Received: 30 June 2022 | Revised: 20 September 2022 | Accepted: 17 October 2022

DOI: 10.4218/etrij.2022-0257

ORIGINAL ARTICLE

ETRI Journal WILEY

Exploring trends in blockchain publications with topic modeling: Implications for forecasting the emergence of industry applications

Jeongho Lee | Hangjung Zo | Tom Steinberger

School of Business and Technology Management, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

Correspondence
Hangjung Zo, School of Business and Technology Management, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea.
Email: joezo@kaist.edu

Abstract

Technological innovation generates products, services, and processes that can disrupt existing industries and lead to the emergence of new fields. Distributed ledger technology, or blockchain, offers novel transparency, security, and anonymity characteristics in transaction data that may disrupt existing industries. However, research attention has largely examined its application to finance. Less is known of any broader applications, particularly in Industry 4.0. This study investigates academic research publications on blockchain and predicts emerging industries using academia-industry dynamics. This study adopts latent Dirichlet allocation and dynamic topic models to analyze large text data with a high capacity for dimensionality reduction. Prior studies confirm that research contributes to technological innovation through spillover, including products, processes, and services. This study predicts emerging industries that will likely incorporate blockchain technology using insights from the knowledge structure of publications.

KEYWORDS

blockchain, dynamic topic model, knowledge spillover, latent Dirichlet allocation, topic modeling

1 | INTRODUCTION

The distributed ledger technology called blockchain enables transparent, secure, and anonymous management of transaction data at scale, potentially disrupting technological and competitive landscapes [1–3]. The 2008 financial crisis, sowing doubt in traditional financial structures, led to the advent of blockchain and cryptocurrencies such as Bitcoin, initially proposed by Nakamoto [4]. Cryptocurrencies are globally traded on various specialized exchanges [5]. Blockchain applications, however, are not restricted to the financial industry [2, 6]. They are among the most frequently discussed subjects in innovation, in academia and industry [6–8].

Innovation research on blockchain's implications has concentrated on cryptocurrencies and related uses [9]. Recently, research has emerged on a much broader set of potential applications for blockchain. It is time to move from analyzing the blockchain as a technological innovation or key technology for cryptocurrencies and finance to investigating how disruptive it may become to diverse industries. Although some studies have identified a broader range of applications, they usually focus on current applications over innovation [10, 11]. This study's analysis develops a topic model for the blockchain literature. This model uses a text mining approach to find the emerging fields most likely to incorporate blockchain technology. We presume that topics emerging in the literature may indicate potential applications.

Potential applications can be sought in papers from relevant disciplines [12, 13], often through text mining, which allows underlying patterns to emerge [14, 15]. Text mining ranges from bibliometric analysis to novel modeling approaches. Advanced topic modeling techniques, such as latent Dirichlet allocation (LDA), are used to reveal patterns and find trends in extensive document collections [16, 17]. This study adopts the unsupervised techniques of LDA and dynamic topic modeling (DTM) to find thematic structures in blockchain-related articles.

DTM is a variant of traditional LDA used to examine topic evolution [18]. In traditional LDA, topics are extracted independently per period. Although labeling topics is possible, using the researcher's own interpretation, it is not possible to identify how a specific topic evolves. This study uses DTM to observe topic evolution in changes in keyword composition.

Previous studies have found that the knowledge produced in academic research positively influences private sector innovation [19, 20]. This study observes this dynamic using technology spillover theory to predict changes in industry use of blockchain.

The combination of two model-based approaches exposed many patterns in the knowledge structure of publications on blockchain publications. This study describes the evolution of research and predicts emerging industries using established theories and identifying relations between academia and industry.

The rest of this study unfolds as follows: Section 2 summarizes works on the blockchain, topic modeling, and academia-industry relations. Section 3 introduces the dataset and methodologies used. The results are presented in Section 4, and Section 5 presents analytical insights. The conclusions are presented in Section 6.

2 | RELATED WORK

2.1 | Blockchain

The blockchain was first proposed by Nakamoto [4] to address the double-spending problem. A blockchain is a transparent, distributed ledger functioning as a peer-to-peer network to record transactions without a centralized authority [1, 4, 6]. It consists of lists of linked blocks in chronological order, where blocks store transaction information [4, 21] and are associated with previous blocks. Network participants verify the transactions, which cannot be modified by unauthorized parties [21]. Its features of transparency, security, anonymity, and data integrity make blockchain difficult to corrupt or alter [2]. Third parties play little role, enabling trust-free economies [22].

The attention blockchain has received has largely focused on the cryptocurrency applications [2, 8, 10, 22, 23]. Until 2015, Bitcoin and similar currencies were the primary topic [9]. However, academic and business research now encompasses Industry 4.0, supporting the development of research on blockchain applications.

The blockchain may transform the landscape of various industries. Its mechanisms for managing transaction data can add value to any transaction or record-keeping type, reducing cost, and increasing security [6]. Industries may be disrupted by smart contracts—code on the blockchain that self-executes when predetermined conditions are met [24]. Several industries already use commercial blockchain-based applications [1, 25]. Blockchain technology was once a near-synonym for Bitcoin, but its applications have grown, and now, many see its potential as a general-purpose technology [8, 11]. However, the range of possible blockchain applications has received limited attention in the innovation literature.

Blockchain has seen several major innovations, divided into three generations [26, 27]. Blockchain 1.0 enables digital transaction systems using cryptocurrencies. The scope of blockchain 2.0 goes beyond cryptocurrency to the entire financial sector. Like blockchain 1.0, blockchain 2.0 ensures a range of finance-related applications, including smart contracts. Blockchain 3.0 creates new values and innovation in healthcare, the Internet of Things (IoT), government, privacy, and security [11, 26–29]. This study provides insight into where the blockchain may expand, using topic modeling of the blockchain literature.

2.2 | Bibliometric analysis and topic modeling

The explosive growth of digitized documents has led to new tools and techniques for automatically organizing and analyzing them. Research has evaluated the ways that how specific academic disciplines develop [12, 13], using bibliometric and content analyses. White and McCain [12] conducted a co-citation analysis in information science. Koufogiannakis and others [30] classified the literature by study type and domain using content analysis to depict the features and evolution of academic research. Topic modeling is increasingly used by researchers for its advantages in examining large, unstructured document collections [13]. The performance and benefits of this approach are established [17, 18, 31].

Text mining is used to analyze large document collections. Using these techniques, high-quality information can be derived, extracting useful knowledge and identifying underlying patterns in the data [14, 15]. Text mining includes information retrieval and extraction, categorization, clustering, and summarization [32].

Topic modeling is among the most common text mining techniques. It identifies latent semantic structures in unstructured document collections from distributions of words and topics [31, 33]. Advanced topic modeling is frequently used in machine learning and statistics for large document collections [16–18, 34].

LDA is a generative probabilistic model used to analyze large text collections proposed by Blei and others [16]. It is an unsupervised technique that automatically extracts unobserved and latent topics from word distributions in a document without labeling [31, 35]. In the LDA approach, words in a document are exchangeable, and sequential orders of terms are unimportant [16]. On this assumption, documents contain multiple topics, and particular topics are distributed over a fixed vocabulary [16, 36]. LDA is a generative probabilistic model for extracting topics. It combines distributions of topics and topic words whose joint probability follows a Dirichlet distribution [16]. LDA creates lists of weighted topics and topic words, and researchers can label topics and interpret them [34], exploring hidden thematic structures [13, 36].

