國立中興大學科技管理研究所
Graduate Institute of Technology Management
National Chung Hsing University

碩士學位論文
A Thesis Submitted in Partial Fulfillment
of The Requirements for The Degree of Master

利用主題式專利探勘模型探索區塊鏈新興技術趨勢

Identifying Emerging Trends of Blockchain Technology Using a Topic-based Patent Mining Model

指導教授：王瑞德博士 Juite Wang, Ph.D.

研究生：劉嘉鴻 Chia-Hung Liu

中華民國一百零八年六月
June 2019

國立中興大學 科技管理 研究所

碩士學位論文

題目： 利用主題式專利探勘模型探索區塊鏈新興技術趨勢

Identifying Emerging Trends of Blockchain Technology

Using a Topic-based Patent Mining Model

姓名： 劉嘉鴻 學號： 7106026108

經 口 試 通 過 特 此 證 明

論文指導教授 王瑞德 王瑞德

論文考試委員 王瑞德 王瑞德

王宏鍇 王宏鍇

蘇信寧 蘇信寧

中華民國 108 年 06 月 17 日

摘要

自從中本聰於 2008 年了發表了區塊鏈的概念，人們對於這項技術的興趣與日俱增，尤其在 2017 年末隨著比特幣價格的大幅上漲之後，更成為新聞媒體的新焦點。然而區塊鏈不只能運用於金融機構，物流業、農業也盼望著利用這項新技術來重塑其作業流程，因此了解區塊鏈的技術趨勢以及制定合適的技術或商業策略，對於有意利用區塊鏈的公司來說相當重要。本研究主要利用潛在狄利克雷分佈並搭配名詞片語的特徵抽取，來挖掘區塊鏈相關專利下潛藏的技術主題。為了發掘技術新興趨勢，我們更進一步利用主題模型所產生的文詞矩陣來衡量各個主題的成長率與專利數量；另外也結合專利中其他結構化資料，如：專利申請人、向前引用數、專利申請範圍等，來描述區塊鏈的技術競爭態勢、技術主題價值以及技術演進地圖。本研究結果揭露區塊鏈領域技術的發展趨勢、透過主題與申請人的網絡圖發現各公司專注於不同的技術主題。最後有意發展區塊鏈的公司，可經由結合分析結果的專利地圖來輔助公司制定技術策略並提升決策效率。

關鍵字：隱含狄利克雷分佈、專利分析、區塊鏈、熱門主題

Abstract

Since Satoshi Nakamoto coined the Blockchain idea, the interest of this technology has been increasing, especially at the end of 2017 as the price of Bitcoin rose dramatically. Not only financial institutes but other industries, such as logistics and agriculture, look forward to using Blockchain technology to redesign their services. Hence, it becomes fairly crucial for these companies to comprehend the emerging trends and propose appropriate development strategies. In this research, we apply a latent Dirichlet allocation (LDA) model with noun phrase extraction to identify the underlying structure of collected patent corpus related to Blockchain technology. In order to uncover the emerging topics, we go one step further by utilizing the document-topic matrix generated from LDA to weigh each topic popularity and growth rate. Also, integrating the unstructured data such as applicants, forward citations, claims to illustrate the competitive landscape, patent value and technology evolution map. Our results reveal the future trends in the Blockchain technology field and find major firms have been focused on different technology topics. Finally, through the patent map combining the analyzed outcomes can help Blockchain companies formulate the patent strategy and facilitate decision process.

Keywords: Latent Dirichlet allocation, Patent analysis, Blockchain, emerging topics

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. Introduction

The concept of Blockchain was developed in 2008, which aims to provide a peer-to-peer version of electronic currency allowing direct payment without going through the financial institution (Nakamoto, 2008). Despite the potential of being applied into diverse applications beyond cryptocurrency, it still remains doubtful about commercial viability of Blockchain, and several studies have investigated for technology challenges it faced (Yli-Huumo, Ko, Choi, Park, & Smolander, 2016; Zheng, Xie, Dai, Chen, & Wang, 2018). However, in recent years, this promising technology has been gaining increasing attention by integrating with existing business processes, including financial transactions, enterprise resource planning, supply chain management, and so on. Instead of taking the technical aspect to examine Blockchain technology, this research proposes an application-oriented classification through patent analysis to discover the underlying topics to identify emerging trends in Blockchain technology for exploring potential technological opportunities.

As a carrier of institution's technological knowledge, patents are one of the most important indicators for detecting technology trends (Zhang, Li, & Li, 2015). According to the statistics from World Intellectual Property Organization (WIPO), 90 ~ 95% of world's inventions can only be found in patented documents and 80% of these techniques do not appear in other professional articles (Liu & Shyu, 1997). However, due to information overload, it remains a great challenge for humans to deal with the huge collection of textual data. Text mining and natural language process (NLP) (Tseng, Lin, & Lin, 2007) comes in handy to tackle this kind of problems to discovery important information from patent corpus.

Patent analysis has been applied for technology planning and forecasting (Liu & Shyu, 1997). Among the relevant studies, technology classification (Fall, Törcsvári, Benzineb, & Karetka, 2003) is one of the considerable issues to finding emerging trends or hot topics. Some of the studies employed citations linkage to build the citation network (Erdi et al., 2013), but this approach may not be suitable for emerging technology, due to limited number of citations available for patents issued in recent years. Hegde and Sampat (2009) also examined the citations to a patent, showing firm's citation choices are strategic likely. On the other hand, International Patent Classification (IPC) can be considered as a popular classification system, and several studies utilized patent classification codes for clustering patents (Teh, Jordan, Beal, & Blei, 2006). While the patent classification code are too general and function-oriented rather than application-oriented to adequately classify technologies for a specific domain, especially for an emerging technology (Hu, Fang, & Liang, 2014).

To overcome the limitation of previous studies, this research applies the latent Dirichlet allocation (LDA) topic model (Blei, Ng, & Jordan, 2003) to discover the latent topics underlying in patent corpus related to Blockchain technology. LDA assumes that each document is a mixture of topics and each topic is a mixture of terms by sampling from prior Dirichlet distribution, using term co-occurrence to calculate the conditional probability. With the result of LDA, this research further applies the generalized linear mixed model (GLMM) (Breslow & Clayton, 1993) that uses the topic assignment on each patent to dig out the emerging topics, performs competitive analysis to identify the relationships between topics and institutions with network analysis (Stanley & Faust, 1994), and conducts technology value analysis that utilizes the citation, patent family and claims to evaluate market value of each patent. Then, we compute the cosine similarity between topics to cluster similar topics, and label the sub-technology by co-occurrence phrases. Through the time series analysis on sub-technologies, technology evolution roadmap can be obtained to understand the technology development trend.

The three primary objectives of this research are (1) to uncover the latent topics under Blockchain technology through related patents, (2) to further analyze the competitive landscape by the topic growth trends and the topic-applicant network, and calculate the market value according to the structured data of patents of each topic, and (3) to construct the technology evolution map via clustering similar topics to help firms discover business opportunities, and facilitate corporate technology strategy formulation.

This research is organized as follows. Chapter 2 discusses the related works on blockchain and patent mining. Chapter 3 describes the data collection and research methodology. Chapter 4 presents the results of Blockchain patent mining. Finally, Chapter 5 concludes this research and discusses the limitations and future works.

2. Literature Review

2.1. Patent Mining

Patents are important intellectual properties of protecting interests of individuals, organizations and companies, and they are also useful for competitive analysis, technological trend analysis, or strategy planning (Abraham & Moitra, 2001). With well-defined format including structured items (like patent number, applicant, and published date) and unstructured items (like claims, abstracts, and description), patents documents are fairly different from web documents and offer more informative message. In addition, because the patent filing process is highly costly and time consuming, patents generally mean there is some potential in economic or technical development (Basberg, 1987). However, a patent document is often lengthy and filled with a lot of technical and legal terminologies, which requires considerable amount of human efforts for patent analysis. Also, it's necessary to have specific domain expert for patent analysis to information retrieval, result interpretation, etc. (Zhang et al., 2015).

Traditional patent analysis is mainly focused on the structured bibliography data in patent documents because it has a standard format that is easier to analyze. As a result, bibliography is the most frequently adopted toll in patent analysis and has been applied to forecast technology trends and asses technology values (Guo, Xu, Huang, & Porter, 2012). Due to technological change is cumulative and path-dependent, one of the appropriate indicators is citation to measure knowledge accumulation and diffusion (Daim, Rueda, Martin, & Gerdsri, 2006; Giovanni, 1982).

Owing to the dramatic increasing of patents, text mining techniques are also required to capture the key information pervading in a large amount of patent documents. Tseng et al. (2007) summarized the typical patent analysis process using both structured and unstructured data format and demonstrate a series of text mining techniques, including text segmentation, feature selection, cluster generation, and so on. Several studies have identified unusual patents as potential technology opportunity. Yoon and Kim (2012a) proposed a subject-action-object (SAO) based semantic patent and calculated the semantic similarities between patents to detect outlier as technological opportunities.

Geum, Jeon, and Seol (2013) represented collected patents with frequency keyword vectors and used novelty detection to identify potential opportunities in near future.

Different techniques have been studied to classify patents, including k-nearest neighbors, Naive Bayes, and support vector machines (SVMs) (Fall, Torcsvari, Fievet, & Karetka, 2004). However, using the IPC or Derwent code (DC) for co-word analysis or classification criterion is more function-oriented and relatively coarse for emerging technologies (Wang, Liu, Ding, Liu, & Xu, 2014). Owing to the limitation, topic-model based method have become increasingly attractive to researchers (Venugopalan & Rai, 2015). Latent Dirichlet allocation, developed by Blei et al. (2003), is a Bayesian topic model using unsupervised learning approach to discover the latent semantic topics in massive documents, and numerous extensions to the standard LDA model were published, such as dynamic topic models, hierarchical Dirichlet process (Teh et al., 2006), and correlated topic models (Blei & Lafferty, 2007).

Griffiths and Steyvers (2004) was the first to demonstrate the LDA topic model and used of Gibb sampling to analyze scientific articles from various disciplines. Chen, Zhang, Zhu, and Lu (2017) applied patent titles and claims as the main data source to forecast the technology trend by fitting univariate quadratic polynomial, and mark the trend tuning points. Greenville, Dickman, and Wardle (2017) undertook the topic model on 75 years of dryland scientific article, further cluster the 25 topics into seven groups on word-base similarity, and calculated the research gaps between each topic visualizing in network (Westgate, Barton, Pierson, & Lindenmayer, 2015). In addition, the advanced textual data preprocessing can apply to LDA before topic modeling, Wang et al. (2014) defined the noun phrase extraction rule to collect informative terms in the corpus and employed one more level in LDA structure.

In practice, LDA has been proved to be more accurate compared to conventional statistic methods, more interpretable through the featured terms in each topic, and easy to apply. Recently, there is an increasing number of studies utilized LDA on analyzing patent data. However, compared to former studies, few researches take a detailed look on hyperparameter of LDA, and the Blockchain technology field is relatively new and more specific than other studies, further research is required.

2.2. Blockchain Technology

This research focuses on Blockchain technology which was came up by Nakamoto (2008), a distributed database solution that maintains a growing list of data records confirmed by the node participant in it. In January of 2009, Satoshi Nakamoto realized the Blockchain idea and introduced Bitcoin as the first peer-to-peer distributed trustless electronic cash, in order to replace the traditional centralized payment system. Although Blockchain is mostly known as the technology behind Bitcoin cryptocurrency by the public, Blockchain is also a suitable solution for company security, integration, cost problem; and Bitcoin is only one of the Blockchain applications to deal with the peer-to-peer transaction.

