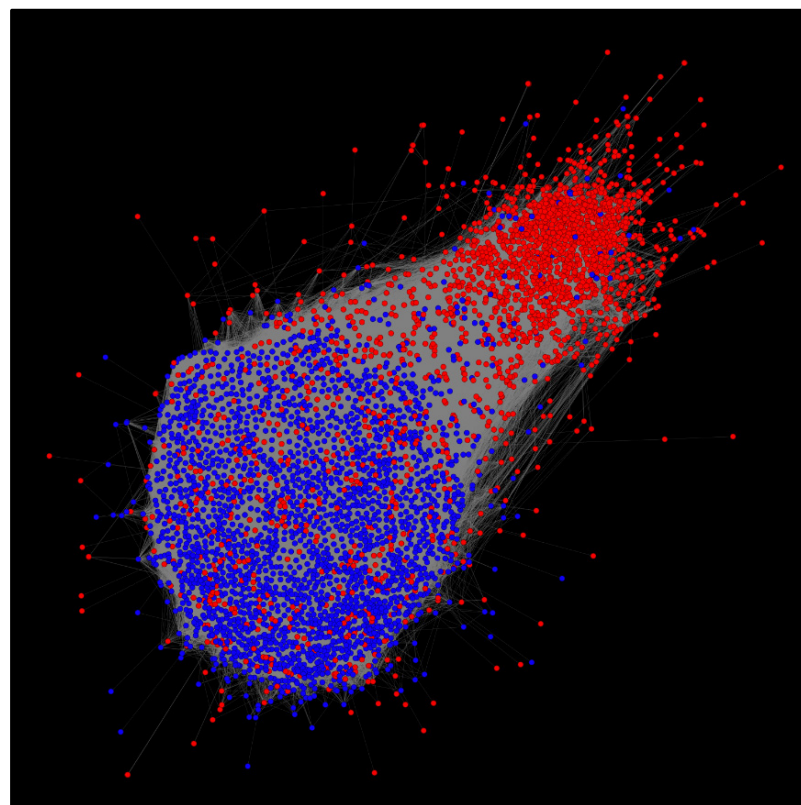


**Figure 4.** Co-citation network among blockchain patents (2008–2019).

With the co-citation network of cryptocurrency patents and distributed ledger patents, we computed two sets of measures, one at the network level and the other at the patent level, to be used in our co-citation analysis. We used the networkx package in Python to compute our measures. Since a sufficient number of distributed ledger patents is only available after 2015 (Figure 1b), we restricted our network analysis to the blockchain patents that were filed after 2015 to enable a comparison of two types of patents on the same timeline.

### 5.2. Network-Level Analysis

Using the quarterly updated co-citation networks described above, we computed three network-level measures, network density, transitivity, and degree assortativity, to examine how the system of blockchain patents has evolved over time. We illustrate below a detailed description of each measure, along with the presentation of the result of our analysis.

#### 5.2.1. Network Density

Network density refers to the degree to which the nodes in a network are directly connected to each other. Typically, network density is computed by calculating the ratio of

actual connections relative to the total number of possible connections among the nodes. Formally, network density is computed as follows:

$$Density = \frac{2m}{n(n-1)} \tag{1}$$

where *m* is the number of edges in the network and *n* is the number of nodes in the network.

In our context, a high level of density of the co-citation network implies that blockchain patents in a given period overall are highly related to each other by sharing common underlying knowledge or backward citation patents. In contrast, network density would be low if they do not share common citation sources, further indicating that they rely on disparate sets of underlying knowledge.

Figure 5 illustrates the changes in the density of our co-citation networks by patent types. The overall density of the co-citation network comprising cryptocurrency patents is much lower than the distributed ledger patents counterpart. Furthermore, the density of the cryptocurrency network does not show much variation across time. This indicates that, throughout the years, cryptocurrency patents have been sparsely connected to others in terms of citing common patents. In contrast, the density of distributed ledger patent networks was high in its early period of emergence. However, the density has drastically decreased in recent years. It is possible to speculate that, in recent years, the distributed ledger technology domain has not only experienced significant growth in the volume of patenting activities but the underlying knowledge of the patents has also become more diversified as well.
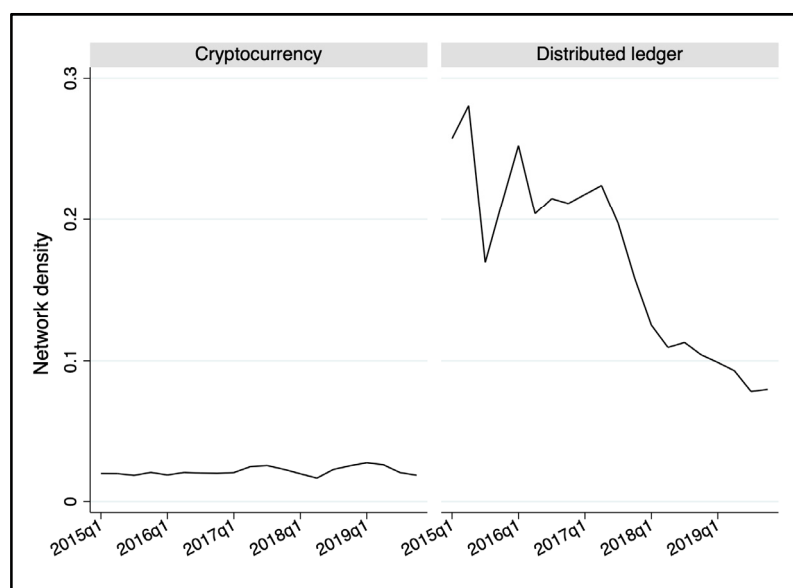


**Figure 5.** Density of co-citation networks by patent type.
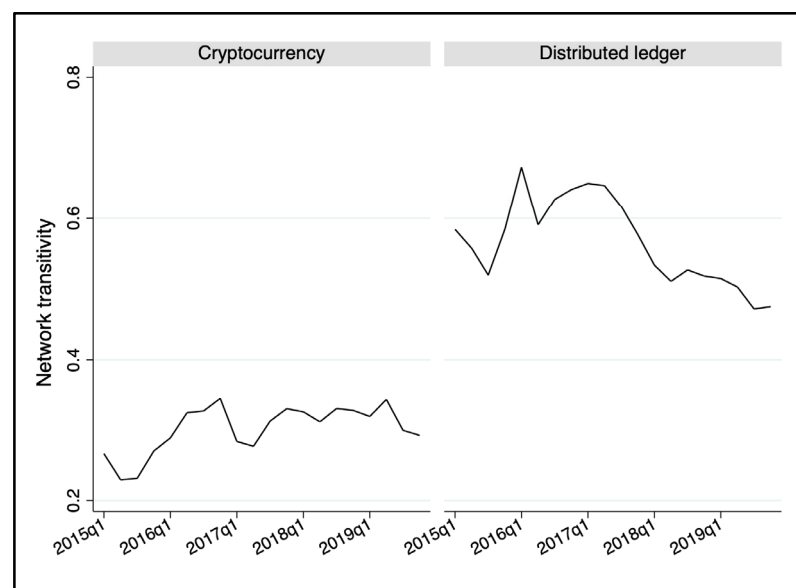
5.2.2. Network Transitivity

Network transitivity reflects the level of interconnectedness among the nodes. Network transitivity often is correlated with network density, but they are distinct from each other. Network transitivity considers not only the direct connections between pairs of nodes but also the indirect ties established between them through sharing a common intermediary. Hence, networks of the same density may still have different levels of transitivity depending on the presence of tightly connected clusters of nodes. Network transitivity is typically computed as the ratio of the realized triads to the possible triads observed in a network. The formula for network transitivity can be written as:

$$Transitivity = 3\frac{\#\ of\ triangles}{\#\ of\ triads} \tag{2}$$

where *triangle* is a set of three nodes with an edge between every pair and *triad* is a set of two edges sharing a common node.

In the context of a blockchain patent co-citation network, high network transitivity indicates that there is a high likelihood that two patents sharing common underlying knowledge will also have many other patents that also share common underlying knowledge with the two.

Figure 6 plots the transitivity of our co-citation networks over time. We find patterns similar to the cases of network density. The transitivity of cryptocurrency networks is generally lower, with relatively fewer fluctuations over time than those of distributed ledger patent networks. The distributed ledger network displays a high level of transitivity in the early period, which steeply declines in the later years. The pattern shows that distributed ledger patents were highly interdependent on common underlying knowledge in the initial phase, but over time, they have evolved to be reliant on a more dispersed set of knowledge.