Blei and Lafferty [18] proposed DTM, a modified and extended LDA, to examine extracted topics. Both LDA and DTM are unsupervised techniques frequently used in machine learning and statistics [16–18, 34]. Adding a time element to LDA, DTM can address LDA's shortcomings to develop an understanding of the dynamics of extracted topics, showing how thematic structures evolve [18]. Although both DTM and traditional time-series analysis require time indicators, DTM requires categorical time data. DTM datasets are split into predetermined periods, such a month or year [18]. An individual LDA model in DTM is trained in each period, but topics determined in the first period remain constant over time [18, 37]. The weighted importance of words fluctuates over time, whereas topics remain fixed, demonstrating changes in trends [5]. Compositional changes and emerging trends can be observed in changes in discrete cases.

This study extracts latent themes from recently published articles, as older papers may be far from contemporary thinking [13]. The topics extracted from LDA and DTM consist of words with probabilities, enabling researchers to interpret, label, and understand them.

2.3 | Academia-industry dynamics

Study of innovation and knowledge has found that knowledge produced in scholarly research positively influences technological innovation [19, 20]. Private sector entities benefit from scientific knowledge spillovers [20]. The theoretical concept of technology spillover grounds our motivation to study how emerging scholarly publications on the blockchain influence the technological and competitive landscapes of different industrial sectors.

Some studies have found no direct link or only weak interaction between early-stage research and industry [38]. However, several studies have argued that academic research helps develop new products and processes in many industries [39, 40]. New theoretical and empirical findings from academic research find a place in an industrial sector, leading to product and process developments [39]. Mansfield [39, 40] found that many new products and processes are based on academic research. Innovation commercialization can lag academic findings approximately 6 or 7 years.

Academic research tends to generate innovation in a range of industries [41]. Firms in various sectors recognize that academic research is an essential input for their research [42, 43]. The influence of scholarly research on industry can be measured with the number of patented inventions or commercial innovations as proxies [44, 45]. The influence of academic research on industrial innovation is stronger than it is on patented inventions [45]. Knowledge spillovers can be measured in terms of citations of academic research in patent applications instead of counting the number of patents or innovations [43]. Branstetter [43] found that citations exhibit a growing trend over time and typically begin soon after publications, peaking after 4 years. Although they vary across sectors and may be concentrated in specific areas, and a trend toward shorter lags between technological innovation and academic research has been identified. Hence, academic research can forecast innovation, including products, processes, and services, contributing to various products and processes [46].

The link between academia and industry may be reinforced through new synergies. Previous studies show that innovative strategies, intermediaries, including universities and technology transfer offices, and university-industry collaboration may accelerate technological dissemination from academia to industry [47–50]. Well-planned strategies and strong intermediaries can enhance knowledge transfer from the public to the private sector, strengthening their links.

Lieberman [51] suggested that the relationship between science and technology diminishes as the technology matures, and a more recent study indicated that potentially stronger and more direct linkages are formed between for new science and technology [52]. Hence, the technology spillover between academia and industry is likely strong in emerging blockchain technology. Knowledge spillovers from academic research to technological innovation provides the theoretical foundation for this study's assumption that private sector blockchain innovations are influenced by academic research findings and can be achieved within a few years.

Using bibliometric approaches, such as co-authorships and citations, previous studies have identified links among interdisciplinary areas [53, 54]. Recent studies show that model-based approaches demonstrate a remarkable capacity for revealing underlying semantic structures and emerging trends beyond bibliometric approaches.

Prior studies have used literature review or topic modeling on blockchain publications to identify the current status of applications [10, 11]. This study investigates academia-industry dynamics as a theoretical background for predicting the state of blockchain-related industries, examining topics extracted with LDA and DTM.

3 | METHODOLOGY

3.1 | Data

Publications relevant to blockchain technology, including articles and conference proceedings, were identified in the Scopus database. We examined these documents to determine how far blockchain technology was discussed in relation to businesses and industry. Scopus is often used as a source for bibliometric analysis and text mining to identify relevant studies [10, 13]. Documents collected using "Blockchain" as a keyword cover many disciplines and non-academic papers. To include publications on fundamental blockchain technology, we used "Blockchain," "Block chain," "Block-chain," "Privacy," and "Security" as keywords to cover a broad range of academic papers, collecting publications in English from 2008 to October 2020. We collected the title, year of publication, abstract, authors, and keywords for each. We obtained this information for 1900 documents.

Several preprocessing steps were taken to clean the dataset. Based on the process suggested by Moher and others [55], Scopus articles were tested for eligibility before the examination. We discarded documents without adequate abstracts or with multiple abstracts and conference proceedings. We also discarded documents with insufficient data before 2018. In all, we obtained 1773 documents from the first screening and excluded an

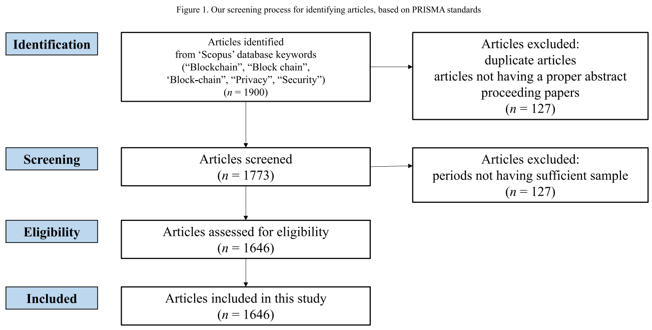

FIGURE 1 Our screening process for identifying articles, based on PRISMA standards [55]

additional 127 from periods without sufficient samples. Although the number of blockchain studies began to grow in 2018, fewer blockchain publications were collected from 2021, as we collected data before 2021 ends. We included the abstracts from 268 documents from 2018, 734 from 2019, and 644 from 2020. Figure 1 presents a flowchart of the data screening process.

The abstracts were analyzed using preprocessing techniques. Using software packages and libraries in Python, such as Pandas, Numpy, and Natural Language Toolkit (NLTK), the abstracts of the screened documents were preprocessed and cleaned [56, 57]. We used lowercasing, stop word removal, tokenization, and lemmatization to eliminate possible distorting factors [58]. For stop word removal, we used NLTK's list of English stop words. Removing extra spaces, punctuations, and stop words eliminates characters that do not have semantic significance. Tokenization converts texts and sentences into word lists, a powerful technique when combined with the above processes. Lemmatization is an excellent technique for use in data dimensionality reduction, which provides part-of-speech (PoS) tagging and reduces all versions of a specific word to one word. Finally, a trigram (*N*-gram) model was built to handle frequently and sequentially occurring words.

3.2 | Analysis

3.2.1 | LDA

LDA is an excellent tool for extracting thematic structures from a collection of documents [18]. Each document has a probability distribution over topics, and each topic has a probability distribution over words, where words belong to a specific document topic [16, 17]. We discovered representative topics with relevant keywords from various documents based on their distribution and relationships.