Through the rapid development of cryptocurrency and the underlying Blockchain technology, another term "smart contract" (Szabo., 1996) was also introduced to the Blockchain field and implemented on Ethereum[1], which is the first public blockchain platform that supports customized smart contracts. This new technology was seen as "Blockchain 2.0", which enables Blockchain to go beyond simple buy/sell cryptocurrency transactions and make it available to fit more business applications through decentralized applications (DApps); e.g., digital asset management, crowdfunding, gambling, etc. Firms can reinvent their services by taking advantage of Blockchain technology with the key features including: decentralization, immutability, transparency, efficiency, security, and anonymity. To date, Blockchain technology has been a popular issue for financial institutions, but it also brings benefits to other areas, e.g., supply chain activities (Kshetri, 2018), Internet of Things (IoT) devices (Christidis & Devetsikiotis, 2016), healthcare data processing (Yue, Wang, Jin, Li, & Jiang, 2016), and so on.

Because the Blockchain technology research is still at the very early stage, about 60 % of Blockchain papers applied case studies in specific domain to point out the advantage of employing this technology, such as supply chain management (Kshetri, 2018), healthcare (Alonso, Arambarri, Lopez-Coronado, & Diez, 2019) etc., while 30% of them used a technical view, for example to analyze the different structure of Blockchain system (Zheng, Xie, Dai, Chen, & Wang, 2017), to compare the performance between

[1] Ethereum. https://www.ethereum.org/.

consensus algorithms (Mingxiao, Xiaofeng, Zhe, Xiangwei, & Qijun, 2017) or to design Blockchain architecture to integrate current technology such as Internet of Things (IoT) (Sharma, Chen, & Park, 2018). Among them, the case study approach is the most popular method used by researchers. To the best of our knowledge, there has not been any patent mining of Blockchain technology.

In order to investigate the Blockchain technology development and its applications, several studies have been conducted. Yli-Huumo et al. (2016) extracted 41 primary papers from scientific databases and classified them into nine categories to identify current status of Blockchain technology and recommend on future research directions. Casino, Dasaklis, and Patsakis (2019) also provided a systematic review across multiple domains using research papers published in high ranked journals to investigate various research gaps and future exploratory directions. One another approach is taking more technical view to analyze Blockchain technology, Zheng et al. (2017) laid out possible trends of Blockchain through comparing different consensus system and algorithm. Besides, Table 2.1 lists the related studies on future trends and challenges on Blockchain. Nevertheless, it is worth noting that Blockchain technology is still an underdevelopment technology and have several limitations. Swan (2015) listed seven technical challenges, including throughput, latency, size and bandwidth, security, wasted resources, usability, and versioning, hard fork, multiple chains, for the adoption of Blockchain technology in the future.

Despite of valuable contributions of above studies, there are still some limitation of analyzing future technical opportunities. First of all, previous researches mainly used bibliometric analysis using scientific papers (Casino et al., 2019; Dabbagh, Sookhak, & Safa, 2019), however as we mentioned before, patent data also have immense technical details and applications that has not been cover in professional articles. Second, previous researches primary used classification code in Web of Science or classify documents manually through reading whole content. However, these approaches cannot deal with more detailed information about technical attributes described in patent documents and handle a large collection of documents. Hence, it is essential to propose another perspective of utilizing patent mining to uncover the underlying technological issues and further discovery technical opportunities.

Table 2.1: Related works on future trends and challenges of Blockchain

Author	Challenges/ Trends
Yli-Huumo et al. (2016)	• To be ready for pervasive use of Blockchain technology, scalability issues like latency and throughput have to be addressed. • More Blockchain based applications beyond Bitcoin and cryptocurrency are essential to leverage Blockchain technology, such as smart contract, property licensing, voting etc.
Zheng et al. (2017)	• The tendency of centralization on Bitcoin network, resulting in 51% mining attacks to threaten Blockchain security. • Combing with big data including data management and data analytics to further secure private data and predict potential user behavior.
Zheng et al. (2018)	• Smart contract applications including develop smart contract itself and smart contract platform like Ethereum. • Establish Blockchain testing criteria to valid if Blockchain works fine as developers claim.
Lu, Huang, Azimi, and Guo (2019)	• Cross-chain technology to integrate different independent Blockchain network and realize digital asset transfer between multiple blockchains. • Hybrid consensus mechanism to improve Blockchain efficiency and security.
Casino et al. (2019)	• Different consensus algorithm or architecture to achieve sustainability and address potential centralization problem. • Quantum resilience hash algorithms to avoid be easily cracked, especially in public key cryptography. • Data management and privacy and security solutions are in demand such as traceability of transactions and smart contract operations.
Wang et al. (2019)	• Formal verification on smart contract application, especially on Ethereum virtual machine (EVM). • Layer 2 technology namely off-chain contract to further handle complex logic execution and implement cross-chain service.

3. Data and Methodology

This chapter describes the research framework consisting of four steps: patent retrieval, textual data feature extraction, topic modeling and competitive and technology value analyses by integrating topic modeling results and structured data, as shown in Figure 3.1. (Wang, 2018).

3.1. Data

This research uses patent titles and abstracts as the primary data source for topic modeling, accessed via the Global Patent Search System (GPSS) collecting three main patent databases, including European Patent Office (EPO), World Intellectual Property Organization (WIPO), and United States Patent and Trademark Office (USPTO). At the same time, we also retrieve the structured items of each patent, e.g., publication date, citation, patent family, to tailor data analysis with topic modeling results. In order to search for all the relevant patents in the Blockchain technology field, the search keywords were essential to define our research scope. After several pilot search, we decided to use "Blockchain" as the only key search term with the wild card and "Block chain" to avoid machine-translation problem, searching for the title, abstract, and claim fields, to extract Blockchain related technology and not extend to the general encryption technology. The same search strategy was also used in studies by Yli-Huumo et al. (2016) and Casino et al. (2019).

Our research returned 2,029 patents through EPO, WIPO, and USPTO from 2015[2] to May 2019. After carefully examining the content of patent manually, filtering out the duplicated patents according to the title and abstract of patents, removing common phrases and combining similar phrases (the detailed textual data preprocessing was described in next section), the final dataset includes 1397 patents and 1237 shared phrases.

---

[2] The first patent using 'Blockchain' term is "Method to securely establish, affirm, and transfer ownership of artworks" published at 2015/2/26.

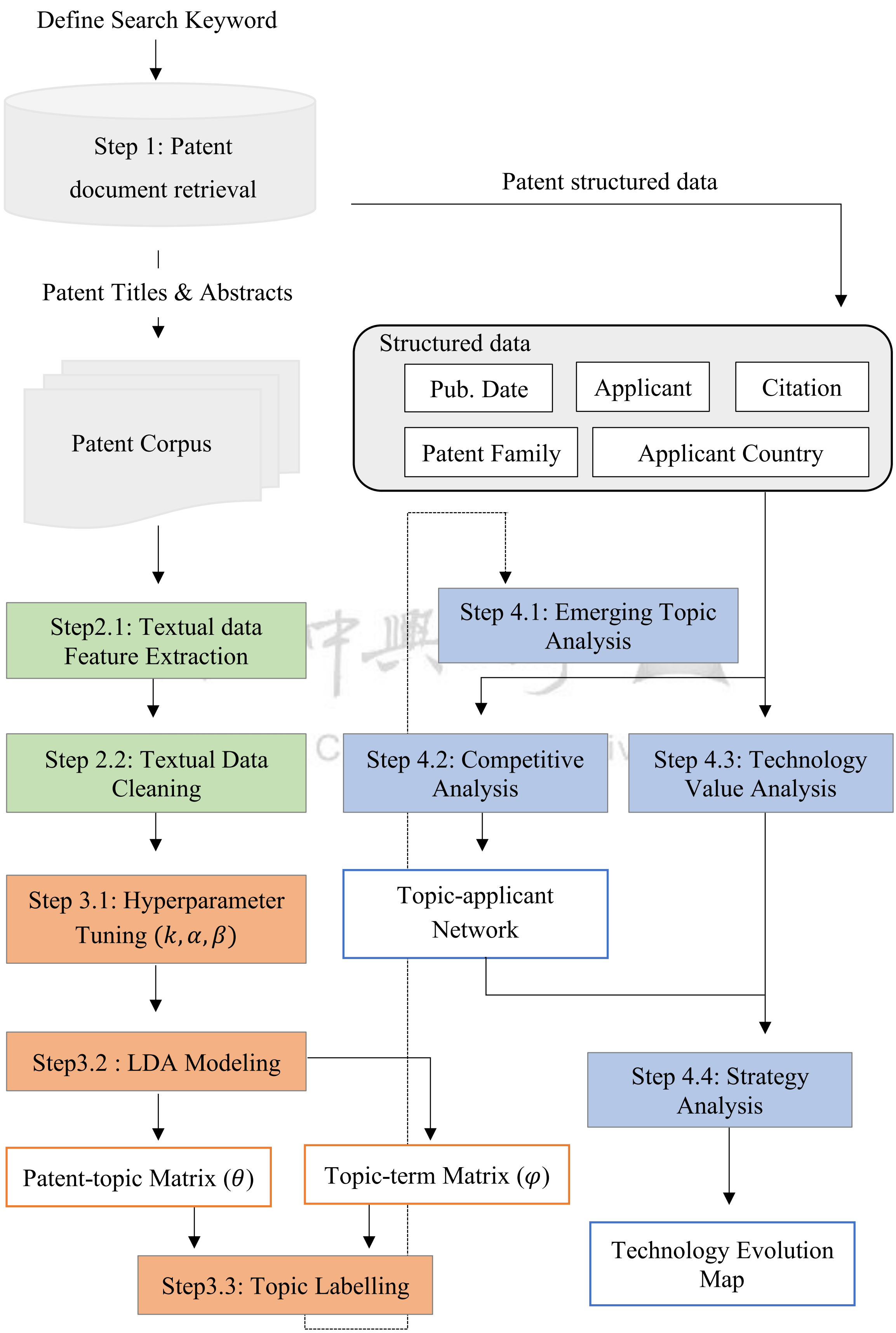

Figure 3.1: Research framework

3.2. Methodology

3.2.1. Textual data feature extraction

We started the natural language process by removing non-English words with regular expressions. Besides, all the hyphen were replaced with space to identify the technical terms and the abbreviations were switched to the original terms, for example, “internet-of-things” was transformed to “internet of things” and “bp” was transformed to “business process”, and all the words in the corpus were converted into lower-case.

Different from ordinary text mining process, our research introduced the noun phrase extraction process. Instead of using all the words in the corpus which usually be less informative, especially in the patent literature, noun phrases extraction can gain the primary information of a sentence according to the construction of a sentence in linguistics (Wang et al., 2014). Furthermore defining the set of keyword or key phrase patterns heavily relies on the efforts of experts, which may be expensive or unavailable (Yoon, Choi, & Kim, 2010). Thus, after basic transformation on textual data, this research utilized textual data preprocessing with lemmatization, the part-of-speech tagging (Straka, Hajic, & Strakova, 2016) and the noun phrase rule to automatically extract the useful phrase from titles and abstracts of the collected patent documents.

In this research, we defined the noun phrase extraction rule as follows (Straka & Strakova, 2017) with a regular expression:

- The length of the noun phrase is limited to a minimum of 1 and a maximum of 3
- The noun phrase pattern is (A|N)*N(P+D*(A|N)*N)*, where N represents noun or proper noun, A represents adjective, P represents preposition, D represents determiner, plus mark represents one or more times, and an asterisk represents zero or more times. All possible combinations in our research as follow:
  1. Noun
  2. Noun + Noun
  3. Noun + Noun + Noun
  4. Adjective + Noun
  5. Adjective + Noun + Noun
  6. Noun + Preposition + Noun
  7. Noun + Determiner + Noun

Afterward, owing to the specific characteristics of patent document that usually consists of transition phrase, such as comprising, having, including, consisting of, and so forth, it is necessary to exclude these common words or phrases of patent terminology (Yoon & Kim, 2012b). Also, excluding the high-frequency words is crucial to gain meaningful information from documents. We applied the stop words list from SMART dictionary (Lewis, 2004), transitional phrase page on Wikipedia (2014), and general academic words list (Haywood, 2003) as follow:

- Stop words such as the, that, which;
- High-frequency words in the patent document such as comprising, claim, consist;
- General academic words such as research, data, access.