**Figure 6.** Transitivity of co-citation networks by patent type.

5.2.3. Degree Assortativity

Degree assortativity represents the extent to which nodes of similar degrees connect with each other. It is often computed with the Pearson correlation coefficient between the degrees of nodes and the degrees of other nodes to which they are connected. The formula for degree assortativity is:

$$Degree\ assortativity = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2} \tag{3}$$

where $M$ is the total number of edges in the network and $j_i$ and $k_i$ are the degrees of nodes at the ends of the $i$th edge.

The degree assortativity measures the correlation between the degrees of nodes that are connected to each other, with a value ranging from $-1$ to 1. A higher value indicates that high-degree nodes tend to connect with other high-degree nodes, while low-degree nodes tend to connect with other low-degree nodes. Conversely, a lower value indicates that high-degree nodes tend to connect with low-degree nodes, and vice versa. A highly assortative network in our context corresponds to networks of blockchain patents, where a highly central patent tends to share common underlying knowledge with other central patents rather than with peripheral patents.

Figure 7 illustrates the degree assortativity of our co-citation networks. For the cryptocurrency network, we observe large fluctuations in the values across time, further obfuscating a clear interpretation of generalizable trends in the networks. This is perhaps due to the inherent instability of emerging technology. For cryptocurrency patents, the patterns of co-citation change drastically from period to period: in some periods, central patents tend to draw on similar sets of underlying knowledge, while in some periods, central patents and peripheral patents are more prone to rely on the same set of previous knowledge. For the distributed ledger network, we find fluctuations in the early period, which we speculate to be originating from the low volume of distributed ledger patents in those periods. However, from 2016 onwards, the network gradually becomes more assortative.
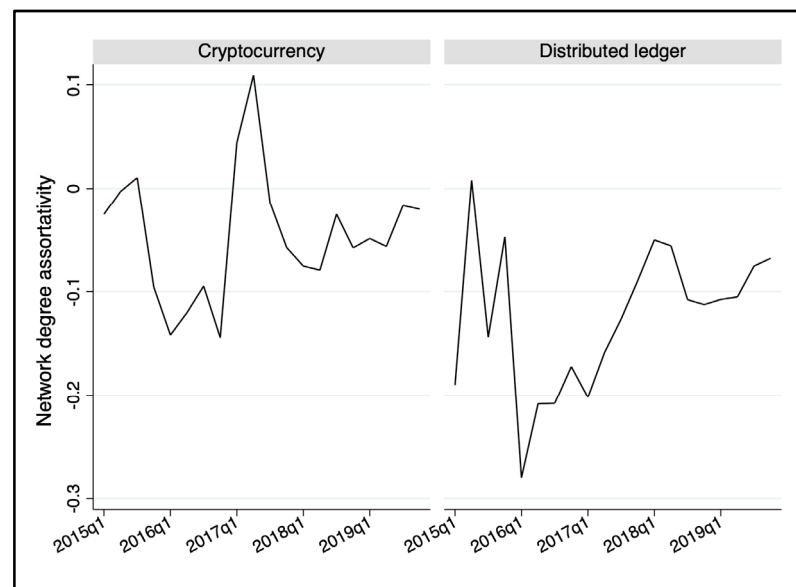


**Figure 7.** Degree assortativity of co-citation networks by patent type.

*5.3. Patent-Level Analysis*

We operationalized three patent-level network measures, degree centrality, eigenvector centrality, and coreness, to capture sample patents' positions in the co-citation networks. Detailed descriptions of the measures and our analysis are presented below.

5.3.1. Degree Centrality

Degree centrality captures a node's level of connections to others in a network, which can be computed as below:

$$Degree\ centrality_i = \frac{\max\left(d_i\right)}{n-1} \tag{4}$$

where $d_i$ is the degree of node $i$ and $n$ is the number of nodes in the network.

In our context, it counts the number of other patents that also cite a given patent's backward citations. We further normalized the degree centrality by dividing the maximum degree centrality observed in the network. A high degree centrality indicates that the focal patent draws on underlying knowledge that is well shared with many others.

We examined how the overall level of degree centrality of patents varies across time. We visualize the temporal trends by plotting the polynomial fit between the degree centrality and the dates of the patents in Figure 8a. The average level of degree centrality is much lower for cryptocurrency patents than the distributed ledger counterparts and varies little across time. In contrast, the average degree centrality of distributed ledger patents was initially high, with a steep decrease in value in recent years. The plot closely follows the pattern of network density. The distributed ledger patents had a high level of connections

with each other in the early period, but they became more disconnected in terms of sharing common underlying knowledge as the technology evolved over time.



(**a**)                                                                      (**b**)
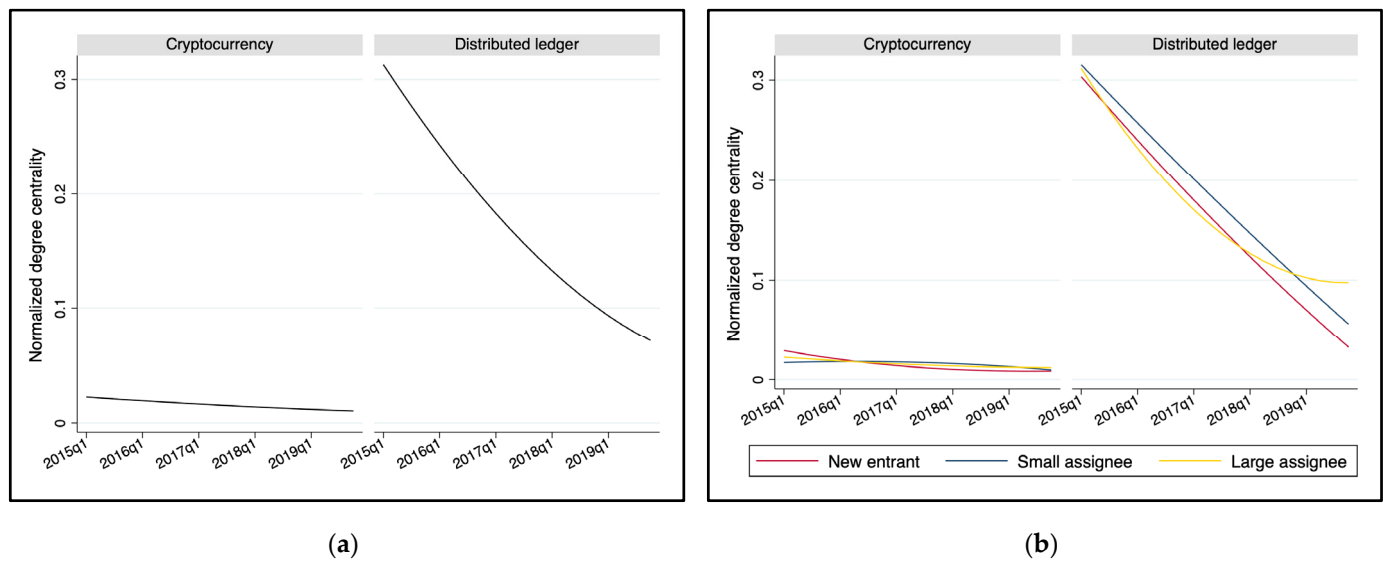
**Figure 8.** (**a**) Degree centrality of blockchain patents by patent type. (**b**) Degree centrality of blockchain patents by patent and assignee type.