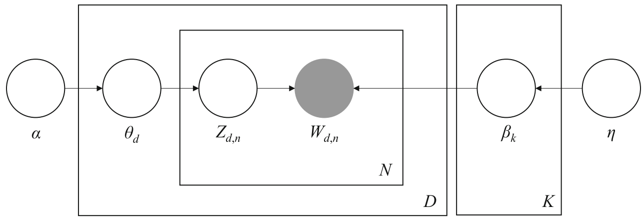

FIGURE 2 Graphical representation of LDA

We performed LDA analysis using the Gensim and Mallet Python libraries, open-source toolkits providing various machine-learning applications, including topic modeling, and these libraries can automatically train and tune hyperparameters (α and η in Figure 2) directly from the dataset.

Figure 2 shows how LDA works in formal notation. $W_{d,n}$, the only observable variable, is highlighted. The other variables are latent [16]. D, K, and N denote numbers of documents, topics, and words. Topic distribution for the dth document θ_d is obtained from the Dirichlet distribution with proportion parameter α. Word distribution for the kth topic β_k is derived from the topic hyperparameter η [13, 16, 59]. Finally, $W_{d,n}$ is a specific word drawn from a document, and $Z_{d,n}$ is the topic assignment for the nth word in the dth document, indicating its word topic assignment [13, 16]. LDA helps researchers identify the underlying semantic structure of documents by interpreting topics extracted from the document-topic and topic-term distributions.

We conducted several tests to build the topic model. Although perplexity measures are frequently used to evaluate topic models and determine optimal topic numbers, optimizing it may distort the evaluation of discovered topics [60]. We optimized the number of topics using coherence scoring to distinguish semantically interpretable topics, as top terms tend to co-occur [61]. We identified eight optimized topics using the coherence score. In addition to this score, we used LDAvis, an interactive graphical representation of inter-topic distance, to determine whether topics are well-scattered. They were scattered throughout the quadrants, and only minimal overlapping existed, indicating an effective topic model.

We tuned the Dirichlet parameter and the topic hyperparameters to ensuring adequate representation of the parameters and model reliability. Steyvers and Griffiths [62] offered guidelines for optimized parameters, but the Gensim library automatically trains the model, providing the optimal α and η values after multiple iterations. Using 1646 abstracts, grouped by year of publication, we extracted the eight most representative topics in each period with the 10 most relevant keywords by relative importance and contributions.

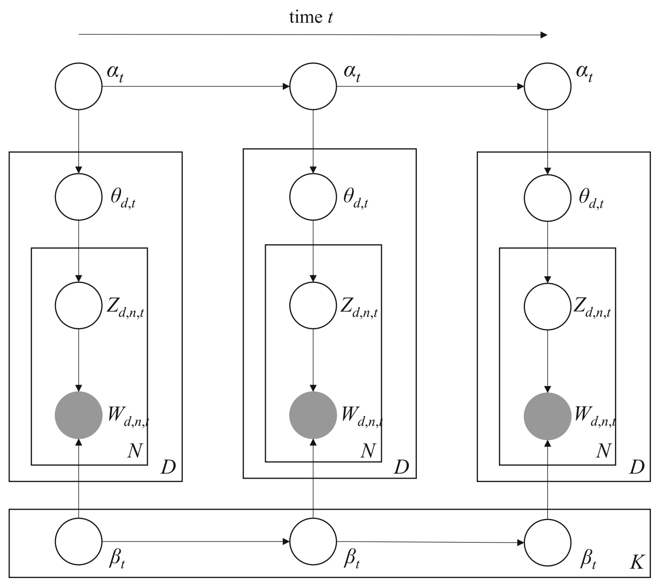

FIGURE 3 Graphical representation of DTM

3.2.2 | Dynamic topic model

DTM is useful for sequentially organized documents, depicting the evolution of topics over time [18]. An individual and independent LDA is modeled in predetermined periods, as shown in Figure 3. When time dynamics are added to the model, topics smoothly evolve from one period to the next [18].

The Gensim library is used as it covers several machine-learning applications. For consistency, we adopted the same dataset as used for LDA; however, the DTM dataset was split by period, as suggested by Blei and Lafferty [18]. We analyzed 268, 734, and 644 articles from 2018, 2019, and 2020, and we set the number of topics to eight to compare across LDA analyses and maintain consistency. Figure 3 presents the generative process of the model.

The notations in Figure 3 are identical to the LDA approach but feature different derivations. The implementation of DTM is similar to LDA but features time dynamics. Here, t and K are integers representing the year and number of topics, respectively. D and N indicate the document number and words in a document, respectively. The difference is shown by the longitudinal arrows of the mean parameter α_t and natural parameter $\beta_{t,k}$, which represents years [18]. These parameters determine the distribution of documents over topics and word distribution of topics, which evolve with Gaussian noise [18]. This generative process recalls LDA's, but the distribution used to extract a topic and a word depends on the time variable.

The generative process of sequentially chained topic models in each period is defined as follows [18, 63]: First, the determination of parameters α_t and $\beta_{t,k}$ is given by the following equations:

$$\alpha_t \mid \alpha_{t-1} \sim N\left(\alpha_{t-1}, \delta^2 I\right), \quad (1)$$

$$\beta_{t,k} \mid \beta_{t-1,k} \sim N\left(\beta_{t-1,k}, \sigma^2 I\right). \quad (2)$$

Distribution over a topic for the dth document at the tth period, $\theta_{d,t}$ depends on α_t, and for each document is obtained with this expression:

$$\theta_{d,t} \sim N\left(\alpha_t, a^2 I\right). \quad (3)$$

From the topic distribution $\theta_{d,t}$, we calculate $Z_{d,n,t}$, which assigns the topic for the nth word in the dth document at time t. $Z_{d,n,t}$ is the topic to which the nth word in the dth document at time t belongs. $W_{d,n,t}$ is the only observable variable in the dynamic model and a specific word of the dth document at the tth period, based on the distribution of $Z_{d,n,t}$ and $\beta_{t,k}$. Hence, a word is extracted from the distribution over words of the kth topic in the tth period. The topic assignment of the nth word in the dth document at time t is as follows:

$$Z_{d,n,t} \sim \mathrm{Mult}(\pi(\theta_{d,t}), \quad (4)$$

$$W_{d,n,t} \sim \mathrm{Mult}\left(\pi\left(\beta_{t,z}\right)\right). \quad (5)$$

4 | RESULTS

4.1 | LDA results

We conducted multiple LDA topic modeling analyses of 1646 abstracts obtained from the Scopus database, and eight topics were derived across 3 years. The topics from the extracted keywords were labeled as seen in Tables 1–3. The computed topic models met the requirements for an effective topic model. Topics on the intertopic distance map were well-scattered, with a small overlapping area across the map. Hence, the extracted topics were independent of each other and could serve as adequate analytical objects.

The LDA model extracted eight dominant topics for each year. We report the top 10 salient keywords for each topic in terms of their relative importance or weighting in Tables 1–3. The keywords are computed based on their contribution to corresponding topics and ordered by their contributions.

Table 1 presents each topic's top 10 salient keywords in 2018: (1) transaction management, (2) user privacy, (3) IoT, (4) secured framework, (5) network architecture, (6) decentralized platform, (7) patient data protection, and (8) cryptocurrency security. By establishing the results of the first period as the evaluation standard, this study observes the evolution of extracted topics over the following periods. For instance, cryptocurrency security disappeared after the first period.