There are two things that worth mentioning: First, we need to examine every phrase in our noun phrase list carefully instead of removing all the phrases including one of the above term, for example, the 'patient data' may important even having general academic words. Second, the removed terms or phrases are excluded from the noun phrases list; namely, we did not vectorize the patent documents or generate document-term matrix yet.

Furthermore, we calculated the TF-IDF(term frequency inverse document frequency) value of each phrase to identify more essential phrases pervading in patent documents, rather than general phrases appearing in every patent. In this research, we remained the phrases with TF-IDF value higher than the median one, and also keep the meaningful phrase after evaluating every excluded phrase. The following list shows some examples of the phrases we excluded and remained:

- Too general phrases or words such as blockchain, transaction, public key;
- Meaningful phrases such as smart contract, digital signature.

On the other hand, patent documents are sensitive to regions and can be described by different languages, and international patents are required to be written by the language accepted worldwide (Zhang et al., 2015). Although machine translation can address the cross-language problem and achieve acceptable performance, sometimes it may have a different expression, especially on technical terms; for example, "Blockchain" would be "block chain" through machine translation. It is imperative to tackle down in text mining project, as well as the same problem takes place on abbreviation. Again, we exploited regular expression to revert the correct translation and combine the same meaning phrases.

3.2.2. Latent Dirichlet allocation

The topic model technique we used is based on a generative probabilistic model named latent Dirichlet allocation (LDA) (Blei et al., 2003), which has been thoroughly explained in the original paper. The LDA model can be denoted as a graphical model in plates notation which mainly following the Blei and Lafferty (2009). The intuition behind LDA is using the co-occurrence words in a collection of document to infer the underlying topic. Figure 3.2 clearly shows that LDA structure is a three-level hierarchical Bayesian model, and math notations are described in Table 3.1.

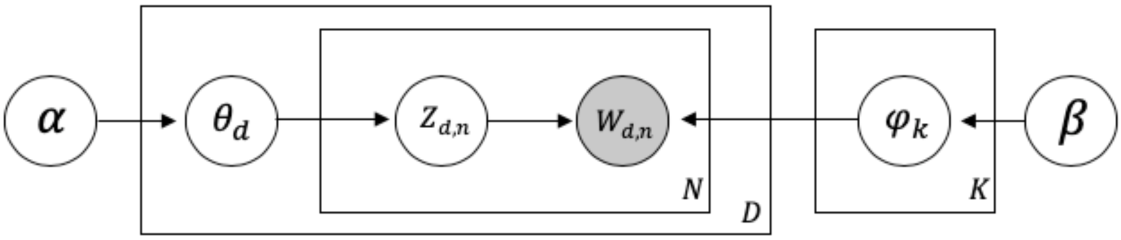

Figure 3.2: Plate notation of LDA

Table 3.1: List of math notation of LDA

Notations	Meaning
K	Number of topics
D	Number of documents
N	Number of words in document
V	Number of words in vocabulary
α	-dimensional parameter vector for Dirichlet prior
β	-dimensional parameter vector for Dirichlet prior
θ	Document-topic matrix
φ	Topic-term matrix
Z	Assigned topics over word s
W	Assigned words

The basic idea of LDA is that each document in corpus is represented as random mixtures over latent topics, where each topic probability is chosen by a Dirichlet distribution. The probabilistic generative process is defined as:

- For each topic, draw a distribution over words $\theta_d \sim Dir(\alpha)$
- For each document, draw a distribution over topics $\varphi_k \sim Dir(\beta)$
- For each word of document:

1. Draw a topic assignment $Z_{d,n} \sim Mult(\theta_d)$
2. Draw a word $W_{d,n} \sim Mult(\varphi_{z,n})$

Obviously, three hyperparameters[3] need to be defined before LDA generative process, namely and . For the Dirichlet distribution parameters α and β, in LDA symmetric prior are used, which means that all topics have the same chance of being assigned to a document and all words have the same chance of being assigned to a topic. Moreover, lowing the values of and will result in more decisive topic associations, thus and will become more sparser (Heinrich, 2008). Figure 3.3 displays the different symmetric parameter of Dirichlet distribution with two-simplex visualization of Dirichlet distribution density plots generated by α = (3, 3, 3), (1, 1, 1), (0.3, 0.3, 0.3) from the left side to the right.

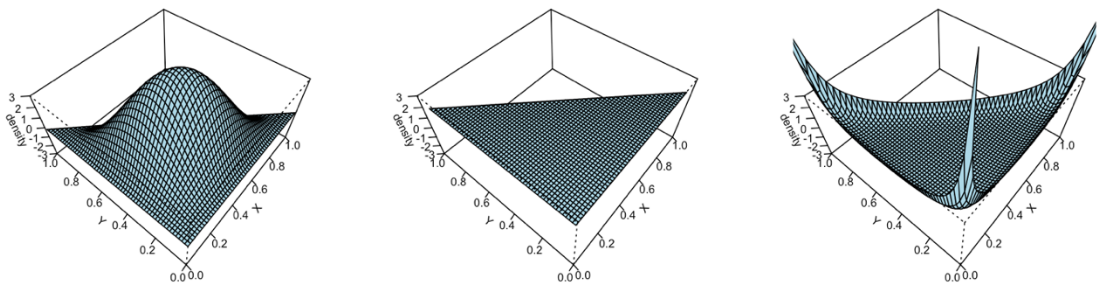

Figure 3.3: Two-simplex (K = 3) visualization

On the other hand, optimal topic number, namely K hyperparameter in LDA, is also a critical issue. So far, There are several metrics have been used in topic modeling: the most well-known perplexity (Rosen-Zvi, Griffiths, Steyvers, & Smyth, 2004), empirical likelihood (Li & McCallum, 2006), and marginal likelihood (Griffiths & Steyvers, 2004). In this research , we use Akaike information criterion (AIC) (Burnham & Anderson, 2016) to measure the model quality to determine the optimal topic number. The equation of AIC is shown below,

$$AIC = 2k - 2ln(L) \quad (1)$$

where L represents the number of estimated parameter in the fitted model and represents the maximum likelihood of $p(w_d)$.

The intuition behind the math equation is that the difference between fitted model and maximum likelihood model is presented as relative entropy, meaning the amount of lost

[3] Hyperparameters are parameters that are not directly optimized through learning process.

information. In addition, with the k parameter in the AIC model, which can be seen as regularization in linear regression, we could avoid overfitting in our model by selecting the minimum point.

3.2.3. Emerging topics

After topic modeling, we have the distribution of each patent over topics (, document-topic matrix) and the distribution of each topic over words (, topic-term matrix). To identify the emerging trends (topics), the number of published patents per topic over time (months) can be utilized. Different from Rosen-Zvi et al. (2004) applying linear model on mean- and year scale, this research used the Generalized Linear Mixed Model (GLMM) (Bates, Machler, Bolker, & Walker, 2015) with random intercept and random slope that can capture the hierarchical characteristics and more close to the multilevel concept of LDA (Barr, Levy, Scheepers, & Tily, 2013).

$$Y_{mt} = \beta_0 + S_{0m} + (\beta_1 + S_{1m})X_t + e_{st} \qquad (2)$$

The equation of GLMM we applied on our research as above, where the subscript t represents the group level meaning the topics assigned over patents, the subscript m represents the time scale that we used month here, and subscript 0 represents the fixed effects and 1 represents the random effects. The model would estimate the fixed effects on intercept and slope of month, and random intercept and random slope for each topic. Moreover, we also used the Bayesian prior to avoid singular fit[4] (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013). Thus, topics with positive random intercepts can be interpreted as having more than average published patents and topics with positive random slope can be interpreted as having more than average growth in the observed time period (Greenville et al., 2017).

3.2.4. Competitive analysis

After topic modeling, each document was treated as a vector of phrases and described as a mixture of topics, and topic-term matrix makes it possible for us to capture the gist of topics and interpret them into understandable and semantic label through the high probability phrases. With reasonable and meaningful labels, we can utilize structured data

[4] Singular fit: https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

like the applicants of patent to present the competitive circumstances against each country and determine the relationship between applicant and each topic.

However, due to the assumption of LDA, which each document has non-zero probability over each topic through sampling from Dirichlet distribution, the applicant-topic network need to be adjusted by edge weight and remove nodes with a threshold of degree centrality. Figure 3.4 shows the topic-applicant network without filtering out any nodes and edges, and green nodes represents topics, pink nodes represents applicant, and purple edges represent the relationship between topics and applicants. Here, ForceAtlas2 layout (Jacomy, Venturini, Heymann, & Bastian, 2014) which is highly efficient and better performance than traditional layout options, is adopted to highlight the characteristics of network, such as the nodes with "Bridge" feature filling a structure hole providing the link between two individuals or clusters (Granovetter, 1973).

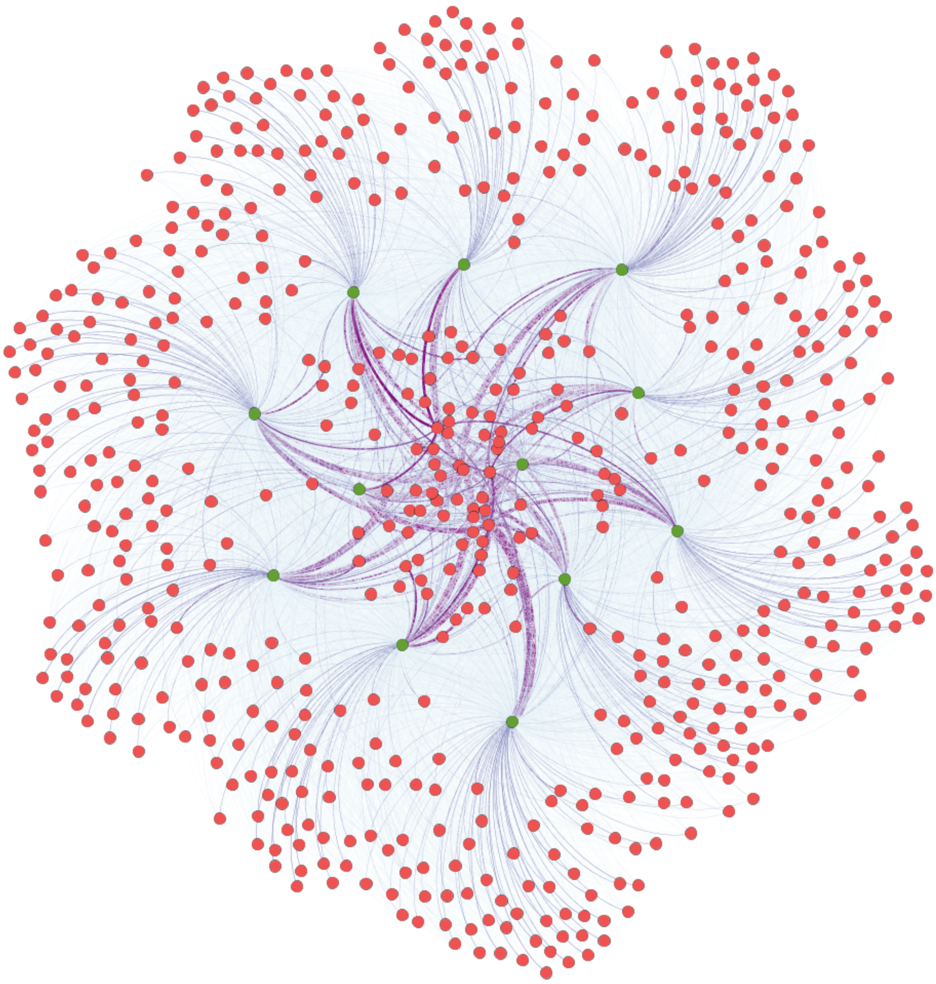

Figure 3.4: Topic-applicant network

3.2.5. Technology value analysis

In order to support decision making and formulate company strategy through patent documents, it is imperative to access the quality of patent to identify the important and valuable topics. The patent valuation can provide a useful measurement for comparing different techniques, in our research, there are three metrics used to evaluate the potential value of each patent, including forward citations, claims, and patent family. The potential value of a patent were calculated as follow, here we also took natural logarithm on citations and patent family (Reitzig, 2004) owing to the right skew characteristics in our dataset:

$$\begin{aligned} &Technology\ value \\ &\quad = z(citation\ count, year) + z(\ln(claims\ count)) \\ &\quad + z(\ln(patent\ family\ size)) \end{aligned} \tag{3}$$

$$f(citation\ count, year) = \frac{citation\ count}{the\ average\ citation\ count\ of\ that\ year} \tag{4}$$

$$z(x) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

First, patent citations may convey two major aspects of innovation through linkage between inventions, inventor or assignees, and as an indicator of the importance of individual patents (Hall, Jaffe, & Trajtenberg, 2005). We take the latter aspect to measure the value of each patent and adjusted the forward citations as above because it is affected by publication year (Breitzman & Thomas, 2015). Second, the patent family, which is a collection of patent applications covering the same or similar technical content, is a different approach to gauge patent value (Kabore & G, 2016). We normalize INPADOC extended patent family[5] size as the second indicator. Third, claims of patents define the scope of protection conferred by a patent. According to Lemley and Shapiro (2005) research, the number of claims is an crucial determinant of patent value, because claims are expensive to draft and prosecute and the more claims a patent has the more difficult it will be to invalidate the patent. Last, we normalize the number of claims for each patent,

[5] INPADOC patent family: https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/patent-families/inpadoc.html

sum the three metrics to measure the potential value of each patent, and multiple with the document-topic matrix to gauge the potential market value of each topic.