One of the benefits of examining patent-level network measures is that we can capture the variation in network positions of each patent and further check whether this variation is related to other patent-level attributes. We sought to check whether different types of assignees tend to invent patents that are more or less central in the network. Figure 8b illustrates the same polynomial fit between degree centrality and time by three different types of assignees. Unfortunately, we do not find significant variations across assignees.
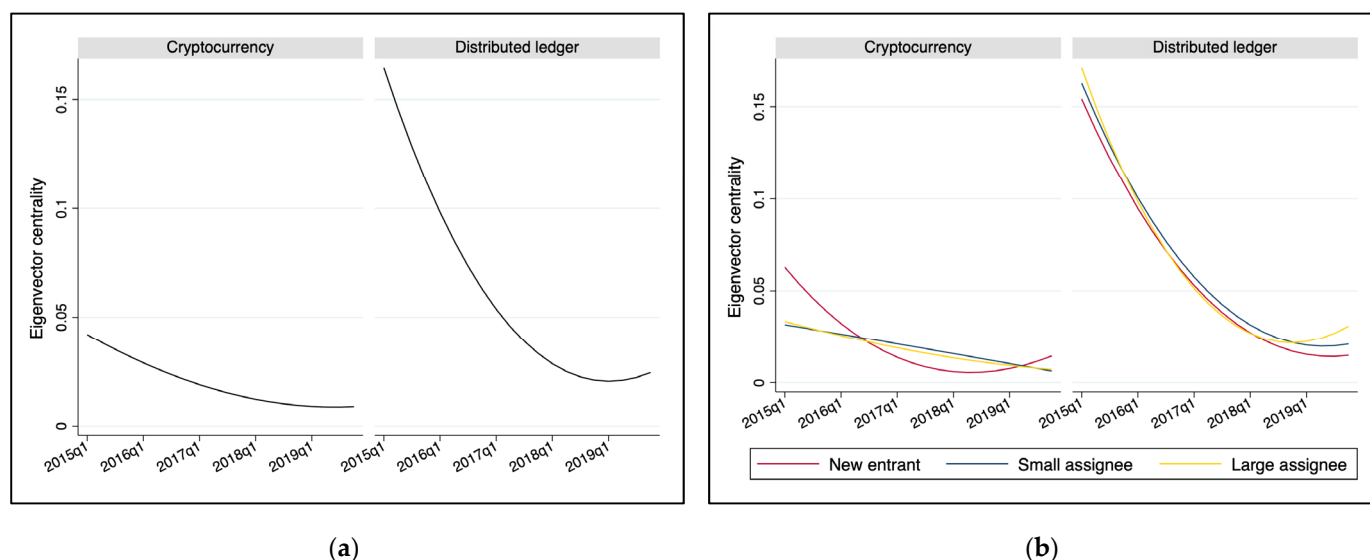
5.3.2. Eigenvector Centrality

Eigenvector centrality reflects a node's influence in the network. Formally,

$$Eigenvector\ centrality_i = \frac{1}{\lambda}\sum_j a_{ij}v_j, \tag{5}$$

where $\lambda$ is the eigenvalue, $a_{ij}$ is an element of the adjacency matrix of the network, and $v_j$ is the neighbor of node $i$.

It is a variant of network centrality measure that additionally considers the quality of a node's connections. The eigenvector is computed by recursively calculating greater weights to be assigned to a node's connections if the node's neighbors are also well connected to their neighbors [54]. The normalized eigenvector centrality ranges from 0 to 1, with a high value capturing, in our context, the degree of influence or importance of the patent in terms of its underlying knowledge.

Figure 9a displays the polynomial fit between eigenvector centrality and time of the sample blockchain patents. We find patterns similar to those of degree centrality. The average eigenvector centrality of cryptocurrency patents is lower than that of distributed ledger patents and has remained low over time. In contrast, the eigenvector centrality of distributed ledge patents declined drastically in recent years. This indicates that distributed ledger patents were prone to rely on key influential underlying knowledge in the early years, but they have become more reliant on disparate sources of knowledge in recent years. Figure 9b plots the same polynomial fit by different types of assignees, but the patterns were not distinguishable across assignee types.

(**a**)                                                    (**b**)

**Figure 9.** (**a**) Eigenvector centrality of blockchain patents by patent type. (**b**) Eigenvector centrality of blockchain patents by patent and assignee type.

### 5.3.3. Coreness

Lastly, we computed the coreness of each patent in this network. Coreness is a measure of its position in the core, a tightly interconnected region of a network. Assuming a global network structure comprises a core, a densely connected component, and a periphery, a sparsely connected component, network researchers attempted to assign each node to each category [55,56]. Several algorithms have been proposed to calculate a continuous score of coreness based on a node's position in the core-periphery spectrum [55,57,58]. In this paper, we implemented an algorithm developed by Rombach et al. [58] to compute the continuous measure of the coreness, ranging from 0 to 1, of each blockchain patent. In our context, patents with a high coreness are those that are positioned in tightly interconnected neighbors of patents that rely on common sets of underlying knowledge.

In Figure 10a, we visualize the polynomial fit between the coreness and the date of sample patents. The average coreness of cryptocurrency has decreased over time, implying that the patents belonged to the densely interconnected clusters in the early years but have begun to occupy more sparsely connected regions of the networks. For distributed ledger patents, a similar pattern followed in the early years, but it has reversed in recent years with a slight increase in the overall coreness of the patents. This implies that, when measured with coreness, distributed ledger patents seem to have formed densely interconnected clusters in recent years in terms of their reliance on common underlying knowledge.

We further investigated whether this pattern is more or less pronounced for different assignee types by plotting the polynomial fit by new entrants, small assignees, and large assignees in Figure 10b. The plot indicates that large assignees in recent years have invented distributed ledger patents that are positioned in the core areas of the network, while new and small assignees filed patents that belonged to the more peripheral areas in the network. The graph illustrates that the system of distributed ledger patents has recently evolved into a core-periphery structure, where large players invent patents in the core, which rely on prior underlying knowledge that is well shared with other patents, and new and small players develop patents in the peripheral region.

(**a**)                                                                                 (**b**)
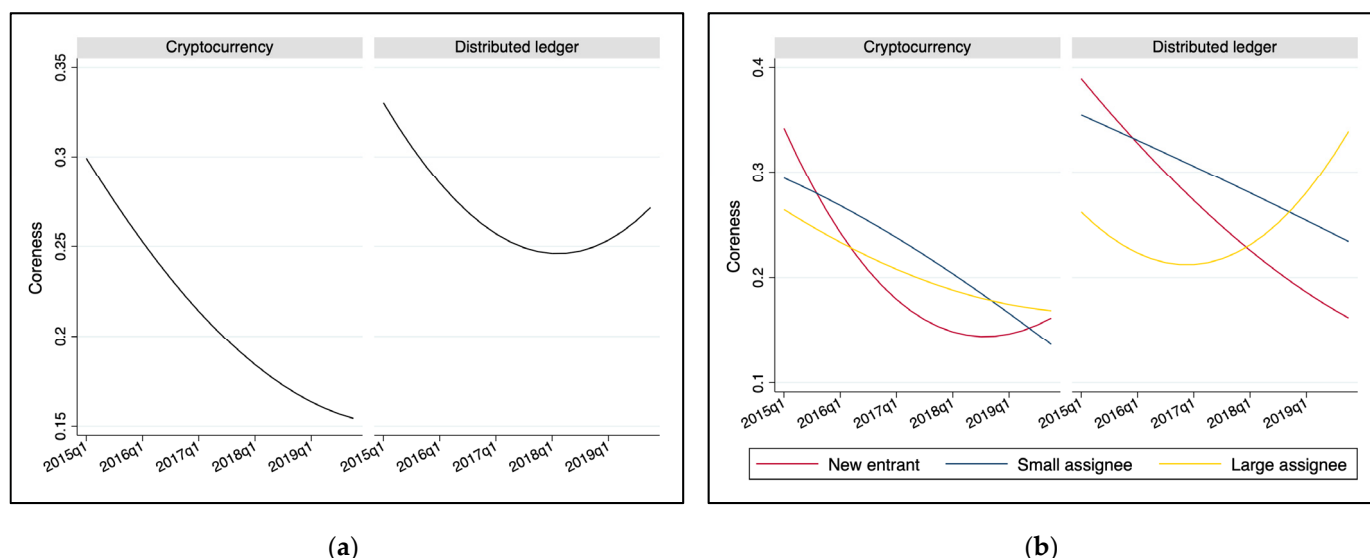
**Figure 10.** (**a**) Coreness of blockchain patents by patent type. (**b**) Coreness of blockchain patents by patent and assignee type.

## 6. Semantic Similarity Analysis

Blockchain technology is a rapidly developing field, with new patents being filed regularly. It is important to go beyond the traditional approach to understand the current state of the field and identify key players. While patent co-citation analysis can provide insights into how knowledge flows within the field and how different types of patents are structurally related, it does not provide a detailed understanding of the similarities and differences in the actual contents of the patents. Therefore, it is beneficial to combine this approach with other methods, such as NLP-based analysis, to gain a more comprehensive understanding of the underlying knowledge structure of blockchain technology.

One advantage of using BERT-based semantic analysis over co-citation analysis for blockchain patent analysis is that it provides a more detailed understanding of the similarities and differences in the contents of the patents. BERT-based semantic analysis can also be used to analyze the patent abstracts and identify the key ideas and concepts that are discussed in the patents to understand the key ideas and concepts that are driving the development of the technology and identify potential areas for future research.