We extracted similar topics and keywords in the second period (Table 2), with slight differences in topic composition: (1) user privacy, (2) IoT architecture, (3) security attack, (4) network security, (5) blockchain platform, (6) electronic medical record (EMR), (7) vehicle communication, and (8) cloud storage are extracted as significant topics in the second period. Topic labeling altered, although keywords imply a consistent tendency. New topics, including vehicular communication and cloud storage, appeared in 2019.

In the third period, most of the keywords extracted in the previous periods continue (Table 3). The transaction platform topic from the first period reappears with minor changes in composition. However, topics related to

TABLE 1 LDA results in 2018

Topic (subject)	Keywords
Topic 1 (Transaction management)	transaction, trust, distribute, energy, problem, identity, management, solution, mechanism, digital
Topic 2 (User privacy)	user, privacy, scheme, base, propose, result, preserve, efficient, security, cloud
Topic 3 (IoT)	technology, IoT, application, challenge, research, blockchain, study, include, discuss, business
Topic 4 (Secured framework)	secure, propose, provide, framework, record, key, enable, communication, base, design
Topic 5 (Network architecture)	base, network, security, paper, device, model, architecture, issue, privacy, smart
Topic 6 (Decentralized platform)	system, service, platform, storage, decentralize, blockchain, make, implement, protocol, decentralized
Topic 7 (Patient data protection)	datum, privacy, share, data, information, access, patient, protection, store, issue
Topic 8 (Cryptocurrency security)	security, privacy, present, current, attack, authentication, cryptocurrency, provide, improve, infrastructure

TABLE 2 LDA results in 2019

Topic (subject)	Keywords
Topic 1 (User privacy)	user, privacy, scheme, base, design, provide, preserve, achieve, propose, protocol
Topic 2 (IoT architecture)	security, IoT, application, device, challenge, technology, issue, architecture, solution, research
Topic 3 (Security attack)	propose, base, model, security, secure, attack, authentication, key, public, node
Topic 4 (Network security)	security, system, network, make, service, identity, paper, problem, address, digital
Topic 5 (Blockchain platform)	technology, blockchain, study, approach, platform, trust, provide, framework, current, industry
Topic 6 (Electronic medical record)	system, information, share, patient, access, privacy, medical, record, healthcare, management
Topic 7 (Vehicle communication)	transaction, propose, privacy, network, vehicle, high, communication, cost, paper, time
Topic 8 (Cloud storage)	datum, system, data, storage, distribute, store, privacy, enable, cloud, ensure

TABLE 3 LDA results in 2020

Topic (subject)	Keywords
Topic 1 (Transaction platform)	system, base, distribute, transaction, increase, energy, management, cost, provide, platform
Topic 2 (Privacy protection)	privacy, model, propose, service, performance, protection, analysis, transaction, preserve, protocol
Topic 3 (Smart contract in IoT)	security, smart, IoT, device, contract, provide, issue, privacy, resource, infrastructure
Topic 4 (Secure authentication)	trust, process, key, identity, authentication, approach, secure, result, technology, framework
Topic 5 (Blockchain challenge)	technology, application, challenge, blockchain, research, paper, security, study, current, present
Topic 6 (Electronic medical record)	datum, system, information, share, medical, patient, record, data, health, healthcare
Topic 7 (Vehicular network)	network, base, attack, mechanism, communication, vehicle, propose, architecture, node, security
Topic 8 (Cloud data protection)	user, scheme, datum, base, propose, storage, cloud, privacy, security, problem

network architecture or security are no longer extracted as a significant topic. Furthermore, two topics linked to keywords about vehicular networks and cloud storage continue to be significant. The following topics are derived: (1) transaction platform, (2) privacy protection, (3) smart contract in IoT, (4) secure authentication, (5) blockchain challenge, (6) EMR, (7) vehicular network, and (8) cloud data protection.

Although extracted from independently conducted LDA analyses, most topics are identical across three periods with only slight differences in keywords. The extracted topics demonstrate similar tendencies and directionality across periods but have slightly different compositions each year, and a few topics disappeared or emerged over this period.

4.2 | DTM results

We examine the evolution of topics using DTM, an extended version of traditional LDA. Unlike conventional LDA, where extracted topics in different periods are independent, topics extracted from the DTM in one period are related to topics in following periods. In other words, the k^{th} topic of the t^{th} time slot should be similar to the k^{th} topic of the $(t-1)^{th}$ time slot [18]. Under these constraints, DTM enables trend analysis, showing how a group of topics can evolve based on keyword changes. As with LDA, the top 10 keywords are computed based on their relative importance to corresponding topics and ordered according to their contributions to topics. Table 4 presents the DTM results.

The topics extracted were: (1) distributed platform, (2) privacy and security, (3) EMR, (4) vehicular network, (5) blockchain technology, (6) cloud storage, (7) IoT, and (8) application. The relative importance of keywords about specific topic changed as they move upward or downward, based on their contributions in different periods (Table 4). However, unlike the keywords for the remaining topics, those constructing the eighth topic changed significantly in terms of relative importance and composition. Our analysis revealed that research on blockchain-based applications underwent significant changes and minor changes in the composition of keywords from the first to the eighth topic.

Although the previous literature has often classified blockchain applications into financial and nonfinancial ones, the scope of blockchain applications has expanded to many areas such as healthcare and IoT [21, 26, 27]. Indeed, various blockchain applications, including cryptocurrency, security, and healthcare, are classified and extracted from prior research that performed a systematic literature review and topic modeling [10, 11]. We

TABLE 4 DTM results from 2018 to 2020

2018 (1st period)	2019 (2nd period)	2020 (3rd period)
Topic 1: Distributed platform		
energy, transaction, system, security, base, propose, privacy, smart, distribute, power	transaction, system, energy, privacy, security, propose, base, smart, power, use	system, transaction, energy, smart, privacy, security, base, propose, use, cost
Topic 2: Privacy and security		
user, privacy, attack, base, transaction, device, location, use, security, mobile	user, privacy, attack, location, security, base, use, propose, device, information	user, privacy, attack, use, location, propose, service, information, security, base
Topic 3: Electronic medical record (EMR)		
datum, privacy, patient, share, use, system, security, information, medical, propose	datum, patient, privacy, medical, share, information, use, system, security, health	datum, patient, medical, system, privacy, use, information, share, health, propose
Topic 4: Vehicular network		
propose, privacy, scheme, base, security, vehicle, network, trust, communication, system	propose, privacy, scheme, base, network, vehicle, security, authentication, use, system	propose, privacy, network, vehicle, base, security, scheme, key, communication, use
Topic 5: Blockchain technology		
technology, security, privacy, system, application, paper, research, blockchain, use, challenge	technology, security, system, privacy, paper, application, research, blockchain, use, challenge	technology, security, privacy, research, system, application, blockchain, challenge, paper, use
Topic 6: Cloud storage		
datum, access, control, cloud, privacy, base, user, security, identity, system	datum, access, control, user, system, base, privacy, security, propose, cloud	datum, access, privacy, base, scheme, propose, user, security, system, control
Topic 7: IoT		
IoT, smart, security, device, contract, privacy, base, application, system, technology	IoT, security, smart, device, system, privacy, architecture, technology, application, network	security, IoT, smart, device, privacy, system, network, datum, technology, application
Topic 8: Application		
IoT, smart, security, device, contract, privacy, base, application, system, technology	IoT, security, smart, device, system, privacy, architecture, technology, application, network	security, IoT, smart, device, privacy, system, network, datum, technology, application

extracted similar representative topics to previous studies, but our results are distinct in detail.