3.2.6. Technology evolution map

After topic modeling, the topic-term matrix and the document-topic matrix were generated, which can be utilized to cluster the similar topics to demonstrate technology evolution map. Hierarchical clustering of multiscale bootstrap resampling with p-value (Suzuki & Shimodaira, 2006) based on topic-term matrix was applied here to assess the uncertainty of each cluster. Figure 3.5 displays the example of hierarchical clustering, providing approximately unbiased (AU) p-value in red and bootstrap probability value in green, and highlighting clusters with high AU p-value in red rectangle.

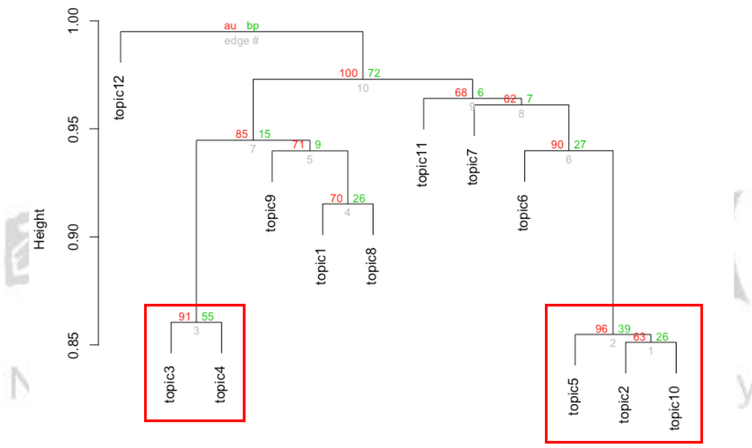

Figure 3.5: Example of hierarchical clustering with p-value

Moreover, the clustering result can further validate the topic label and label each clusters through the co-occurrence phrases of topics in same cluster. The cosine similarity is also employed in our research to calculate the distance to present the clustering via topic vector with the following formula:

$$cosine\ similarity(Topic_1, Topic_2) = \frac{Topic_1 \times Topic_2}{|Topic_1| \times |Topic_2|} \quad (6)$$

With dendrogram results and co-word analysis, we can construct the technology evolution map according to the patent publication date and then explore the technology diffusion pattern for each cluster to understand technology development trends in distinct clusters.

4. Results

Our goal is to find the emerging trends in the Blockchain technology field from various latent topics from the collected patent corpus, and demonstrate the competitive landscape, potential market value, and technology diffusion patterns through the developed methodology. In this chapter, using the corpus collected from the search keyword, we present the topic model after textual data preprocessing. Altogether, we found 1,237 unique noun phrases within a total of 1,397 patents from Jan 2015 to May 2019 in EPO, WIPO, and USPTO databases.

4.1. Topic modeling

As mentioned before, the optimal number of topics and prior Dirichlet distribution play a critical role in LDA. Zhao, Chen, and Zou (2015) and W. Zhao et al. (2015) suggested that the sweet spot of Dirichlet parameter is around 0.01. We take into account the characteristic of our corpus, more focusing on a specific field comparing to other patent analysis using LDA model in the related works, like Griffiths and Steyvers (2004) research using scientific paper in PNAS and Chen et al. (2017) using Australia patents. Therefore, adopting a relatively lower value on α and β is more suitable in our research and be more interpretable. After determining range of the prior parameter, we applied Akaike information criterion (AIC) as a model quality estimator, and chose the minimum AIC value point of topic number K to deal with the trade-off between the overfitting and the simplicity of the model. Here, the grid search was used to perform the hyperparameter tuning, and the grid of our model for K is from 2 to 40 with one step, and α and β were from 0.01 to 0.1 with 0.1 step. For readability, Figure 4.1 only shows the gird search process of three parameter setting (0.01, 0.05 and 0.1) with fixed α and β, and the minimum AIC value of topic number K is around 10 to 15. Also, the top 10 terms ranked by probability of all topics are listed in Table A.1. However, the shortcoming of LDA or other text mining techniques is that the model validation is mainly subjective, in order to enhance the robustness of our research, we interviewed expert working in Fintech domain more than five years to evaluate the classification results with questionnaire in Appendix C. After carefully examined the terms and corresponding patent documents for each topic, and revised the labels according to the advices of experts, we labelled the 12 topics as shown in Table 4.2. The configuration of the crucial parameters in LDA (Grun & Hornik, 2011) is described in Table 4.1.

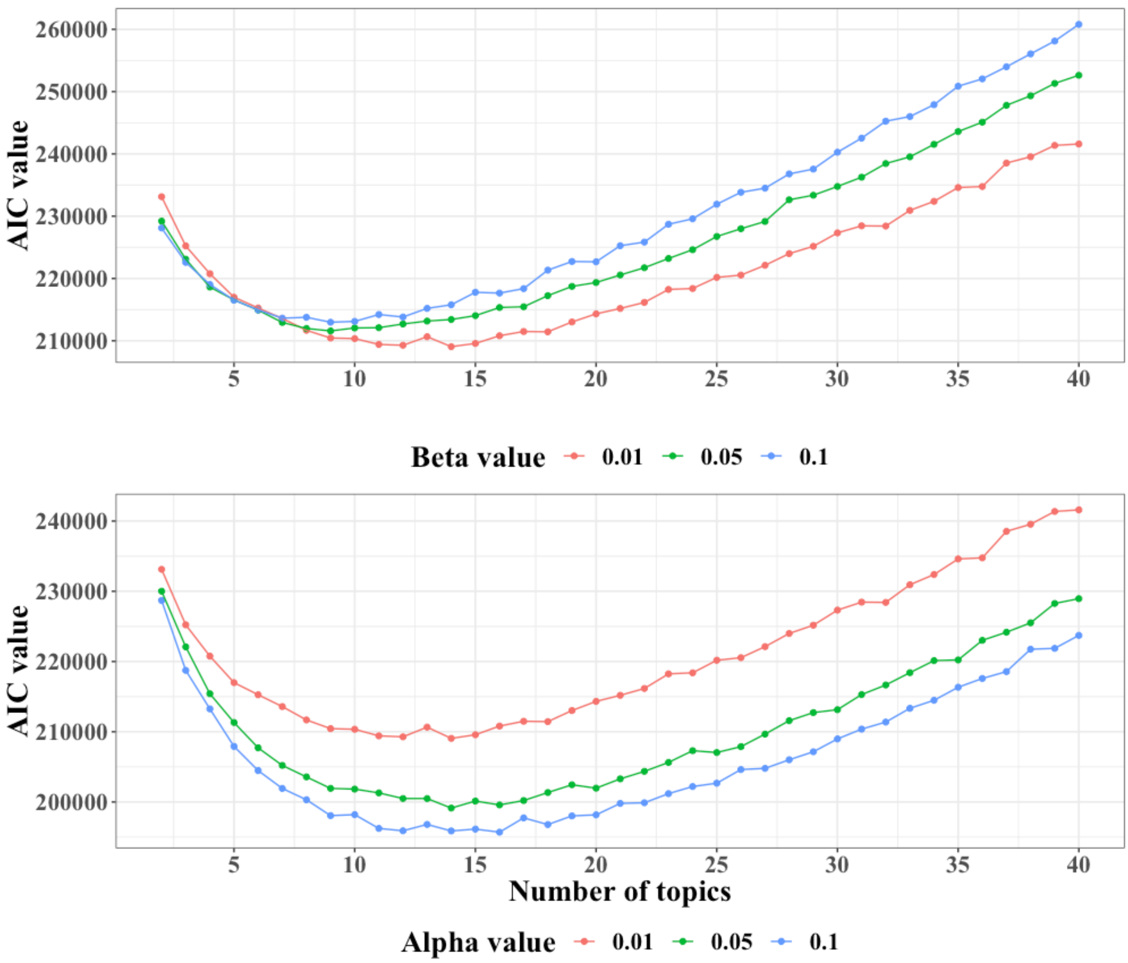

Figure 4.1: Grid search of K with fixed alpha at 0.01 and fixed beta at 0.01

Table 4.1: The configuration of parameters in LDA

Parameter	value	Meaning
K	12	The number of topics
Alpha	0.1	The parameter of Dirichlet distribution over topic distribution
Beta	0.01	The parameter of Dirichlet distribution over terms distribution
Iteration	3000	The number of iterations of Gibbs sampling
Burnin	5000	The number of omitted Gibbs iteration at beginning

Table 4.2: Label and quantity of each topic

Topic label	Patent quantity
Topic 1 - Document and payment verification related technologies on Blockchain	134.19
Topic 2 - Database management and data storage based on Blockchain related technologies	97.91
Topic 3 - Consensus system building and digital asset management related technologies on Blockchain	118.70
Topic 4 - Cross-chain transaction related technologies on Blockchain	105.34
Topic 5 - Database information processing based on Blockchain related technologies	118.58
Topic 6 - Integrate information across various resources to Blockchain related technologies	109.20
Topic 7 - Identity management related technologies on Blockchain	106.21
Topic 8 - Identity verification related technologies on Blockchain	115.34
Topic 9 - Healthcare applications using Blockchain and electronic currency management related technologies	115.22
Topic 10 - Database access authentication and data synchronization related technologies on Blockchain	121.54
Topic 11 - Cryptocurrency payment, token distribution and Blockchain security related technologies	125.96
Topic 12 - Digital asset management using smart contract related technologies	128.80

There are two interesting findings in topic model hyperparameter setting. First, Figure 4.1 not only shows the gird search process but also present the changing tendency of AIC value, for example, when fixing the value at 0.01, the lower value returns the lower AIC value meaning the better model and there is a total different case when fixing . Second, with lower and value than default setting[6], instead of nearly equal quantity on each topic, LDA can present the clustering characteristic according to terms appearance; namely as we discussed above, the lower value of Dirichlet distribution tending to have higher probability on each vertex.

The patents related to Blockchain technology have been divided into twelve topics. The document-topic matrix containing the topic distribution over document and top 5 ranked patents of each topic are listed in Table B.1, and the topic-term matrix containing

[6] The default setting of LDA in topicmodel package in R are $\alpha = 50/K$ and $\beta = 0.1$

the term distribution over topic and top 10 ranked terms and corresponding probabilities are listed in Table A.1.

4.2. Emerging topic analysis

In the earlier section, we utilized the natural language processing technique and the topic model to uncover the latent topic under Blockchain related patents and capture the core concept of each topic via keyword-based interpretation. The following analysis, we are going to integrate bibliographic information such as applicant and publication date to visualize the technology development trends.