Additionally, BERT-based semantic analysis allows researchers to analyze the relative positions of patents in the semantic space. The semantic space refers to the representation of the meaning or content of patents in a numerical form, which, in this case, are the vectors generated by the BERT model. By reducing the dimensionality of these vectors using techniques such as uniform manifold approximation and projection (UMAP), it is possible to project them onto a two-dimensional graph, which can be visualized and analyzed. This approach allows for a detailed understanding of the similarities and differences in the underlying knowledge structure of patents and can be useful for identifying key concepts, trends, and connections within the blockchain patent landscape.

Importantly, by analyzing the relative positions of patents in the semantic space, it is also possible to identify which patents are central to the field and which are on the periphery. Central patents are those that are closely related to many other patents in the field and have a high degree of similarity in terms of their underlying knowledge, while patents on the periphery are less related to other patents in the field and have a lower degree of similarity in terms of their underlying knowledge.

Overall, the use of an NLP-based approach, such as BERT, in combination with co-citation analysis can provide a more comprehensive understanding of the underlying knowledge structure of blockchain technology. It can provide insights into the relationships and connections between different patents and technologies, as well as a detailed understanding of the similarities and differences in the contents of the patents. This can

be useful for researchers and practitioners in the field to understand the state of the art, identify potential areas for future research, and follow the evolution of the technology.

### 6.1. PatentSBERTa

In a recent study, Bekamiri et al. [59] proposed a new method for classifying patents that combines K-nearest neighbors (KNN) (K-nearest neighbors (KNN) is a traditional machine learning algorithm that is used for classification tasks. It works by comparing an input sample to the k closest samples in the training set and then classifying the input sample based on the majority class of those k closest samples. It is a simple algorithm that is easy to understand and interpret, but it can be computationally expensive for large datasets.) and Sentence-BERT (SBERT) (Sentence-BERT (SBERT) is a pretrained transformer model that is designed to generate semantically meaningful sentence embeddings. It is based on the BERT model, but it is trained on a large corpus of sentence-pair data. SBERT uses cosine similarity to compare sentences and find the most similar pair in a collection of sentences, and it can be fine-tuned on specific domains, such as patent data, to make it more effective at understanding that domain-specific language. This method is more efficient than BERT and RoBERTa, which can take up to 65 h to find the most similar pair in a collection of 10,000 sentences; SBERT can perform it in just 5 s.) to achieve higher accuracy and efficiency than previous methods. The study proposed an augmented version of SBERT, which is fine-tuned to the domain of patent claims to increase its performance. The study also showed that fine-tuning SBERT to domain-specific language in textual patent data could improve the performance of the model even without labeled examples, making the process faster and more cost-effective [59].

The study's approach is to first use transformer models to understand the text in the patent claims and create a numerical representation of it called embeddings. These embeddings capture the meaning of the text in a way that can be compared to other embeddings. The study then uses the KNN algorithm to classify the patent claims by comparing the embedding of the patent claim in question to the embeddings of the closest k patents in the training dataset. The algorithm then assigns the patent claim to the most common category among the closest k patents. This approach is different from the traditional approach of using metadata, keywords, or citation information to classify patents [59].

The study utilized a dataset of 1,492,294 patents from 2013 to 2017, and 8% of the patents were used as a test dataset to evaluate the model's performance. The proposed framework predicts individual input patent classes and subclasses by finding the top k semantic similarity patents. The study used transformer models based on augmented SBERT and RoBERTa and used a different approach to predict patent classification by finding top k similar patent claims and using the KNN algorithm to predict the patent class or subclass [59].

### 6.2. Vector Representation and Visualization of Semantic Analysis Using PatentSBERTa

We conducted an in-depth analysis of the content of the sample blockchain patents reflected in their abstracts. To do so, we utilized PatentSBERTa, introduced above. We transformed the texts of the sample patents into 768-dimensional vectors. We visualized the positions of the blockchain patents represented in the semantic space. To do so, we transformed the 768-dimensional vectors into 2-dimensional vectors UMAP, a dimension reduction technique often used in transforming high-dimensional vectors [60].

Figure 11 represents the plot of the sample patents mapped onto the two-dimensional coordinates. Red crosses represent cryptocurrency patents, while the blue dots show the positions of the distributed ledger patents. In the plot, we observe that patents of the same type tend to cluster together, while different types of patents are more distanced from each other. To statistically test this observation, we computed all pairwise cosine distances among 768-dimensional vectors of sample patents. The average cosine distance between the vector representations of patent texts of different types was significantly greater than

those of the same types (*t*-test, *p*-value < 0.001). The visual inspection and the simple statistical test demonstrate BERT's ability to discern the difference between the two types of blockchain patents by only using their textual information. This is especially surprising because these two types of patents fall under the umbrella of blockchain technology and the categorization of these patents often needs close examinations and validations from subject-matter experts [49].
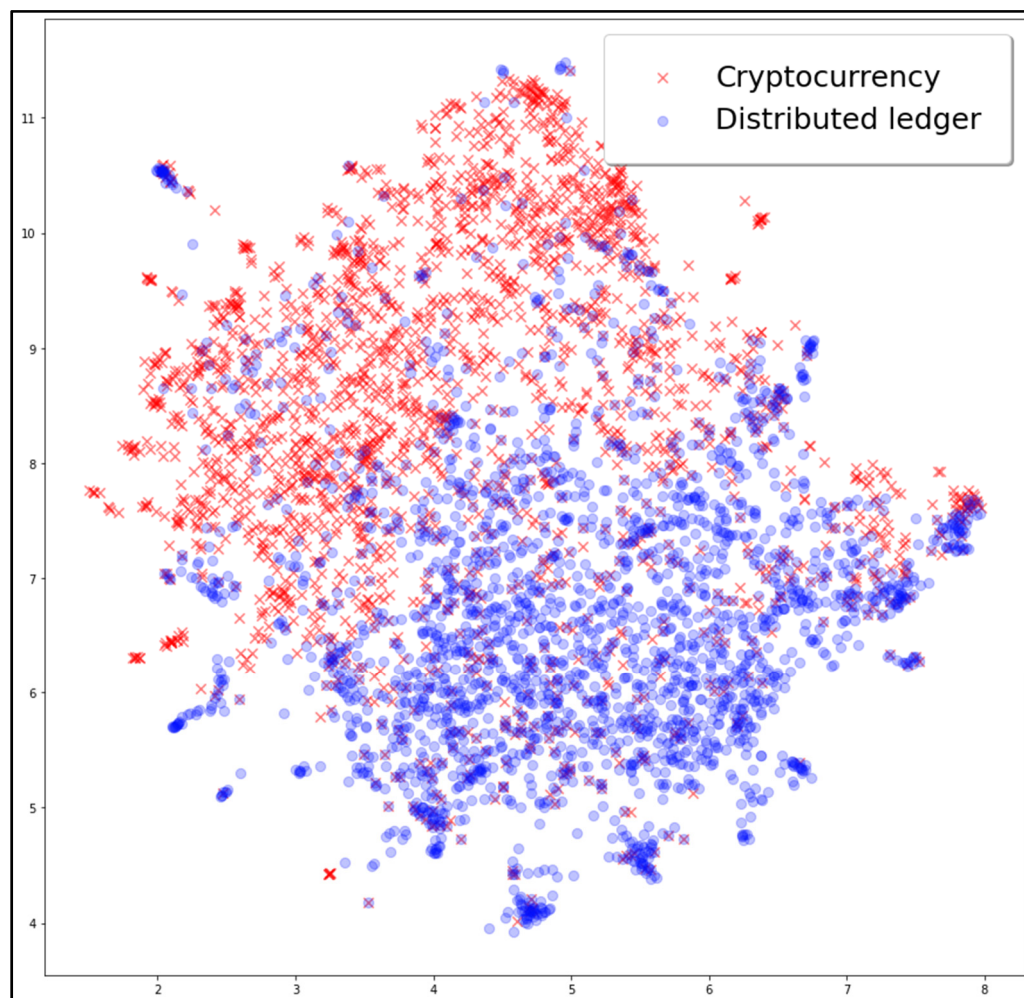


**Figure 11.** Two-dimensional representation of patent abstracts using BERT and UMAP.

*6.3. System-Level Analysis*

We used the BERT's vector representations of patent abstracts to examine how the system of blockchain patents has evolved. Specifically, we quantified how the contents of the patents filed in each time period closely resemble or deviate from the contents of the patents that were filed in the past. In so doing, we aimed to excavate whether the blockchain patents have evolved over time towards a convergence (more similar to the ones in the past) or a divergence (more dissimilar to the ones in the past) in technology. We detail the two steps we took in this analysis.