We derived similar but not identical topics from three individual LDA analyses and a single DTM. The LDA analyses identify the dominant topics for each period, regardless of the topics extracted in the previous period. However, DTM provides evidence for changing trends in each topic, addressing variations in keywords.

5 | DISCUSSION

This study examined changes in blockchain-related publications and its implications for the emergence of industrial developments based on two complementary topic modeling approaches. This section further discusses the labeled topics obtained from the LDA and DTM analyses in Section 4.

We ran models across the 3 years and found slightly different results from those of the DTM. These differences reflect the absence of time dynamics in LDA. DTM describes how a specific topic evolves by looking at the relative importance of extracted keywords; independent LDA analyses derived eight topics separately in separate years. Although the topics were independently derived, we describe the development of a discipline by labeling extracted topics and observing their emergence and disappearance. The keyword composition and fluctuations provided us with rich insights into the evolution of blockchain-related publications.

Below, we describe insights derived from individual LDA analyses and an interpretation of topic evolution using keyword compositions derived from the DTM analyses. We then forecast emerging blockchain-related fields in industrial areas by comparing the LDA and DTM results. Finally, we discuss how far the literature on the relationship between academia and industry supports our forecasts.

5.1 | Dominant topics from 2018 to 2020 from LDA

Eight topics were extracted every year and are arranged for comparison in Table 5. Although LDA analysis extracted the same number of topics for the 3 years, we observed differences across the three periods. Some topics disappeared, and others emerged, indicating transitions in the research interests of blockchain-related research. This study suggests that the composition of extracted topic would change over time, reflecting the fact that the researchers value subject areas differently at different times.

Remarkably, the topic of the security of cryptocurrency (the eighth topic in 2018) and keywords related to cryptocurrency were no longer present after 2018. High volatility characterized the cryptocurrency market in 2018, but Bitcoin and blockchain continued to attract public interest and grew in popularity. The term "blockchain" was once used interchangeably with "Bitcoin" as there were few cryptocurrencies or blockchain uses [11]. Bitcoin was the first cryptocurrency based on blockchain technology [4]. As understanding of blockchain increases, the term "blockchain" was no longer considered synonymous to "cryptocurrency." This study reflects this transition, as the eighth topic in 2018 related to (1) whether cryptocurrencies are reliable in security and privacy and (2) cyberattacks on cryptocurrencies.

Another group of topics extracted for 2018 and 2019, but not apparent in 2020, was related to network security. Security issues have long a major problem for systems, services, and devices. The fifth topic suggested this phenomenon in 2018, and the fourth topic did so in 2019. Both were related to security issues that can affect systems or devices. In 2018, the fifth topic associated with networks was related to building an architecture to protect a network. By contrast, the topic extracted the following year was concerned with network systems security and measures to deal with system issues.

By contrast, the network topic was no longer significant in 2020. This study suggests that after 2019, networks were a fundamental element of a system or service and were no longer regarded a separate issue or a topic. Other extracted topics were vehicular networks, smart contracts in IoT, and transaction platforms, which support this interpretation.

Communication among vehicles and other network-enabled devices emerged as a significant topic in the 2019 data in terms of network, and communication-related functions embedded in vehicles are expected to approach intelligent transportation systems (ITS) [64]. Vehicle-related topics are extracted in 2019 and 2020 due to their increasing importance, and their coverage expands from mere communication to vehicular networks. These networks require secure transactions across IoT networks, the keywords "vehicle," "communication," "transaction," and "cost" were extracted in 2019. In 2020, this topic encompassed (1) mechanisms and architectures of vehicular networks and (2) means of protection from external attacks.

The extracted keywords of cloud storage systems indicated an interesting transition. One fundamental purpose of cloud systems is data storage. Topics from 2019 and 2020 dealt data storage means. The keyword "user" was extracted in 2020, indicating means to storing "user data" attracted more importance than "data." The distribution or decentralization of data in cloud systems was emphasized in 2019 but not in 2020. However, privacy and security-related issues were seen in 2020.

Although the analysis already extracted a network topic, new topics on vehicular networks and cloud storage, a subcategory of network architecture, emerged in

TABLE 5 Topics extracted through LDA from 2018 to 2020

2018 (1st period)	2019 (2nd period)	2020 (3rd period)
Topic 1: Transaction management	-	Topic 1: Transaction platform
Topic 2: User privacy	Topic 1: User privacy	Topic 2: Privacy protection
Topic 3: IoT	Topic 2: IoT architecture	Topic 3: Smart contract in IoT
Topic 4: Secured framework	Topic 3: Security attack	Topic 4: Secure authentication
Topic 5: Network architecture	Topic 4: Network security	-
Topic 6: Decentralized platform	Topic 5: Blockchain platform	Topic 5: Blockchain challenge
Topic 7: Patient's data protection	Topic 6: Electronic medical record	Topic 6: Electronic medical record
Topic 8: Cryptocurrency security	-	-
-	Topic 7: Vehicle communication	Topic 7: Vehicle network
-	Topic 8: Cloud storage	Topic 8: Cloud data protection

2019. This suggests detailed and specific aspects of network system gained significance and influenced blockchain applications in 2019.

Transactions involving blockchain were the most popular and commercialized applications in finance for their transparency, reliability, and immutability [2]. Therefore, transaction management was extracted in the first period. "energy" also emerged, indicating that blockchain technology was usable for transparency, security, and smart grids in energy transactions [11]. The keyword "transaction" was associated with vehicle-related topics in the second period, as blockchain technology is increasingly being involved in vehicular networks. The importance of "system," "distribute," and "cost" was heavily emphasized in the third period. Beyond financial transactions, distributed transaction platforms formed an emerging research area and an application with a broad range.

Other topics, such as user privacy, IoT, security, decentralization with blockchain, and electronic medical records, were extracted throughout the study period. We extracted user privacy topics, and keywords for these topics were also consistent. "Design" and "scheme" for user privacy were emphasized in 2019, and "model" and "service" gained traction in 2020, reflecting interest in blockchain for privacy protection in transactions [2, 6, 10].

IoT requires a rapid, such as 5G, reliable, and highly secure connection [24]. IoT topics were continuously extracted during the study period. By contrast, the keywords for each topic reflect the evolution and development of IoT applications for the blockchain. The top terms of IoT-related topics demonstrated substantial change, evolving from "business" and "application" in 2018 to "security" and "architecture" in 2019 and "smart," "contract," and "infrastructure" in 2020. The scope of research expanded from simple IoT applications to an architecture of solutions based on IoT. In addition, smart contracts on blockchain platforms are key enablers of IoT [65], where privacy and security are vital.