To determine topic popularity and growth, the number of published patents over time (month) can be analyzed using Generalized Linear Mixed Model (GLMM). By treating topic as a categorical variable and estimating the random slope and random intercept to predict the counts of published patent, we can depict every topic on 2-dimensional plot as Figure 4.2 (GLMM: fixed effect estimate = 1.089, z-value = 3.171, p-value < 0.01). Topics with positive random intercept can be explained as having higher than average number of published patents and topics with positive random slope as having higher than average growth in publication during the time period. However, it is worth noting that as Figure 4.3 displays that each topic in Blockchain technology is still rapidly evolving and the results of GLMM present relatively growth rate compared to average.

Two topics - "Cross-chain transaction related technologies on Blockchain (Topic 4)" and "Database information processing based on Blockchain related technologies (Topic 5)", were identified as relatively "hot topics" marked in red receiving higher than average growth[7]. On the other hand, two topics - "Identity management related technologies on Blockchain (Topic 7)" and "Digital asset management using smart contract related technologies (Topic 12)", were identified as "cold topics" marked in blue having lower than average growth rate of published patents.

Although, Blockchain technology are in the initial growth periods of their technological lifecycle, the technology s-curve can still be applied to interpret the development of technologies (Ernst, 1997) and formulate patent strategy (Berkowitz, 1993). In our case, the hot topics are all in Quadrant II, which is relatively sparse with patents and change rate is high, and can be interpreted as in early growth period. Also,

[7] Here, we defined average growth as random slope approximated as normal distribution within 50% data value, namely 0 ± 0.177

the cold topics are in Quadrant III and IV, which is relatively crowed and close to average with patents respectively, and change rate is lower, and can be interpreted as in relative maturity period.

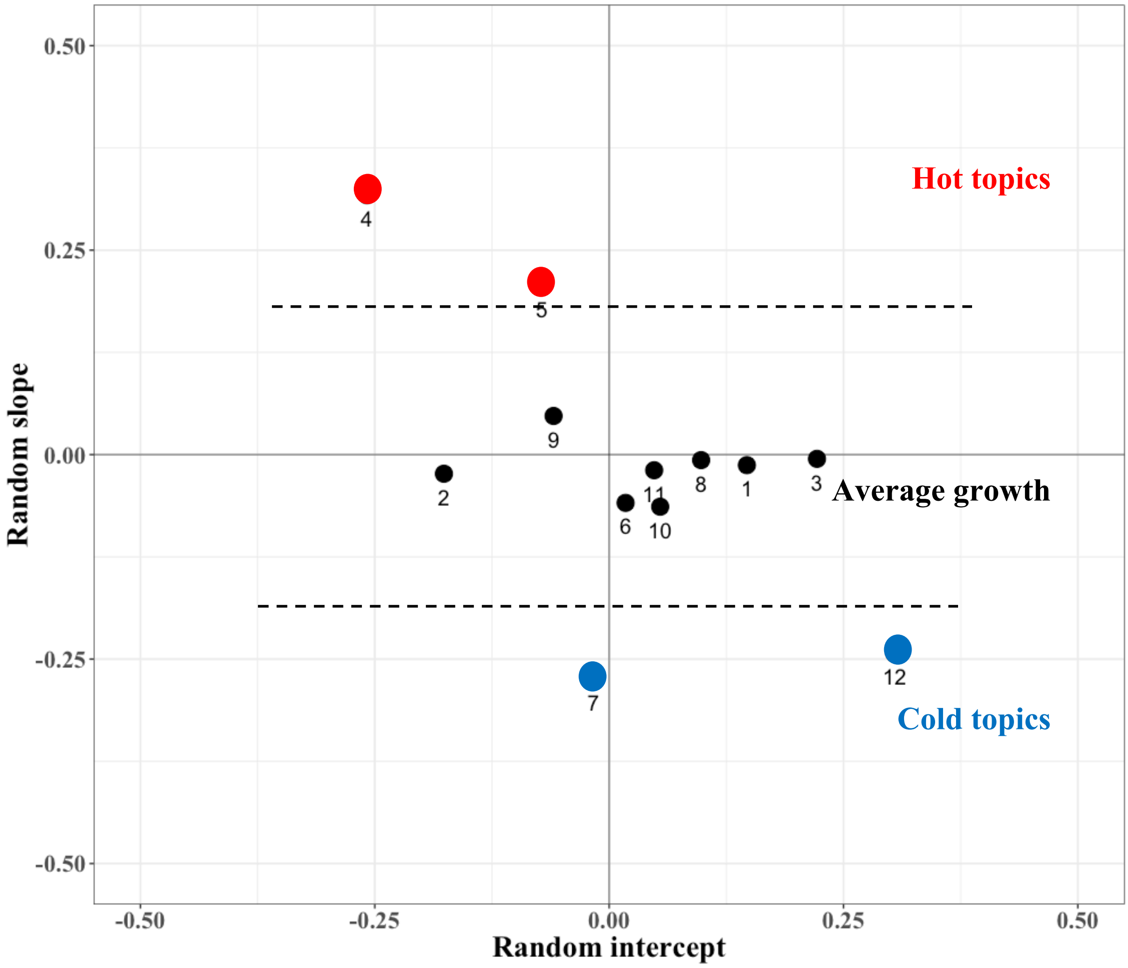

Figure 4.2: Emerging topics in Blockchain technology

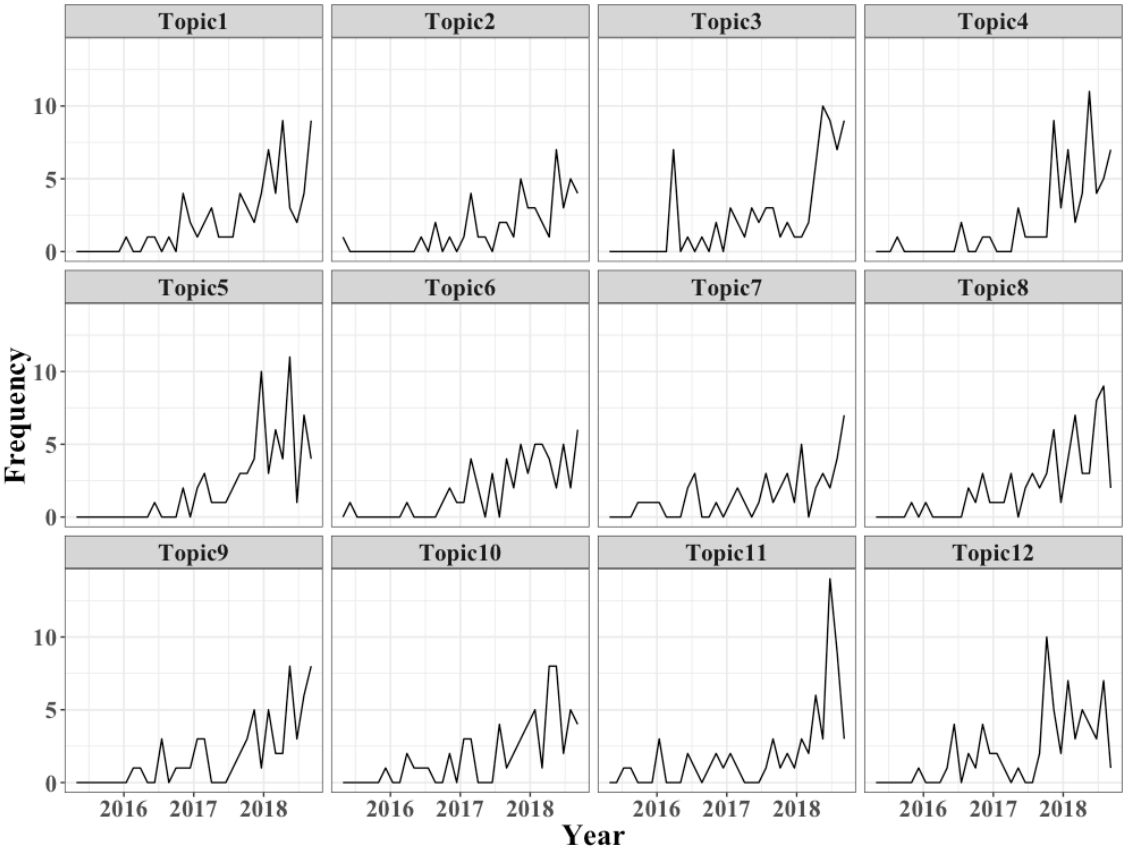

Figure 4.3: Growth in Blockchain related patents per topic

4.3. Competitive analysis

In order to explore the competitive landscape and point out the key applicants for each topic, we utilized the applicant of patents to demonstrate the competitor country and relationship between topics and applicants. After textual data preprocessing of the applicant of patents, we present the country competitive landscape as visualized in Figure 4.4. It is significant that almost more than 50% applicant's country of all topics is United States, and heavily right skewed distribution. Surprisingly, there are slightly different propensity in Topic3 and Topic11, which Antigua and Barbuda (AG)[8] has relative high proportion. However, it cannot reflect clear country competitive advantages through Figure 4.4, because of some patents were filed by the branch office, for example, Alibaba Group Holding Limited applied patent through headquarter in China (CN) and also branches in United States (US).

---

[8] Because of nChain, which is a company offering Blockchain-based solution and announcing Bitcoin SV.

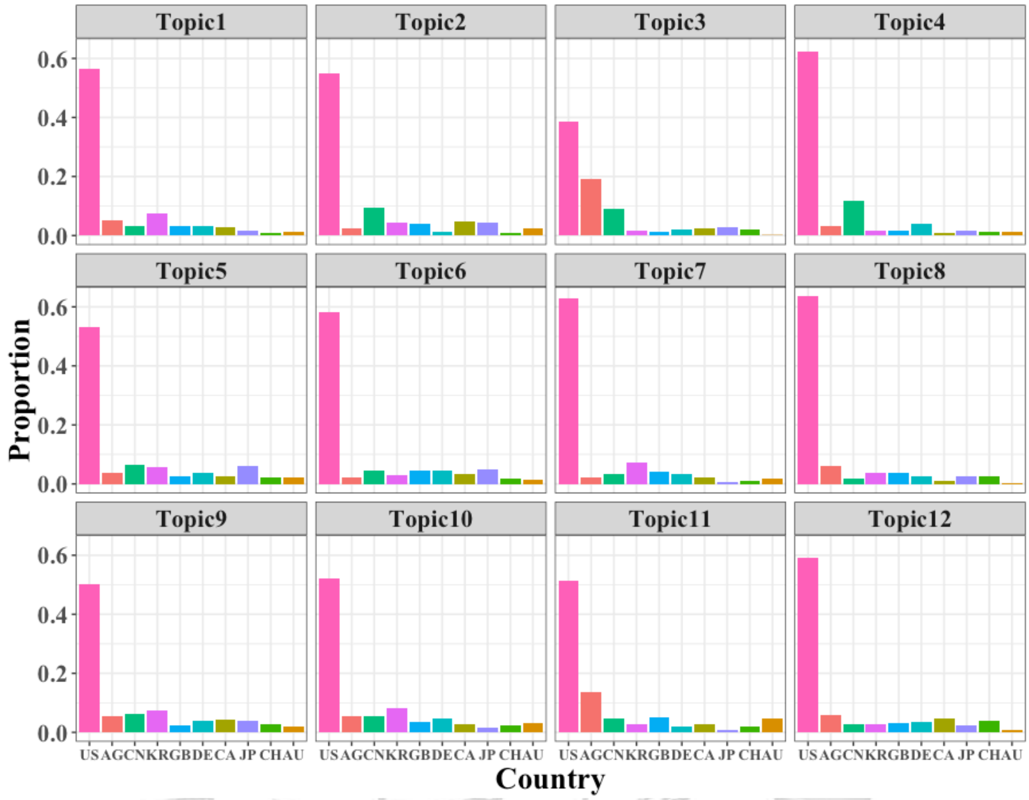

Figure 4.4: Top-10 countries[9] of applications for each topic

Hence, we applied social network analysis to demonstrate the relationship between applicants and corresponding topics, and for readability, we first filter out the edge weight (the relationship strength) less than 2.82, and node with less than 1 degree. Also, ForceAtlas 2 layout is used to enhance the efficacy of the network to present the subgroup and emphasize the network features of the applicants. In Figure 4.5, green nodes represent applicants, pink nodes represent topics, the thickness of edges between a topic and an applicant represents the sum of topic probabilities over an applicant, and the size of node represents the summation of weighted degree.