First, for each time period, we computed a centroid, the average of the vector representations, of the patents that were filed in the recent past. Since each patent's vector representation denotes its position in the semantic space, the numeric aggregation of these vector representations, through averaging them, provides the center point in the semantic space that is densely populated by the patents. Hence, a centroid of patents' vector representations corresponds to a hypothetical vector that captures the popular features of patents that were recently filed in a given period. Specifically, for each quarter in the

sample period, we computed a centroid of vector representations of the patents filed within the past three years.

Figure 12 provides the visualization of the centroids across sample periods. The red line with crosses follows the centroids of cryptocurrency patents, while the blue line with dots represents the distributed ledger counterparts. The absolute coordinates of centroids alone do not yield substantive interpretations. Still, the movements of the centroids show that the popular features in the contents of the patents have changed over time at different paces and in different directions.
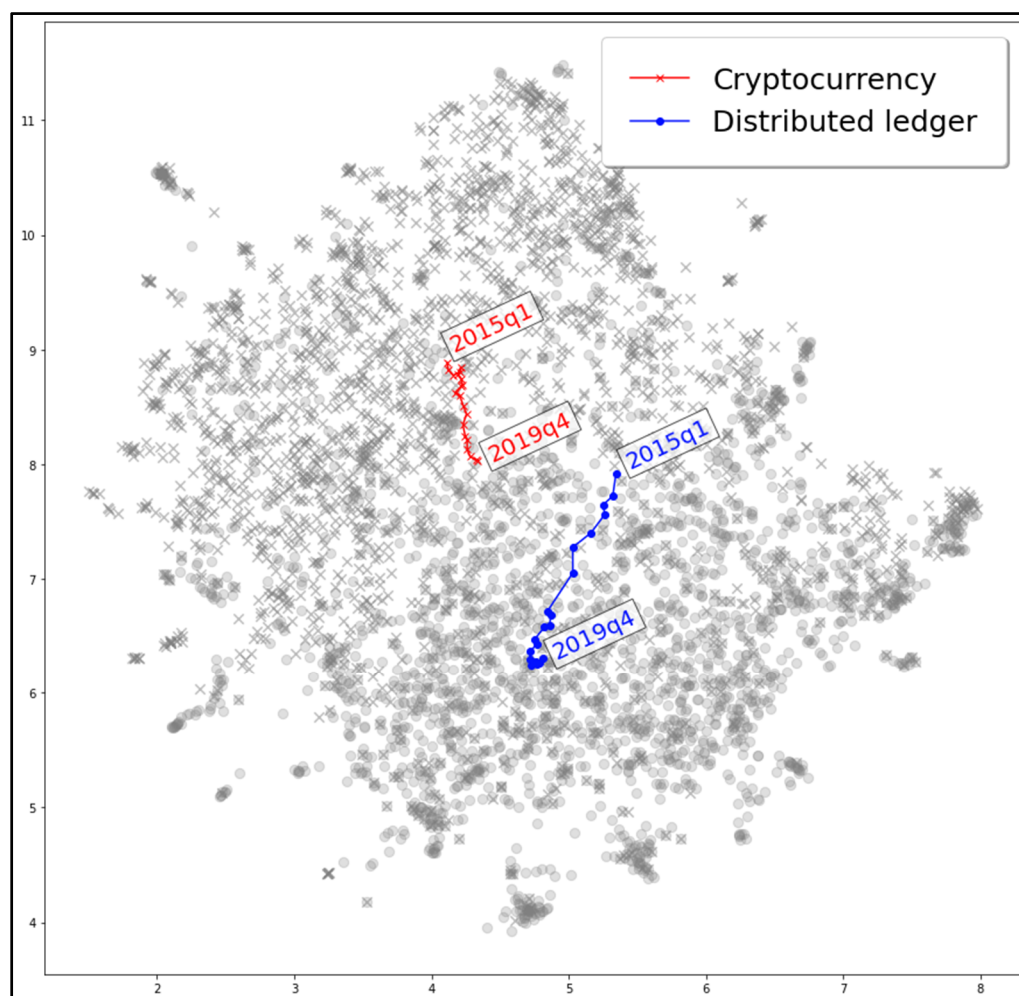


**Figure 12.** Movement of centroids of vector representations of patent abstracts across time by type.

Second, we calculated the cosine distance between the vector representation of each patent in a given period with the centroid of the patents in the recent past. Hence, the distance captures the extent to which each patent's textual description of its technology in the abstract is dissimilar from the overall textual features of the patents of the recent past.

Figure 13 plots the polynomial fit between the patents' distance from the centroid and their dates. In general, the distance is higher for cryptocurrency patents. This implies that cryptocurrency patents have a more diverse set of knowledge reflected in their abstract texts in comparison to distributed ledger patents—an observation consistent with the result of the co-citation network analysis with a low level of network density. Additionally, the plot shows that the distance has been gradually increasing over the years. Patents filed in recent years tend to be more semantically dissimilar from the ones in the past, implying a trend of divergence in cryptocurrency technology.
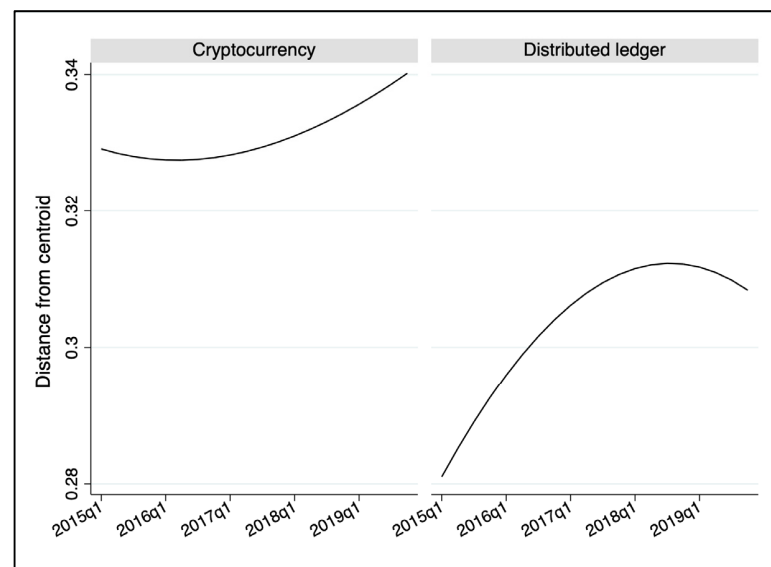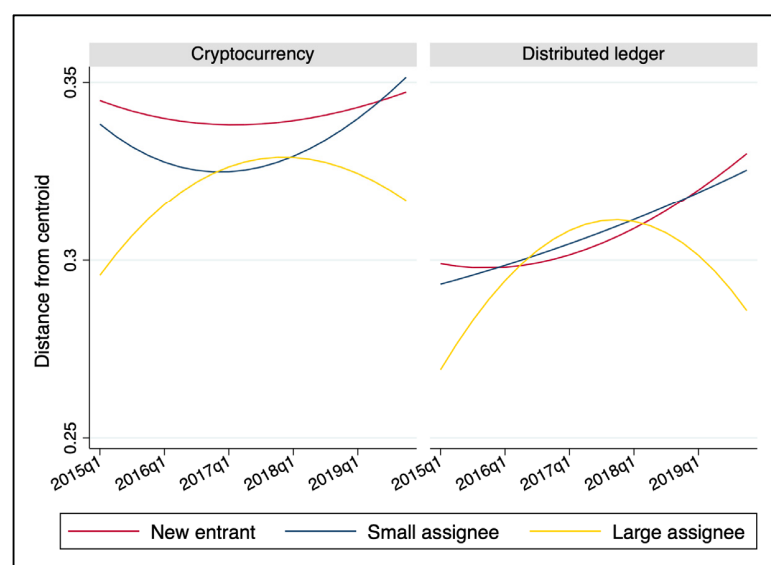
**Figure 13.** Distance from centroid by patent type.

For distributed ledger patents, the increase in the distance seems to be steeper, indicating a greater speed of divergence in this technological domain compared to the cryptocurrency counterpart. However, in the most recent two years, the rate of increase in the distance has decreased. This perhaps suggests a saturation of patents in the semantic space or a reversing trend toward convergence in distributed ledger technology in recent years.