Beyond transparency, security in blockchain technology is frequently mentioned, and topics related to blockchain security were consistently extracted in the three periods. Secure communication and means of keeping records were important in 2018, so establishing appropriate frameworks and designs to fulfill these objectives was a high priority. In 2019, the topic had advanced further and extended to types of authentication and means of maintaining security against attacks on nodes and models. "Authentication" was an essential keyword for 2020, when keywords such as "trust" and "identity" related to components of the framework gained significance.

The introduction of blockchain technology has led to various discussions on decentralized and distributed platforms. The decentralization of traditional services and platforms attracted attention, as blockchain has proven safe and transparent [2]. Related extracted topics initially refer to a decentralized platform or service; then, it evolves and covers (1) how blockchain technology is applied and incorporated into industries and (2) scientific research and studies on blockchain and upcoming challenges that it may face over time.

Blockchain is also increasingly seen as a reliable and safe means for preserving patients' data in hospitals. Topics related to patient data are extracted with an almost identical keyword composition. Topics consistently include means of keeping patients' data safe using blockchain technology. The first period focused on safely maintaining and managing patient information over establishing comprehensive systems. This topic focused on managing, sharing, and accessing patient data in systemized electronic medical records in the second period. Coverage expanded to individuals' health records, beyond healthcare.

Extracted topics and top keywords provided rich information on changes in research interest in the blockchain. For instance, as interest in cryptocurrency fell, the topic disappeared. Likewise, LDA analysis reflects emerging topics and keywords.

5.2 | Evolution of topics based on DTM

The DTM results from 2018 to 2020 indicate that keyword weighting for a specific topic change over time. Predetermined topics in the first period are labeled, and changes in the keyword composition represent their smooth evolution over time. Table 4 shows how a particular topic evolved from 2018 to 2020.

In the first topic (distributed platform) in the first period, "energy" was the top term, followed by "transaction" and "system." Together with "distribute" and "transaction," the first period focused on distributed energy trading or resource use platforms with smart grids. However, the importance of the term "energy" decreased over time. In addition, increased weighted importance for "privacy" and decreased weighted importance for "security" explains the current situation, where solutions to privacy issues have gained traction. As "system" became critical, "cost" became a top term in 2020. This accounts for new circumstances, in which reducing costs in transactions and distributed platforms was essential. As the smart contract became better-known, the term "smart" appeared in the second period and increased in importance. Enlarged and extended measures of applying smart contracts to systems and platforms explain this tendency.

The second topic, "privacy and security," showed the same top three terms throughout the sample period. Except for a few instances, most keywords shared a distribution and composition over time. The weighted importance of "location" grew and decreased, whereas "information" was a new top term in the second period and grew in importance. This composition variation indicated that properly managing individual or private information become crucial in 2019. Recently, privacy and security of services became critical, explaining the decreased importance of "device" and "mobile" and the emergence of "service" in the third period.

The term "datum" remained the highest position among the keywords related to the third topic (electronic medical records) in the study period. The weighted significance of "medical" increased. The concept of electronic medical records expanded from the second period to embrace the health category as "health" emerged as a prominent keyword. The importance of "privacy" and "security" fluctuated and diminished as patient-centric, medical and health systems became crucial. The upward movement of "patient," "health," and "system" accounts for such circumstances, as suggested by individual LDA analyses.

Among the keywords for the fourth topic (vehicular network), the term "vehicle" increased over the entire sample period. The term "network" also gained significance, suggesting that blockchain technology is increasingly absorbed in vehicular networks. The constant rise in the importance of "network" implies that vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-grid (V2G) networks became crucial.

The composition of terms referring to the fifth topic (blockchain technology) indicated an interesting transition. The transition of "application" and "research" in the third period suggests that more academic research was produced, and its focus shifted from creating blockchain applications to generating research-based knowledge. The upward movement of "challenge" suggests that obstacles and challenges exist in applying blockchain technology across industries.

A few top keywords under the sixth topic (cloud storage) showed dramatic fluctuation in the weighted importance of keywords. The term "cloud" lost importance after the first-time slot, and "privacy" and "scheme" gained significance among the top terms. These results indicate the importance of appropriate schemes for operating systems and managing privacy issues.

The IoT is a widely acknowledged technology in theory and practice [24, 65]. The seventh topic (IoT) covers keywords and terms regarding IoT; "privacy" and "security" exhibited growing importance compared to the first introduction of the concept. The smart contract technique is grafted onto the IoT network by the presence of "smart" and "contract" across the periods. In the new application of smart contracts, keywords such as "security" and "privacy" gained importance. In addition, "application" lost traction as being more relevant to commercialization than research. More research-centric terms, such as "network" and "data," persisted over time. As smart contracts demonstrated high potential for various practical uses, related terms also gained importance.

The last topic (application) dealt with more comprehensive issues than previous topics and exhibited the most significant transitions in the composition of top terms. Blockchain-based services, including e-residency and e-voting, are as well-known as cryptocurrency, and blockchain technology is now an easily recognized concept [26]. In the first period, "government" was the top term, suggesting that many attempts have been made in government-related fields. Different terms replaced "government" over time. After the first period, the term "financial" became more important, as financial-related applications became crucial commercially and academically. The emergence of novel applications, platforms, and cryptocurrencies explains this phenomenon. Regarding applications, "information" remained significant despite fluctuations in its importance.

In DTM, topics were predetermined in the first period. Changes in the ranks of keywords represented the evolution of topics, and keyword composition differed slightly from one period to another. From the keywords, this study confirms evolution of and trends in topics.

5.3 | Emerging industrial fields based on analyses

Because topics were independently derived through LDA, topics in the last period can be considered as evolving from previous time slices. Hence, the LDA results for 2020 are comparable with the DTM results (Table 6). Keywords were separately extracted using various techniques; they were labeled later with almost identical themes. The two model-based approaches complement each other, demonstrating the development of blockchain-related publications through (1) changes in extracted topics and (2) evolution in the keyword composition.

It is possible to take a further step based on the dynamics between academia and industry. Although the linkage may be stronger or weaker, according to the role of intermediaries, the previous literature provides a strong link between academia and industry, with knowledge spillover from the public to the private sector.

TABLE 6 Comparison between the LDA and DTM results: Topics in blockchain-relating publications

DTM from 2018 to 2020	LDA in 2020
Topic 1: Distributed platform	Topic 1: Transaction platform
Topic 2: Privacy and security	Topic 2: Privacy protection
Topic 3: Electronic medical record	Topic 3: Electronic medical record
Topic 4: Vehicular network	Topic 4: Vehicular network
Topic 5: Blockchain technology	Topic 5: Blockchain challenge
Topic 6: Cloud storage	Topic 6: Cloud data protection
Topic 7: IoT	Topic 7: Smart contract in IoT
Topic 8: Application	Topic 8: Secure authentication

Hence, this study forecasts several industrial sectors to incorporate blockchain technology in the future based on model-based analyses.

6 | CONCLUSION

6.1 | Contributions

This study examines how blockchain-related research topics evolved using yearly independent LDA analyses and DTM over 3 years and considers the implications for industrial applications. Similar topics were extracted from LDA and DTM analyses, with small exceptions. As a result of technological innovation, some research topics have either been entirely absorbed by others or became outdated. For instance, cryptocurrencies are no longer the center of attention, whereas vehicular networks are an emerging research interest. LDA provides independently extracted topics across three periods, allowing trends to be observed. The DTM captures the evolution of the extracted topics, observing fluctuations in keyword compositions.