Taking applicant of patent into consideration, we can further investigate the company patent strategy whether focusing on certain topic or not. In addition, a node with a high betweenness centrality means that the node takes crucial position and owns more diverse technologies in Blockchain field. As shown in Figure 4.5, IBM, nChain, and Coinplug are positioned in the center of network holding the most diversification topics of patents. In contrast, the applicants that are far from the center of the network indicates more

[9] Country codes in top 10 countries: United States (US), Antigua and Barbuda (AG), China (CN), Korea (KR), Germany (DE), United Kingdom (Great Britain) (GB), Canada (CA), Japan (JP), Switzerland (CH), Australia (AU)

concentrate on specific topics. All applicants are also listed in Table D.1 with industry category and strongest linked topic.

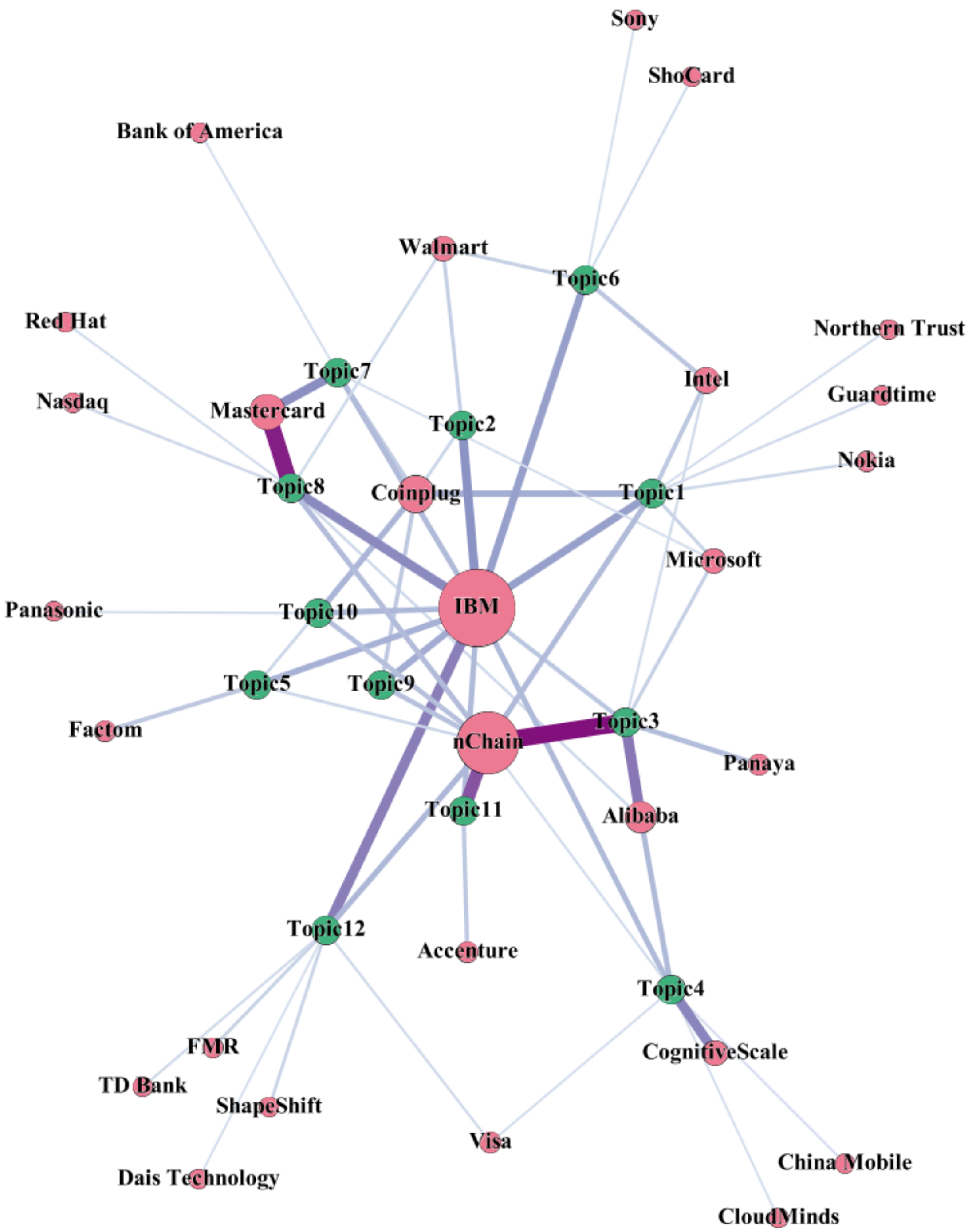

Figure 4.5: Topic-applicant network

4.4. Technology value analysis

We estimated the technology value by forward citations, claims and patent family with normalization, and removed the patent with less than 70% proportion of each topic in order to eliminate noises and emphasize the average value of topics. It returned 755 patents to evaluate technology value, and after calculating the technology value of each patent, the average value per topic can be computed from dividing by the amount of topic.

With the result displayed in Table 4.4, we can observe the topic having relatively high technology value meaning that this kind of topics are more valuable. When integrating technology value and GLMM finding, it is interesting that 'cold topics', such as Topic 12 – 'Digital asset management using smart contract related technologies' and Topic 7 – 'Identity management related technologies on Blockchain' hold the higher technology value, and in contrast 'hot topics' including Topic 4 – 'Cross-chain transaction related technologies on Blockchain' and Topic 5 – 'Database information processing based on Blockchain related technologies' receive relatively lower technology value. The reason may be that the emerging topics haven't present the patent value yet due to they are still young technology and lack of citation. This scenario is worth researching to establish corresponding patent strategy, in order to achieve or maintain competitive advantage, which we will further discuss in below. Besides, we also listed the top 20 ranked patents according to technology value in Table E.1.

Table 4.4: Market value of each topic

	Value	Amount of topic	Average value
Topic1	11.46	69.10	0.7712
Topic2	7.88	52.65	0.7415
Topic3	12.55	73.67	0.7392
Topic4	10.13	63.65	0.7607
Topic5	10.22	69.07	0.7257
Topic6	10.24	60.56	0.7350
Topic7	8.78	49.93	0.7571
Topic8	9.10	65.79	0.7172
Topic9	10.28	59.29	0.7548
Topic10	10.69	60.24	0.7539
Topic11	11.13	62.74	0.7677
Topic12	12.76	68.32	0.7641

4.5. Technology evolution map

Before further clustering topics into broad groups, we collected the top 20 featured phrases according to topic-term matrix to enhance the efficiency of hierarchical clustering. Clustering of topic content through hierarchical clustering with bootstrap sampling indicated two groups having approximately unbiased p-value more than 0.9 in Figure 4.6. The first cluster consists of Topic3 and Topic4, and via examining the co-phrases appearing in both topics such as blockchain, member and consensus, we labeled this group as “Consensus infrastructure in Blockchain related techniques”. The second cluster including Topic2, Topic5 and Topic10 contains topics representing database applications using Blockchain techniques and the co-phrase including authentication, platform, transaction information and database, we labeled this group as “Database application in Blockchain related techniques”. Last, Table F.1 presents the cosine similarity between various topics, Topic1 and Topic 8 have the highest score in the matrix and the shortest distance excepting for the topics that were already clustered. As a result, we also labeled this group as “Verification in Blockchain related technologies” and took this group into consideration for technology evolution map.

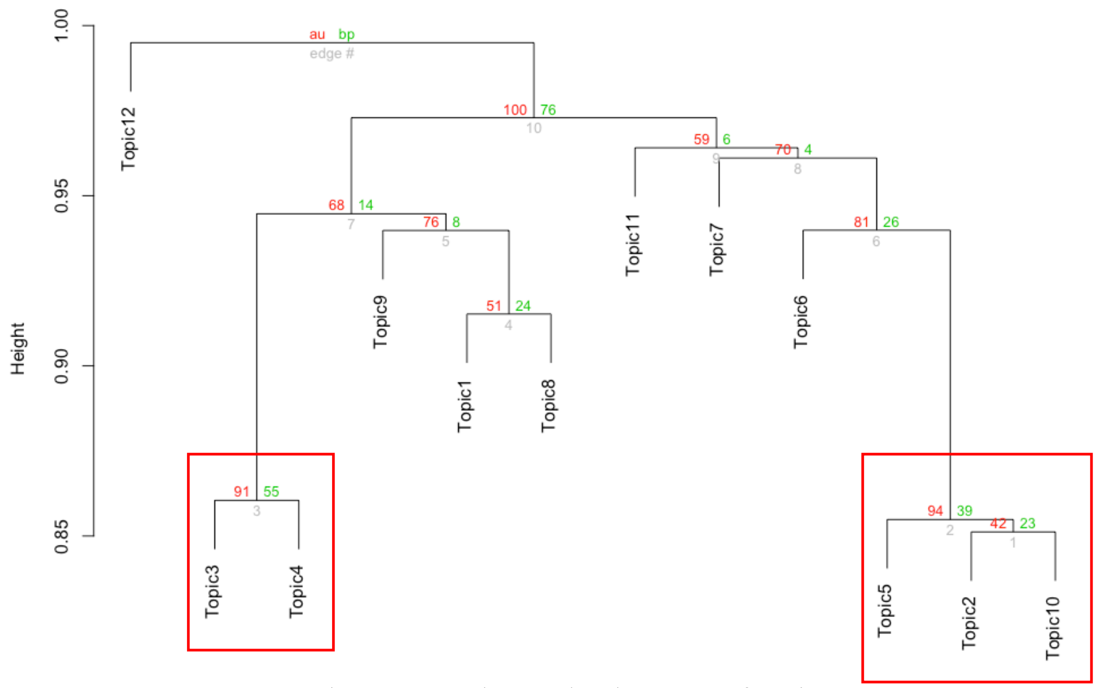

Figure 4.6: Cluster dendrogram of topics

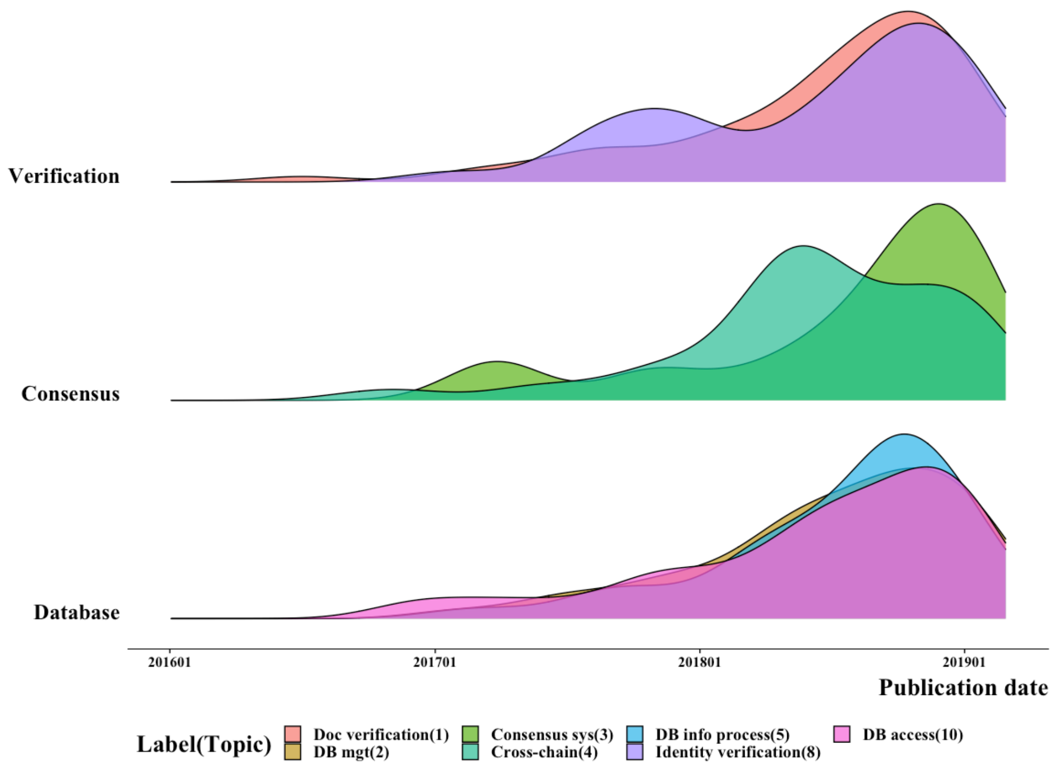

Figure 4.7: Technology evolution roadmap (clustered)