*6.4. Patent-Level Analysis*

We further examined whether the distance from the centroid varies across different types of assignees. Figure 14 plots the same polynomial fit by different assignee types. In both technological domains, the patterns clearly illustrate that new and small assignees have been filing patents that are more dissimilar from previous patents. The large assignees, on the other hand, in more recent years, perhaps after the burst of the bubble, tend to develop patents that are less dissimilar to the preceding patents. Our patent-level analysis of vector representations of patents using BERT shows that different types of players active in the blockchain technology domain engage in patenting activities with different levels of innovations.
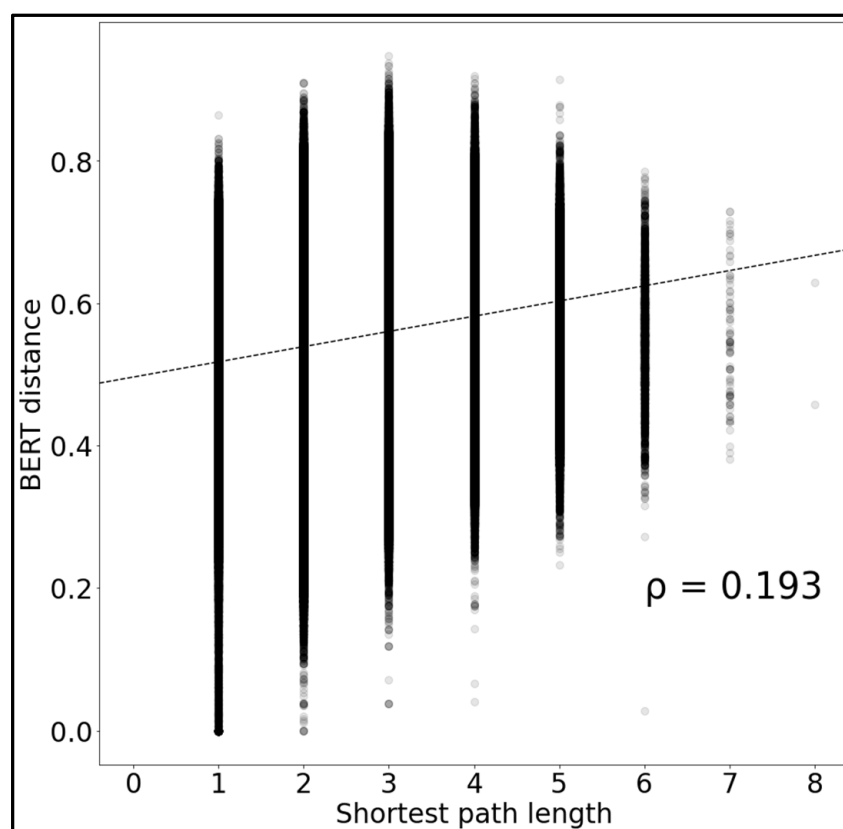


**Figure 14.** Distance from centroid by patent and assignee type.

### 7. Comparison of Co-Citation Network Analysis and Semantic Similarity Analysis

The co-citation network analysis and semantic similarity analysis both captured similar dynamics in the evolution of blockchain technology. We conducted one additional analysis to check how the two approaches relate to each other systematically. To do so, we compared the distance metrics obtained from the network space and semantic space. First, we computed the shortest path length between every pair of patents in the co-citation network. A path length of $i$ is assigned to a pair of patents if they can be reached in $i$th step to each other, with a path length of 1 indicating a direct connection (i.e., presence of a co-citation tie) between them. Second, we computed the cosine distance between every pair of patents using their vector representations obtained from PatentSBERTa.

Figure 15 illustrates the scatter plots of patents across the two distance metrics. The plot indicates that patents that are more distant in terms of their co-citation network ties are actually more dissimilar in the textual descriptions of their underlying technology. The correlation between the two metrics was 0.193. Hence, the semantic similarity approach, to some extent, can capture the characteristics of patents that emerge from their structural relationships based on citations. At the same time, there are variations in the patents that the semantic similarity approach can uniquely capture as well. For instance, among the pairs of patents with a path length of one, we still observe a lot of variation in their semantic similarity. The semantic similarity approach using BERT proves to be useful in capturing more variations in patents beyond their positions in the co-citation network structures.



**Figure 15.** Scatterplot of distances in network and semantic spaces.

The co-citation network analysis and semantic similarity analysis are two widely used methods for analyzing the evolution of technological fields. Both methods have been shown to be effective in capturing the dynamics of the development of various technologies. In the case of blockchain technology, our results showed that both the co-citation network analysis and semantic similarity analysis captured similar dynamics.

Given the complementary nature of these two methods, we conducted a systematic evaluation of their relationship to determine the extent to which they capture similar or different aspects of the evolution of blockchain technology. To achieve this, we conducted a comparative analysis by comparing the distance metrics obtained from the network space and semantic space. The rationale behind this comparative analysis is to understand the complementarity of the two approaches and determine the extent to which they capture unique aspects of the evolution of blockchain technology [61–63]. By comparing the results obtained from the two methods, we aimed to gain a more comprehensive understanding of the dynamics of the development of blockchain technology.

Our results showed that the semantic similarity approach, to some extent, captured the characteristics of patents that emerged from their structural relationships based on citations. At the same time, the semantic similarity approach was able to capture unique variations in the patents beyond their positions in the co-citation network structures. Accordingly, our comparative analysis provides valuable insights into the relationship between the co-citation network analysis and semantic similarity analysis and highlights the importance of combining multiple approaches for a comprehensive analysis of the evolution of technological fields.

## 8. Discussion and Conclusions

In this study, we aimed to depict the knowledge landscape of the blockchain technological domain through an examination of blockchain technology patent data and citation data. Our focus was on understanding the evolution of relationships among various knowledge areas that constitute the technological space, as well as examining these subdomains based on their respective owners (assignees). Additionally, we introduced a novel, state-of-the-art NLP-based approach for extracting and analyzing the technological knowledge embedded in the textual descriptions of patents within the blockchain technology field.

Our analysis showed that the number of patents related to blockchain technology has been increasing in recent years. This increase is not only related to the specific use case of cryptocurrency but also to the underlying technology of distributed ledgers. The increase in volume was driven by the inflow of new entrants into the field and an increase in the patenting activities of small assignees. Furthermore, the increase in patent filings for distributed ledgers was more consistent, indicating a more sustained expansion in this subdomain. This suggests that as the technology of distributed ledgers becomes more widely understood and adopted, companies and public organizations are beginning to see more potential uses for the technology beyond just cryptocurrency.

To further explore the knowledge landscape of the blockchain technology domain, we used the co-citation network to trace relationships between patents based on their reliance on common citation sources. We analyzed the evolving patterns of the blockchain technology domain using various network-level metrics, such as density, transitivity, and assortativity. Additionally, we used patent-level network measures, such as degree centrality, eigenvector centrality, and coreness, to understand how different groups of assignees invent patents that occupy different structural positions in the network over time.

While co-citation network analysis has traditionally been used to study patent data and identify the relationships between different technologies, it has limitations, particularly when it comes to analyzing the content of patent documents. To address these limitations, we conducted an additional semantic analysis using our NLP-based approach to analyze the textual content of patent documents and extract the underlying knowledge and relationships within the blockchain technology field. Specifically, we introduced the patent-specific NLP algorithm PatentSBERTa, which is a variant of the Sentence-BERT model that is specifically trained with patent data for the purpose of calculating patent distance and classification tasks. Based on PatentSBERTa, we calculated the cosine distance among patents to understand whether the contents of the blockchain technology have evolved to become more dissimilar from the past. Finally, we compared the results of our NLP-based approach to those obtained from traditional co-citation network analysis

and evaluated the effectiveness of this approach in understanding patents beyond their structural positions in the network.

Our results demonstrate the advantages of using an NLP-based approach for analyzing technological knowledge and relationships within the blockchain technology field. Our findings show that the field of blockchain technology is expanding and diverging, with increasing patent filings in both cryptocurrency and distributed ledger technologies, and that there is a growing knowledge similarity between the two subdomains. Additionally, we also found that the ways in which patent assignees engage in innovative activities in the blockchain technology domain vary by their relative prior experience in this domain. Our study can inform future research and guide industry and policy decisions related to blockchain technology by providing insights into the current state and future direction of the blockchain technology industry and identifying the key players and the most influential patents in the field.