Model-based approaches are frequently used to show the underlying semantic structures of documents to understand a specific academic discipline [17, 18, 31]. This study takes one step further, combining previous approaches to explore insights into the dynamics of academic research for industrial applications with model-based analyses, suggesting that multiple industrial fields may soon incorporate blockchain technology. The derivation of emerging industrial applications for blockchain technology is supported and confirmed by two model-based techniques, through which we achieved high consistency in topics and keywords.

6.2 | Limitations and future research

We suggest future research directions from the limitations of this study. Blockchain is a relatively new concept, and an understanding of it has only recently emerged. Hence, there are still a fairly limited number of blockchain-related publications. The relatively short history of blockchain research could affect its robustness as a discipline, making it difficult to precisely forecast emerging industrial applications.

This study examines the evolution of academic research by observing changes in topics and terms in discrete periods, but other methods, such as the continuous-time dynamic topic model (cDTM) and topic over time (ToT), using time-series data without discretization of time [66, 67]. Adopting one of these methods may provide valuable insights from a different perspective.

Additional insight into emerging blockchain-related applications may also be obtained by building on the results of our study. Future research could investigate the eight fields extracted from model-based approaches finding how each area incorporates blockchain technology in practice, using data obtained from the relevant industries.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Hangjung Zo https://orcid.org/0000-0002-2892-1659

REFERENCES

1. W. Mougayar, *The business blockchain: Promise, practice, and application of the next internet technology*, John Wiley & Sons, 2016.
2. J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, *Where is current research on blockchain technology?-a systematic review*, PLoS ONE **11** (2016), e0163477.
3. H. Halaburda, *Blockchain revolution without the blockchain?* Commun. ACM. **61** (2018), 27–29. https://doi.org/10.1145/3225619
4. S. Nakamoto, *Bitcoin: A peer-to-peer electronic cash system*, Decentralized Business Review, 2008, 21260.
5. M. Linton, E. G. S. Teo, E. Bommes, C. Chen, and W. K. Härdle, *Dynamic topic modelling for cryptocurrency community forums, in Applied Quantitative Finance*, Springer, 2017, pp. 355–372.
6. K. R. Lakhani and M. Iansiti, *The truth about blockchain*, Harv. Bus. Rev. **95** (2017), 119–127.
7. D. Yermack, *Donor governance and financial management in prominent us art museums*, J. Cult. Econ. **41** (2017), 215–235.
8. M. V. Fernandes and J. R. Verschoore, *How blockchain affects the technological strategy of the financial industry: An analysis*

based on knowledge discovery in text, Fut. Stud. Res. J.: Trend. Stra. **12** (2020), 311–334.
9. S. Miau and J.-M. Yang, *Bibliometrics-based evaluation of the blockchain research trend: 2008–march 2017*, Technol. Anal. Strateg. Manag. **30** (2018), 1029–1045.
10. S. Kim, H. Park, and J. Lee, *Word2vec-based latent semantic analysis (W2v-Lsa) for topic modeling: A study on blockchain technology trend analysis*, Expert Syst. Appl. **152** (2020), 113401.
11. F. Casino, T. K. Dasaklis, and C. Patsakis, *A systematic literature review of blockchain-based applications: Current status, classification and open issues*, Telemat. Inform. **36** (2019), 55–81.
12. H. D. White and K. W. McCain, *Visualizing a discipline: An author co-citation analysis of information science, 1972–1995*, J. Am. Soc. Inf. Sci. **49** (1998), 327–355.
13. X. Han, *Evolution of research topics in Lis between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model*, Scientometrics. **125** (2020), 2561–2595.
14. C. Zhai, A. Velivelli, and B. Yu, *A cross-collection mixture model for comparative text mining*, (Proc. 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA), 2004, pp. 743–748.
15. M. Ponweiser, *Latent Dirichlet allocation in R*, Diploma Thesis, Univ. of Vienna, 2012.
16. D. M. Blei, A. Y. Ng, and M. I. Jordan, *Latent Dirichlet allocation*, J. Mach. Learn. Res. **3** (2003), 993–1022.
17. T. L. Griffiths and M. Steyvers, *Finding scientific topics*, Proc. Natl. Acad. Sci. **101** (2004), 5228–5235.
18. D. M. Blei and J. D. Lafferty, *Dynamic topic models*, (*Proc. 23rd Int. Conf. Machine Learning, Pittsburgh, PA, USA*), 2006, pp. 113–120.
19. A. B. Jaffe, *Characterizing the "technological position" of firms, with application to quantifying technological opportunity and research spillovers*, Res. Policy. **18** (1989), 87–97.
20. J. D. Adams, *Fundamental stocks of knowledge and productivity growth*, J. Polit. Econ. **98** (1990), 673–702.
21. M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, *Blockchain technology: Beyond bitcoin*, Appl. Innov. **2** (2016), 71.
22. F. Glaser, *Pervasive decentralisation of digital infrastructures: a framework for blockchain enabled system and use case analysis*, (*Proc. 50th Hawaii Int. Conf. System Sciences, Waikoloa, HI, USA*). 2017.
23. M. Risius and K. Spohrer, *A blockchain research framework*, Bus. Inf. Syst. Eng. **59** (2017), 385–409.
24. K. Christidis and M. Devetsikiotis, *Blockchains and smart contracts for the internet of things*, IEEE Access. **4** (2016), 2292–2303.
25. Y. Abuidris, R. Kumar, T. Yang, and J. Onginjo, *Secure large-scale e-voting system based on blockchain contract using a hybrid consensus model combined with sharding*, ETRI J. **43** (2021), 357–370.
26. M. Swan, *Blockchain: Blueprint for a new economy*, O'Reilly Media, Inc, 2015.
27. J. L. Zhao, S. Fan, and J. Yan, *Overview of business innovations and research opportunities in blockchain and introduction to the special issue*, Financ. Innov. **2** (2016), 28.
28. M. Alharby and A. Van Moorsel, *Blockchain-based smart contracts: a systematic mapping study*, arXiv preprint. 2017. https://doi.org/10.48550/arXiv.1710.06372
29. M. H. Miraz and M. Ali, *Applications of blockchain technology beyond cryptocurrency*, arXiv preprint, 2018. https://doi.org/10.48550/arXiv.1801.03528
30. D. Koufogiannakis, L. Slater, and E. Crumley, *A content analysis of librarianship research*, J. Inf. Sci. **30** (2004), 227–239.
31. C. Jacobi, W. Van Atteveldt, and K. Welbers, *Quantitative analysis of large amounts of journalistic texts using topic modelling*, Digit. J. **4** (2016), 89–106.
32. S. Dang and P. H. Ahmad, *Text mining: Techniques and its application*, Int. J. Eng. Technol. Innov. **1** (2014), 22–25.
33. K. Hornik and B. Grün, *Topicmodels: An R package for fitting topic models*, J. Stat. Softw. **40** (2011), 1–30.
34. J. C. Campbell, A. Hindle, and E. Stroulia, *Latent Dirichlet allocation: extracting topics from software engineering data*, In *The art and science of analyzing software data*, Morgan Kaufmann, 2015, 139–159.
35. S. I. Nikolenko, S. Koltcov, and O. Koltsova, *Topic modelling for qualitative studies*, J. Inf. Sci. **43** (2017), 88–102.
36. D. M. Blei, *Probabilistic topic models*, Commun. ACM. **55** (2012), 77–84.
37. L. Du, W. Buntine, H. Jin, and C. Chen, *Sequential latent Dirichlet allocation*, Knowl. Inf. Syst. **31** (2012), 475–503.
38. D. J. de Solla Price, *Is technology historically independent of science? A study in statistical historiography*, Technol. Cult. **6** (1965), 553–568.
39. E. Mansfield, *Academic research and industrial innovation*, Res. Policy. **20** (1991), 1–12.
40. E. Mansfield, *Academic research and industrial innovation: An update of empirical findings*, Res. Policy. **26** (1998), 773–776.
41. R. R. Nelson, *Institutions supporting technical advance in industry*, Am. Econ. Rev. **76** (1986), 186–189.
42. W. M. Cohen, R. R. Nelson, and J. P. Walsh, *Links and impacts: The influence of public research on industrial R&D*, Manag. Sci. **48** (2002), 1–23.
43. L. Branstetter, *Measuring the impact of academic science on industrial innovation the case of California's research universities*, NBER, 2003, working paper.
44. A. B. Jaffe, *Real effects of academic research*, Am. Econ. Rev. **79** (1989), 957–970.
45. Z. J. Acs, D. B. Audretsch, and M. P. Feldman, *Real effects of academic research: Comment*, Am. Econ. Rev. **82** (1992), 363–367.
46. J. H. Grossman, P. P. Reid, and R. P. Morgan, *Contributions of academic research to industrial performance in five industry sectors*, J. Technol. Transf. **26** (2001), 143–152.
47. J. E. Jankowski, *Trends in academic research spending, alliances, and commercialization*, J. Technol. Transf. **24** (1999), 55–68.
48. B. Marciniec and J. Gulinski, *Knowledge and technology transfer (KTT) from academia to industry in central European countries: The case of Poland*, Mol. Cryst. Liq. Cryst. **374** (2002), 13–22.
49. A. Lockett, M. Wright, and S. Franklin, *Technology transfer and universities' spin-out strategies*, Small Bus. Econ. **20** (2003), 185–200.
50. S. Yusuf, *Intermediating knowledge exchange between universities and businesses*, Res. Policy. **37** (2008), 1167–1174.