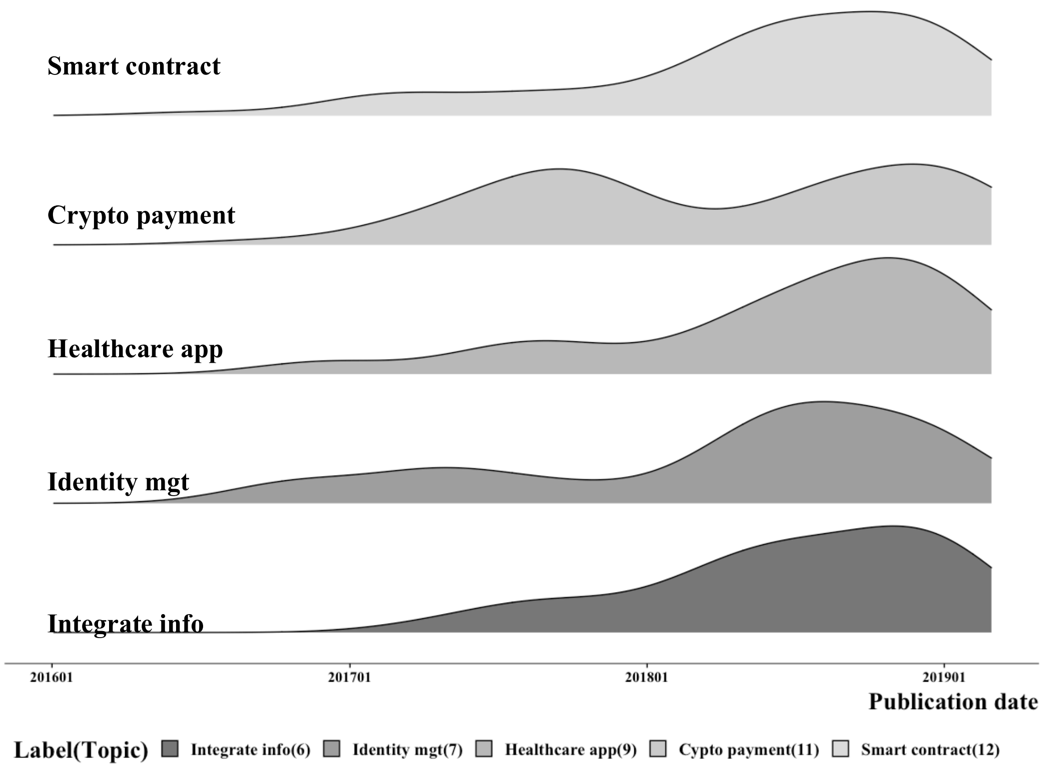

Figure 4.8: Technology evolution map (unclustered)

Technology evolution map can be constructed using the clustering results as Figure 4.7. Owing to that LDA classification tends to generate the topic with equal quality during the sampling process especially in more specific domain, the clustering step is essential to demonstrate the widely researched technology issues[10] in patent documents. Obviously, the third group – "Database application in Blockchain related techniques" including three topics is the most distinctive development issue and reflect that the private chain or consortium chain application take a huge proportion in Blockchain related patents. Figure 4.7 and Figure 4.8 also indicate the development trend of each group. Nevertheless, it is inevitable that these groups have the similar growth tendency because Blockchain technology is still in the initial growth period.

[10] Compared to the GLMM results about topic popularity, the technology we discussed here means this technology group was extensive researched and more firms are interested in.

5. Conclusion and Discussion

Blockchain technology has been seen as a revolution of traditional business process, providing properties like transparency, auditability and security (Christidis & Devetsikiotis, 2016). Companies across a wide span of industries also invest in this technology as they see the potential of making their infrastructure decentralized and minimizing the communication cost, as they become inherently safer and faster in some cases (Casino et al., 2019). This research applied the patent analysis view to identifying emerging trends in Blockchain technology. LDA is used to uncover the latent topics underlying Blockchain technology with natural language processing techniques enhancing the efficacy of information retrieve, and hyperparameter discussion also improve the interpretation of topics. The growth and popularity of topics were investigated via GLMM analysis and indicated two hot topics having higher than average growth rate: "Cross-chain transaction related technologies on Blockchain" and "Database information processing based on Blockchain related technologies" and two cold topics: "Identity management related technologies on Blockchain" and "Digital asset management using smart contract related technologies". Furthermore, we integrated the structured and unstructured data to visualize the relationship between topics and applicants and generate patent value and technology evolution map: (1) The topic-applicant network demonstrates the relationship between topics and applicants and reveals the main applicants in each topic, (2) technology value analysis presents that certain topics were well-protected and have more influence power, and (3) technology evolution map shows the topic clustering results and more popular issues in Blockchain patent documents.

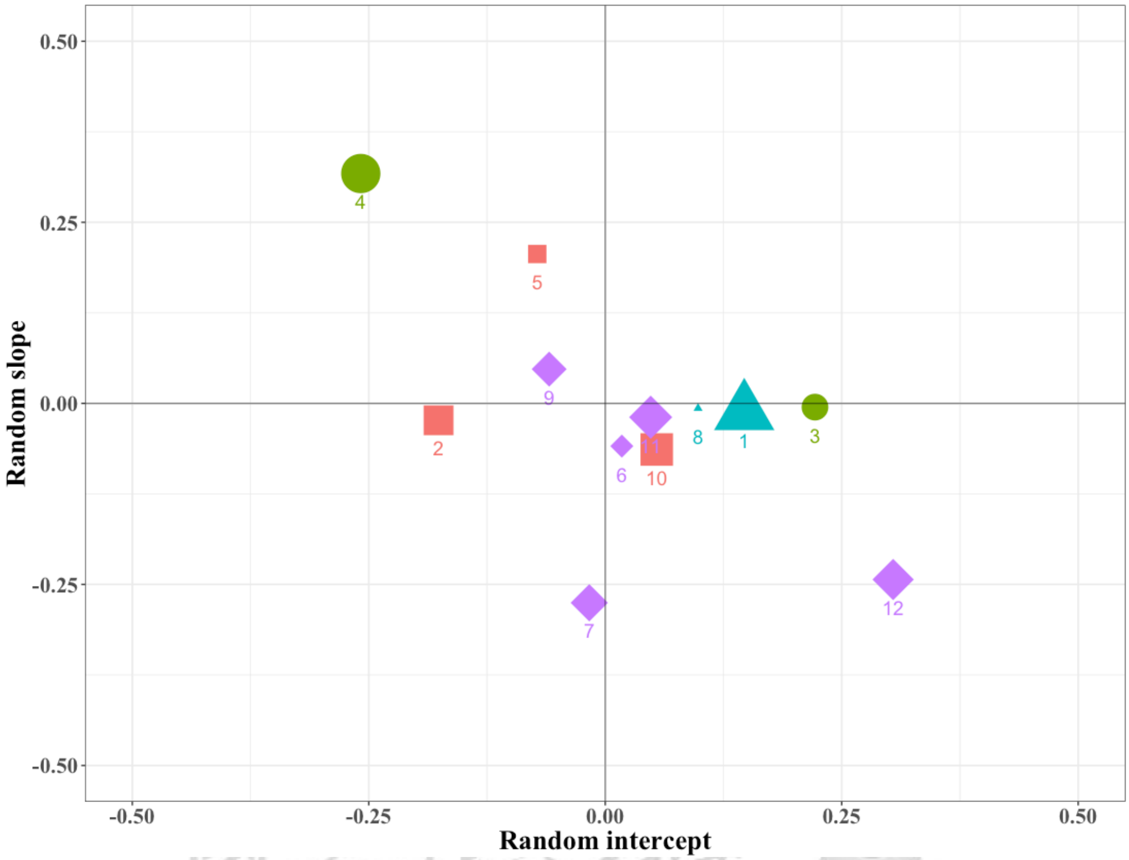

Figure 5.1: Technology strategy map

Figure 5.1 combines the results from emerging topic analysis, technology value analysis, and cluster analysis, where the color and shape[11] of the point represent the clusters, and the size of the point represents the technology value per topic, and can be utilized to formulate patent strategy.

In Quadrant II, as patents are less crowed and growth rate is higher, it may offer a paradigm shift opportunities for firms to explore new technology or market territories. Yet it also need to keep in mind that once the invention is public, many others can follow and make further improve on it. Therefore, firms must to cover their invention both broadly and deeply, and it is able to be achieved through well-defined claims; anyone who infringes any one of your patent claims is liable for patent infringement and can be prohibited from competing firms against you after the patent application is granted. Besides, taking technology value into consideration can further formulate the patent strategy; for example, although Topic 4 and Topic 5 have relatively higher growth rate as

[11] In Figure 5.1, circles represent the “Consensus infrastructure techniques” group, squares represent the “Database application” group, triangles represent “Verification related techniques” and diamonds represent others.

hot topics, Topic 4 obtains more technology value which implies published patents in Topic 4 were protected more broadly but also reflect this technique are more influential. In contrary, Topic 5 having lower technology value may be less risky in terms of infringement and litigation, but more possible that this technique cannot be realized or commercialized in the near future. In addition, these two hot topics - Topic 4 and Topic 5 are information processing techniques in cross-chain and database, respectively, showing the necessity of the interoperability and compatibility across different blockchain. Among them, Topic 4 more focus on processing information on public chain, and Topic 5 is dealing with the communication between databases based on Blockchain namely private chain or consortium chain. However, especially through increasing usage of established blockchain network like Bitcoin, Ethereum, and Ripple, most blockchains operate on isolated ecosystems because they try to address on specific needs. The inability of different blockchain to communicate with one another has made it impossible for users or firms to enjoy the full benefits of Blockchain technology. In Fridgen, Radszuwill, Urbach, and Utz (2018) research, they also demonstrated the potential of Blockchain to serve as infrastructure for cross-organizational workflow management in a joint effort with a German Bank. Besides, the potential competitors or leading companies can be found in Figure 3.4 and Table D.1. For example, in Topic 4 one of the main applicants is CognitiveScale who offers commerce-related blockchain processing data from plurality sources; in Topic 5, Factom provides solutions for enterprises to integrate company data, documents and transactions into Blockchain.

On the other hand, two cold topics receiving lower growth rate are Topic 7 and Topic 12 which lie in Quadrant III and IV respectively. First, Quadrant IV is crowed with patents and the change rate is slow, implying that it might be a highly competitive technology field and unlikely to be a source of competitive advantage for firms. Besides, Topic 12 also holds higher technology value, therefore companies who sought to launch related technology can be achieved via licensing or cooperation with companies who already published related licenses to mitigate R&D expenses. For instance, ShapeShift is a Blockchain company that offers global trading of a variety of digital assets via web and mobile platform and smart contract applications to manage multi-asset, companies who sought to build related technology can further cooperate with them to shorten the development time. In addition, other potential cooperation opportunities can be found in

Table D.1 and topic-applicant network. The other cold topic, Topic 7 is a total different case owing to lower number of published patents. For topics in Quadrant III, companies should understand the reasons of sparse patenting, and the technologies in this field might be old-fashioned. However, taking Blockchain technology into account, it may rather be a relative fresh technological areas and an opportunity to gain great advantages by patenting business-specific embodiments. Furthermore, identity management in Blockchain technology provides the fundamental infrastructure to facilitate further applications such as registration, legal documents, voting and so on. As a result, Topic 7 can be one another field that companies can take into consideration, but also need to realize that this technique may not have the significant impacts near in time. The leading companies in this area are Mastercard and Coinplug, the former applies blockchain to identify users and distribute electronic vouchers and the latter offers a blockchain-based solution of providing a safer and more convenient identity authorization.