The findings of this study provide valuable insights into the blockchain technology domain. However, it is important to acknowledge some limitations of our study. Our analysis was limited to the examination of patent data and did not take into account other sources of information such as technical papers, news articles, and trade publications. Although we used patent data to categorize blockchain technologies into three categories— cryptocurrency technology, distributed ledger technology, and smart contract technology— it may not align with the understanding of business practitioners. Patent data serve as a useful source of information, but they may not necessarily reflect the actual implementation and adoption of these technologies in the industry. This aspect should be kept in mind when interpreting the results of our study. In addition, the sample of patent data used in this study was limited to a specific period and location, which may not accurately reflect the entire blockchain technology landscape. The NLP-based approach used in this study also has limitations in accurately capturing and representing the complex relationships between patents and knowledge areas.

To build on the findings of this study, future research could consider incorporating additional sources of data and exploring alternative methods of measuring the similarity and relationships between patents. Additionally, research could be conducted to improve the NLP-based approach used in this study by incorporating state-of-the-art models such as transformer-based models.

## References

1. Sadawi, A.A.; Madani, B.; Saboor, S.; Ndiaye, M.; Abu-Lebdeh, G. A Comprehensive Hierarchical Blockchain System for Carbon Emission Trading Utilizing Blockchain of Things and Smart Contract. *Technol. Forecast. Soc. Chang.* **2021**, *173*, 121124. [CrossRef]
2. Omar, I.A.; Hasan, H.R.; Jayaraman, R.; Salah, K.; Omar, M. Implementing Decentralized Auctions Using Blockchain Smart Contracts. *Technol. Forecast. Soc. Chang.* **2021**, *168*, 120786. [CrossRef]
3. Chang, S.E.; Chen, Y.-C.; Lu, M.-F. Supply Chain Re-Engineering Using Blockchain Technology: A Case of Smart Contract Based Tracking Process. *Technol. Forecast. Soc. Chang.* **2019**, *144*, 1–11. [CrossRef]

4. Hou, J.; Wang, C.; Luo, S. How to Improve the Competiveness of Distributed Energy Resources in China with Blockchain Technology. *Technol. Forecast. Soc. Chang.* **2020**, *151*, 119744. [CrossRef]

5. Kowalski, M.; Lee, Z.W.Y.; Chan, T.K.H. Blockchain Technology and Trust Relationships in Trade Finance. *Technol. Forecast. Soc. Chang.* **2021**, *166*, 120641. [CrossRef]

6. Puthal, D.; Malik, N.; Mohanty, S.P.; Kougianos, E.; Yang, C. The Blockchain as a Decentralized Security Framework [Future Directions]. *IEEE Consum. Electron. Mag.* **2018**, *7*, 18–21. [CrossRef]

7. Köhler, S.; Bager, S.; Pizzol, M. Sustainability Standards and Blockchain in Agro-Food Supply Chains: Synergies and Conflicts. *Technol. Forecast. Soc. Chang.* **2022**, *185*, 122094. [CrossRef]

8. Agbo, C.C.; Mahmoud, Q.H.; Eklund, J.M. Blockchain Technology in Healthcare: A Systematic Review. *Healthcare* **2019**, *7*, 56. [CrossRef]

9. Tandon, A.; Kaur, P.; Mäntymäki, M.; Dhir, A. Blockchain Applications in Management: A Bibliometric Analysis and Literature Review. *Technol. Forecast. Soc. Chang.* **2021**, *166*, 120649. [CrossRef]

10. Balasubramanian, S.; Shukla, V.; Sethi, J.S.; Islam, N.; Saloum, R. A Readiness Assessment Framework for Blockchain Adoption: A Healthcare Case Study. *Technol. Forecast. Soc. Chang.* **2021**, *165*, 120536. [CrossRef]

11. Abdel-Basset, M.; Chang, V.; Nabeeh, N.A. An Intelligent Framework Using Disruptive Technologies for COVID-19 Analysis. *Technol. Forecast. Soc. Chang.* **2021**, *163*, 120431. [CrossRef]

12. Sundarakani, B.; Ajaykumar, A.; Gunasekaran, A. Big Data Driven Supply Chain Design and Applications for Blockchain: An Action Research Using Case Study Approach. *Omega* **2021**, *102*, 102452. [CrossRef]

13. Lu, Y. The Blockchain: State-of-the-Art and Research Challenges. *J. Ind. Inf. Integr.* **2019**, *15*, 80–90. [CrossRef]

14. Barirani, A.; Agard, B.; Beaudry, C. Discovering and Assessing Fields of Expertise in Nanomedicine: A Patent Co-Citation Network Perspective. *Scientometrics* **2013**, *94*, 1111–1136. [CrossRef]

15. Wang, X.; Duan, Y. Identifying Core Technology Structure of Electric Vehicle Industry through Patent Co-Citation Information. *Energy Procedia* **2011**, *5*, 2581–2585. [CrossRef]

16. Castriotta, M.; Di Guardo, M.C. Disentangling the Automotive Technology Structure: A Patent Co-Citation Analysis. *Scientometrics* **2016**, *107*, 819–837. [CrossRef]

17. Smojver, V.; Štorga, M.; Zovak, G. Exploring Knowledge Flow within a Technology Domain by Conducting a Dynamic Analysis of a Patent Co-Citation Network. *J. Knowl. Manag.* **2021**, *25*, 433–453. [CrossRef]

18. Garg, P.; Gupta, B.; Chauhan, A.K.; Sivarajah, U.; Gupta, S.; Modgil, S. Measuring the Perceived Benefits of Implementing Blockchain Technology in the Banking Sector. *Technol. Forecast. Soc. Chang.* **2021**, *163*, 120407. [CrossRef]

19. Chang, V.; Baudier, P.; Zhang, H.; Xu, Q.; Zhang, J.; Arami, M. How Blockchain Can Impact Financial Services—The Overview, Challenges and Recommendations from Expert Interviewees. *Technol. Forecast. Soc. Chang.* **2020**, *158*, 120166. [CrossRef]

20. Liu, J.; Li, X.; Wang, S. What Have We Learnt from 10 Years of Fintech Research? A Scientometric Analysis. *Technol. Forecast. Soc. Chang.* **2020**, *155*, 120022. [CrossRef]

21. Ali, O.; Momin, M.; Shrestha, A.; Das, R.; Alhajj, F.; Dwivedi, Y.K. A Review of the Key Challenges of Non-Fungible Tokens. *Technol. Forecast. Soc. Chang.* **2023**, *187*, 122248. [CrossRef]

22. Sarker, I.; Datta, B. Re-Designing the Pension Business Processes for Achieving Technology-Driven Reforms through Blockchain Adoption: A Proposed Architecture. *Technol. Forecast. Soc. Chang.* **2022**, *174*, 121059. [CrossRef]

23. Wan, Y.; Gao, Y.; Hu, Y. Blockchain Application and Collaborative Innovation in the Manufacturing Industry: Based on the Perspective of Social Trust. *Technol. Forecast. Soc. Chang.* **2022**, *177*, 121540. [CrossRef]

24. Rodríguez-Espíndola, O.; Chowdhury, S.; Dey, P.K.; Albores, P.; Emrouznejad, A. Analysis of the Adoption of Emergent Technologies for Risk Management in the Era of Digital Manufacturing. *Technol. Forecast. Soc. Chang.* **2022**, *178*, 121562. [CrossRef]

25. Kamble, S.S.; Gunasekaran, A.; Kumar, V.; Belhadi, A.; Foropon, C. A Machine Learning Based Approach for Predicting Blockchain Adoption in Supply Chain. *Technol. Forecast. Soc. Chang.* **2021**, *163*, 120465. [CrossRef]

26. Dal Mas, F.; Massaro, M.; Ndou, V.; Raguseo, E. Blockchain Technologies for Sustainability in the Agrifood Sector: A Literature Review of Academic Research and Business Perspectives. *Technol. Forecast. Soc. Chang.* **2023**, *187*, 122155. [CrossRef]

27. Pincheira, M.; Vecchio, M.; Giaffreda, R. Characterization and Costs of Integrating Blockchain and IoT for Agri-Food Traceability Systems. *Systems* **2022**, *10*, 57. [CrossRef]

28. Devine, A.; Jabbar, A.; Kimmitt, J.; Apostolidis, C. Conceptualising a Social Business Blockchain: The Coexistence of Social and Economic Logics. *Technol. Forecast. Soc. Chang.* **2021**, *172*, 120997. [CrossRef]

29. Marsal-Llacuna, M.-L. The People's Smart City Dashboard (PSCD): Delivering on Community-Led Governance with Blockchain. *Technol. Forecast. Soc. Chang.* **2020**, *158*, 120150. [CrossRef]