51. M. B. Lieberman, *A literature citation study of science-technology coupling in electronics*, Proc. IEEE. **66** (1978), 5–13.
52. M. R. Darby and L. G. Zucker, *Grilichesian breakthroughs: Inventions of methods of inventing and firm entry in nanotechnology*, Ann. Econ. Stat. (2005), 143–164.
53. Y. W. Chang and M. H. Huang, *A study of the evolution of interdisciplinarity in library and information science: Using three bibliometric methods*, J. Am. Soc. Inf. Sci. Technol. **63** (2012), 22–33.
54. P. Liu and H. Xia, *Structure and evolution of co-authorship network in an interdisciplinary research field*, Scientometrics. **103** (2015), 101–134.
55. D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group, *Preferred reporting items for systematic reviews and meta-analyses: The prisma statement*, Ann. Intern. Med. **151** (2009), 264–269.
56. S. Bird, E. Klein, and E. Loper, *Natural language processing with python: Analyzing text with the natural language toolkit*, O'Reilly Media, Inc, 2009.
57. W. McKinney, *Python for data analysis: Data wrangling with Pandas, Numpy, and Ipython*, O'Reilly Media, Inc, 2012.
58. M. Umer, I. Ashraf, A. Mehmood, S. Ullah, and G. S. Choi, *Predicting numeric ratings for Google apps using text features and ensemble learning*, ETRI J. **43** (2021), 95–108.
59. E. S. Negara, D. Triadi, and R. Andryani, *Topic modelling Twitter data with latent Dirichlet allocation method*, (International Conference Electrical Engineering and Computer Science, Batam, Indonesia), 2019, pp. 386–390.
60. D. O'Callaghan, D. Greene, J. Carthy, and P. Cunningham, *An analysis of the coherence of descriptors in topic modeling*, Expert Syst. Appl. **42** (2015), 5645–5657.
61. S. Syed and M. Spruit, *Full-text or abstract? examining topic coherence scores using latent Dirichlet allocation*, (IEEE International Conference on Data Science and Advanced Analytics, Tokyo, Japan), 2017, pp.165–174.
62. M. Steyvers and T. Griffiths, *Probabilistic topic models*, In *Handbook of latent semantic analysis*, Psychology Press, 2007, 439–460.
63. A. Bhadury, J. Chen, J. Zhu, and S. Liu, *Scaling up dynamic topic models*, (Proc. 25th Int. Conf. World Wide Web, Montreal, Canada), 2016, pp. 381–390.
64. E. C. Eze, S.-J. Zhang, E.-J. Liu, and J. C. Eze, *Advances in vehicular ad-hoc networks (vanets): Challenges and road-map for future development*, Int. J. Autom. Comput. **13** (2016), 1–18.
65. S. Huh, S. Cho, and S. Kim, *Managing Iot devices using blockchain platform, 2017*, (19th International Conference Advanced Communication Technology, PyeongChang, Rep. of Korea), 2017, pp. 464–467.
66. X. Wang and A. McCallum, *Topics over time: a non-markov continuous-time model of topical trends*, (Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA), 2006, pp. 424–433.
67. C. Wang, D. Blei, and D. Heckerman, *Continuous time dynamic topic models*, arXiv preprint, 2012. https://doi.org/10.48550/arXiv.1206.3298

AUTHOR BIOGRAPHIES

Jeongho Lee received his BS in materials science and engineering from the Department of Materials Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2013, and his MS degree in business and technology management from the School of Business and Technology Management, KAIST, Daejeon, Republic of Korea, in 2016. Since 2016, he has been at the School of Business and Technology Management, KAIST, Daejeon, Republic of Korea, where he is now a PhD candidate. His main research interests are blockchain, cryptocurrency, and text mining.

Hangjung Zo is a professor of digital innovation management at the School of Business and Technology Management, KAIST, Daejeon, Republic of Korea. He obtained his PhD in IT management from the University of Wisconsin-Milwaukee, USA. His research interests include digital social innovations and the impacts of digital technology on business and society. His papers have appeared in the IEEE Transaction on Systems Man and Cybernetics, Decision Support Systems, Journal of Business Research, Electronic Commerce Research and Applications, and Computers and Education, among others. He is the director of the Center for Science-based Entrepreneurship at KAIST.

Tom Steinberger is an Assistant Professor at the School of Business and Technology Management, KAIST, Daejeon, Republic of Korea. He received his PhD from the University of California, Irvine, USA. His current research focuses on the strategic management and organization of data and intelligent technologies.

How to cite this article: J. Lee, H. Zo, and T. Steinberger, *Exploring trends in blockchain publications with topic modeling: Implications for forecasting the emergence of industry applications*, ETRI Journal **45** (2023), 982–995. DOI 10.4218/etrij.2022-0257