30. Chin, T.; Shi, Y.; Singh, S.K.; Agbanyo, G.K.; Ferraris, A. Leveraging Blockchain Technology for Green Innovation in Ecosystem-Based Business Models: A Dynamic Capability of Values Appropriation. *Technol. Forecast. Soc. Chang.* **2022**, *183*, 121908. [CrossRef]

31. Friedman, N.; Ormiston, J. Blockchain as a Sustainability-Oriented Innovation?: Opportunities for and Resistance to Blockchain Technology as a Driver of Sustainability in Global Food Supply Chains. *Technol. Forecast. Soc. Chang.* **2022**, *175*, 121403. [CrossRef]

32. Pazaitis, A.; De Filippi, P.; Kostakis, V. Blockchain and Value Systems in the Sharing Economy: The Illustrative Case of Backfeed. *Technol. Forecast. Soc. Chang.* **2017**, *125*, 105–115. [CrossRef]

33. Pólvora, A.; Nascimento, S.; Lourenço, J.S.; Scapolo, F. Blockchain for Industrial Transformations: A Forward-Looking Approach with Multi-Stakeholder Engagement for Policy Advice. *Technol. Forecast. Soc. Chang.* **2020**, *157*, 120091. [CrossRef]

34. Grida, M.O.; Abd Elrahman, S.; Eldrandaly, K.A. Critical Success Factors Evaluation for Blockchain's Adoption and Implementing. *Systems* **2023**, *11*, 2. [CrossRef]

35. Liu, F.; Yang, C.; Yang, J.; Kong, D.; Zhou, A.; Qi, J.; Li, Z. A Hybrid with Distributed Pooling Blockchain Protocol for Image Storage. *Sci. Rep.* **2022**, *12*, 3457. [CrossRef] [PubMed]

36. Nguyen, P.T.; Nguyen, L.T.M. Understanding Platform Market Value through Decentralization Governance—An Integrative Model from Signaling and Mechanism Design Theory. *Technol. Forecast. Soc. Chang.* **2022**, *183*, 121913. [CrossRef]

37. Nam, J.-H.; Lee, S.-J.; Park, S.-P.; Lee, I.-G. IP DLedger—Decentralized Ledger for Intellectual Property Administration. *Technol. Forecast. Soc. Chang.* **2023**, *186*, 122105. [CrossRef]

38. Liu, F.; Li, Z.; Jia, K.; Xiang, P.; Zhou, A.; Qi, J.; Li, Z. Bitcoin Address Clustering Based on Change Address Improvement. *IEEE Trans. Comput. Soc. Syst.* **2023**, 1–12. [CrossRef]

39. Santana, C.; Albareda, L. Blockchain and the Emergence of Decentralized Autonomous Organizations (DAOs): An Integrative Model and Research Agenda. *Technol. Forecast. Soc. Chang.* **2022**, *182*, 121806. [CrossRef]

40. Pereira, J.; Tavalaei, M.M.; Ozalp, H. Blockchain-Based Platforms: Decentralized Infrastructures and Its Boundary Conditions. *Technol. Forecast. Soc. Chang.* **2019**, *146*, 94–102. [CrossRef]

41. Jovanovic, M.; Kostić, N.; Sebastian, I.M.; Sedej, T. Managing a Blockchain-Based Platform Ecosystem for Industry-Wide Adoption: The Case of TradeLens. *Technol. Forecast. Soc. Chang.* **2022**, *184*, 121981. [CrossRef]

42. Daim, T.; Lai, K.K.; Yalcin, H.; Alsoubie, F.; Kumar, V. Forecasting Technological Positioning through Technology Knowledge Redundancy: Patent Citation Analysis of IoT, Cybersecurity, and Blockchain. *Technol. Forecast. Soc. Chang.* **2020**, *161*, 120329. [CrossRef]

43. Zhang, H.; Daim, T.; Zhang, Y. (Peggy) Integrating Patent Analysis into Technology Roadmapping: A Latent Dirichlet Allocation Based Technology Assessment and Roadmapping in the Field of Blockchain. *Technol. Forecast. Soc. Chang.* **2021**, *167*, 120729. [CrossRef]

44. Yu, D.; Pan, T. Tracing the Main Path of Interdisciplinary Research Considering Citation Preference: A Case from Blockchain Domain. *J. Informetr.* **2021**, *15*, 101136. [CrossRef]

45. Zanella, G.; Liu, C.Z.; Choo, K.-K.R. Understanding the Trends in Blockchain Domain Through an Unsupervised Systematic Patent Analysis. *IEEE Trans. Eng. Manag.* **2021**, 1–15. [CrossRef]

46. Someda, H.; Akagi, T.; Kajikawa, Y. An Analysis of the Spillover Effects Based on Patents and Inter-Industrial Transactions for an Emerging Blockchain Technology. *Scientometrics* **2022**, *127*, 4299–4314. [CrossRef]

47. Miau, S.; Yang, J.-M. Bibliometrics-Based Evaluation of the Blockchain Research Trend: 2008—March 2017. *Technol. Anal. Strateg. Manag.* **2018**, *30*, 1029–1045. [CrossRef]

48. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. Available online: https://bitcoin.org/bitcoin.pdf (accessed on 14 January 2023).

49. Clarke, N.S.; Jürgens, B.; Herrero-Solana, V. Blockchain Patent Landscaping: An Expert Based Methodology and Search Query. *World Pat. Inf.* **2020**, *61*, 101964. [CrossRef]

50. Corbet, S.; Lucey, B.; Yarovaya, L. Datestamping the Bitcoin and Ethereum Bubbles. *Financ. Res. Lett.* **2018**, *26*, 81–88. [CrossRef]

51. Bamakan, S.M.H.; Bondarti, A.B.; Bondarti, P.B.; Qu, Q. Blockchain Technology Forecasting by Patent Analytics and Text Mining. *Blockchain Res. Appl.* **2021**, *2*, 100019. [CrossRef]

52. Zhong, B.; Hei, Y.; Jiao, L.; Luo, H.; Tang, J. Technology Frontiers of Building-Integrated Photovoltaics (BIPV): A Patent Co-Citation Analysis. *Int. J. Low-Carbon Technol.* **2020**, *15*, 241–252. [CrossRef]

53. Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE* **2014**, *9*, e98679. [CrossRef]

54. Bonacich, P. Power and Centrality: A Family of Measures. *Am. J. Sociol.* **1987**, *92*, 1170–1182. [CrossRef]

55. Borgatti, S.P.; Everett, M.G. Models of Core/Periphery Structures. *Soc. Netw.* **2000**, *21*, 375–395. [CrossRef]

56. Zhang, X.; Martin, T.; Newman, M.E. Identification of Core-Periphery Structure in Networks. *Phys. Rev. E* **2015**, *91*, 032803. [CrossRef] [PubMed]

57. Boyd, J.P.; Fitzgerald, W.J.; Mahutga, M.C.; Smith, D.A. Computing Continuous Core/Periphery Structures for Social Relations Data with MINRES/SVD. *Soc. Netw.* **2010**, *32*, 125–137. [CrossRef]

58. Rombach, M.P.; Porter, M.A.; Fowler, J.H.; Mucha, P.J. Core-Periphery Structure in Networks. *SIAM J. Appl. Math.* **2014**, *74*, 167–190. [CrossRef]

59. Bekamiri, H.; Hain, D.S.; Jurowetzki, R. PatentSBERTa: A Deep NLP Based Hybrid Model for Patent Distance and Classification Using Augmented SBERT. *arXiv* **2021**, arXiv:2103.11933. [CrossRef]

60. McInnes, L.; Healy, J.; Melville, J. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**, arXiv:1802.03426.

61. Feng, Y.; Zhu, Q.; Lai, K.-H. Corporate Social Responsibility for Supply Chain Management: A Literature Review and Bibliometric Analysis. *J. Clean. Prod.* **2017**, *158*, 296–307. [CrossRef]

62.  Li, G.; Jiang, Z.; Li, X. A Text Mining-Based Approach for Modelling Technical Knowledge Evolution in Patents. *Int. J. Technol. Policy Manag.* **2020**, *20*, 318–339. [CrossRef]

63.  Ranaei, S.; Suominen, A.; Porter, A.; Kässi, T. Application of Text-Analytics in Quantitative Study of Science and Technology. In *Springer Handbook of Science and Technology Indicators*; Glänzel, W., Moed, H.F., Schmoch, U., Thelwall, M., Eds.; Springer Handbooks; Springer International Publishing: Cham, Switzerland, 2019; pp. 957–982, ISBN 978-3-030-02511-3.