Overall, Blockchain remains a young technology both from a legislation and a patenting perspective and Gartner (2018) also applied the hype cycle on Blockchain and positioned it in “Trough of Disillusionment”, which is the stage before significant value will be realized. Although Blockchain application are being widely deployed in finance, internet of things, healthcare, and so on, many issues have yet to be addressed (Casino et al., 2019). However, the potential of Blockchain technology must not be overlooked and technology strategy can be used to establish competitive advantage. Amidst the proliferation of Blockchain patents, the interoperability characteristic stands out to be emerging topics and widely discussed smart contract applicants are a relatively mature technique. With above discussion, firms can formulate patent strategy according to different applications and development orientation or seek out cooperation companies through topic-applicant network.

However, we still have limitations which possibly merit the future research for further improvements. First, in the data collection period, some data fields of patents are hard to be retrieved, for example, using the assignee attribute would be better than applicants, because the owner of patent could be transfer to other company or individual and the assignee is the entity that has the property right to the patent. Second, the discussion of hyperparameter tuning in LDA is seldom mentioned in previous studies, especially in more specific technology field. Though we use AIC metric to estimate the quality of model, there are still several measurements like perplexity, cosine similarity, etc. to find the best model. Last, this research applied patent mining method to identify emerging trends, but in Blockchain technology field the cryptocurrencies and related services may not be presented in patent documents and have immense influence. According to CoinMarketCap[12], the number of cryptocurrencies is currently exceeding 2,000 and still growing. Hence, the whitepaper of cryptocurrencies can be one another choice to explore technology opportunities and explore the Blockchain applications.

[12] CoinMarketCap: https://coinmarketcap.com/

6. Reference

1. Abraham, B. P., & Moitra, S. D. (2001). Innovation assessment through patent analysis. *Technovation, 21*(4), 245-252. doi:10.1016/s0166-4972(00)00040-7
2. Alonso, S. G., Arambarri, J., Lopez-Coronado, M., & Diez, I. D. (2019). Proposing New Blockchain Challenges in eHealth. *Journal of Medical Systems, 43*(3). doi:10.1007/s10916-019-1195-7
3. Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:10.1016/j.jml.2012.11.001
4. Basberg, B. (1987). Patents and the measurement of technological change: a survey of the literature. *Research Policy*, 131-141.
5. Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.
6. Berkowitz, L. (1993). GETTING THE MOST FROM YOUR PATENTS. *Research-Technology Management, 36*(2), 26-31.
7. Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet Allocation. *Machine Learning Research, 30*, 993-1022.
8. Blei, D. M., & Lafferty, J. D. (2007). A CORRELATED TOPIC MODEL OF SCIENCE. *Annals of Applied Statistics, 1*(1), 17-35. doi:10.1214/07-aoas114
9. Blei, D. M., & Lafferty, J. D. (2009). Topic models.
10. Breitzman, A., & Thomas, P. (2015). The Emerging Clusters Model: A tool for identifying emerging technologies across multiple patent systems. *Research Policy, 44*(1), 195-205. doi:10.1016/j.respol.2014.06.006
11. Breslow, N. E., & Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association, 88*(421), 9-25. doi:10.2307/2290687
12. Burnham, K. P., & Anderson, D. R. (2016). Multimodel Inference. *Sociological Methods & Research, 33*(2), 261-304. doi:10.1177/0049124104268644
13. Casino, F., Dasaklis, T. K., & Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and Informatics, 36*, 55-81. doi:10.1016/j.tele.2018.11.006

14. Chen, H., Zhang, G., Zhu, D., & Lu, J. (2017). Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. *Technological Forecasting and Social Change, 119*, 39-52. doi:10.1016/j.techfore.2017.03.009

15. Christidis, K., & Devetsikiotis, M. (2016). Blockchains and Smart Contracts for the Internet of Things. *Ieee Access, 4*, 2292-2303. doi:10.1109/access.2016.2566339

16. Chung, Y. J., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. C. (2013). A Nondegenerate Penalized Likelihood Estimator for Variance Parameters in Multilevel Models. *Psychometrika, 78*(4), 685-709. doi:10.1007/s11336-013-9328-2

17. Dabbagh, M., Sookhak, M., & Safa, N. S. (2019). The Evolution of Blockchain: A Bibliometric Study. *Ieee Access, 7*, 19212-19221. doi:10.1109/access.2019.2895646

18. Daim, T. U., Rueda, G., Martin, H., & Gerdsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change, 73*(8), 981-1012. doi:10.1016/j.techfore.2006.04.004

19. Erdi, P., Makovi, K., Somogyvari, Z., Strandburg, K., Tobochnik, J., Volf, P., & Zalanyi, L. (2013). Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics, 95*(1), 225-242. doi:10.1007/s11192-012-0796-4

20. Ernst, H. (1997). The use of patent data for technological forecasting: The diffusion of CNC-technology in the machine tool industry. *Small Business Economics, 9*(4), 361-381. doi:10.1023/a:1007921808138

21. Fall, C. J., Törcsvári, A., Benzineb, K., & Karetka, G. (2003). Automated categorization in the international patent classification. *SIGIR Forum, 37*(1), 10-25. doi:10.1145/945546.945547

22. Fall, C. J., Torcsvari, A., Fievet, P., & Karetka, G. (2004). Automated categorization of German-language patent documents. *Expert Systems with Applications, 26*(2), 269-277. doi:10.1016/s0957-4174(03)00141-6

23. Fridgen, G., Radszuwill, S., Urbach, N., & Utz, L. (2018). Cross-Organizational Workflow Management Using Blockchain Technology - Towards Applicability, Auditability, and Automation. *HICSS*.

24. Gartner. (2018). *Hype Cycle for Blockchain Technologies*. Retrieved from

25. Geum, Y., Jeon, J., & Seol, H. (2013). Identifying technological opportunities using the novelty detection technique: a case of laser technology in semiconductor manufacturing. *Technology Analysis & Strategic Management, 25*(1), 1-22. doi:10.1080/09537325.2012.748892

26. Giovanni, D. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy, 11*(3), 147-162. doi:https://doi.org/10.1016/0048-7333(82)90016-6

27. Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360-1380. doi:10.1086/225469

28. Greenville, A. C., Dickman, C. R., & Wardle, G. M. (2017). 75 years of dryland science: Trends and gaps in arid ecology literature. *Plos One, 12*(4), e0175014. doi:10.1371/journal.pone.0175014

29. Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proc Natl Acad Sci U S A, 101 Suppl 1*, 5228-5235. doi:10.1073/pnas.0307752101

30. Grun, B., & Hornik, K. (2011). Topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software, 40*(13), 1-30.

31. Guo, Y., Xu, C., Huang, L., & Porter, A. (2012). Empirically informing a technology delivery system model for an emerging technology: illustrated for dye-sensitized solar cells. *R & D Management, 42*(2), 133-149. doi:10.1111/j.1467-9310.2012.00674.x

32. Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *Rand Journal of Economics, 36*(1), 16-38.

33. Haywood, S. (2003). Academic Vocabulary. .

34. Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters, 105*(3), 287-289. doi:10.1016/j.econlet.2009.08.019

35. Heinrich, G. (2008). Parameter estimation for text analysis.

36. Hu, Z. Y., Fang, S., & Liang, T. (2014). Empirical study of constructing a knowledge organization system of patent documents using topic modeling. *Scientometrics, 100*(3), 787-799. doi:10.1007/s11192-014-1328-1
37. Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *Plos One, 9*(6). doi:10.1371/journal.pone.0098679
38. Kabore, F., & G, P. W. (2016). Can Patent Family Size and Composition Signal Patent Value?
39. Kshetri, N. (2018). Blockchain's roles in meeting key supply chain management objectives. *International Journal of Information Management, 39*, 80-89. doi:10.1016/j.ijinfomgt.2017.12.005
40. Lemley, M. A., & Shapiro, C. (2005). Probabilistic patents. *Journal of Economic Perspectives, 19*(2), 75-98. doi:10.1257/0895330054048650
41. Lewis, D., Yang, Y., Rose, T., Li, F. (2004). SMART stopword list. *Journal of Machine Learning Research*.
42. Li, W., & McCallum, A. (2006). Pachinko Allocation DAG-Structured Mixture Models of Topic Correlations.
43. Liu, S., & Shyu, J. (1997). Strategic planning for technology development with patent analysis. *INTERNATIONAL JOURNAL OF TECHNOLOGY MANAGEMENT, 13*(5-6), 661-680.
44. Lu, H. F., Huang, K., Azimi, M., & Guo, L. J. (2019). Blockchain Technology in the Oil and Gas Industry: A Review of Applications, Opportunities, Challenges, and Risks. *Ieee Access, 7*, 41426-41444. doi:10.1109/access.2019.2907695
45. Mingxiao, D., Xiaofeng, M., Zhe, Z., Xiangwei, W., & Qijun, C. (2017, 5-8 Oct. 2017). *A review on consensus algorithm of blockchain.* Paper presented at the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC).
46. Nakamoto, S. (2008). Bitcoin: a peer-to-peer electronic cash system. *9*, 1-9.
47. Reitzig, M. (2004). Improving patent valuations for management purposes - validating new indicators by analyzing application rationales. *Research Policy, 33*(6-7), 939-957. doi:10.1016/j.respol.2004.02.004

48. Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). *The author-topic model for authors and documents*. Paper presented at the Proceedings of the 20th conference on Uncertainty in artificial intelligence, Banff, Canada.
49. Sharma, P. K., Chen, M., & Park, J. H. (2018). A Software Defined Fog Node Based Distributed Blockchain Cloud Architecture for IoT. *Ieee Access, 6*, 115-124. doi:10.1109/ACCESS.2017.2757955
50. Stanley, W., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*.
51. Straka, M., Hajic, J., & Strakova, J. (2016). UD- Pipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological anal- ysis, POS tagging and parsing.
52. Straka, M., & Strakova, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe.
53. Suzuki, R., & Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics, 22*(12), 1540.
54. Swan, M. (2015). *Blockchain: Blueprint for a new economy*: O'Reily Media, Inc.
55. Szabo., N. (1996). Smart Contracts: Building Blocks for Digital Markets.
56. Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association, 101*(476), 1566-1581. doi:10.1198/016214506000000302
57. Tseng, Y. H., Lin, C. J., & Lin, Y. I. (2007). Text mining techniques for patent analysis. *Information Processing & Management, 43*(5), 1216-1247. doi:10.1016/j.ipm.2006.11.011
58. Venugopalan, S., & Rai, V. (2015). Topic based classification and pattern identification in patents. *Technological Forecasting and Social Change, 94*, 236-250. doi:10.1016/j.techfore.2014.10.006
59. Wang, B., Liu, S., Ding, K., Liu, Z., & Xu, J. (2014). Identifying technological topics and institution-topic distribution probability for patent competitive intelligence analysis: a case study in LTE technology. *Scientometrics, 101*(1), 685-704. doi:10.1007/s11192-014-1342-3

60. Wang, J. (2018). A topic modeling approach for technological opportunity analysis. *Working Paper, Graduate Institute of Technology Management, National Chung Hsing University, Taichung, Taiwan.*

61. Wang, S., Ouyang, L., Yuan, Y., Ni, X., Han, X., & Wang, F. (2019). Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 1-12. doi:10.1109/TSMC.2019.2895123

62. Westgate, M. J., Barton, P. S., Pierson, J. C., & Lindenmayer, D. B. (2015). Text analysis tools for identification of emerging topics and research gaps in conservation science. *Conserv Biol, 29*(6), 1606-1614. doi:10.1111/cobi.12605

63. Wikipedia. (2014). Transitional phrase.

64. Yli-Huumo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where Is Current Research on Blockchain Technology?-A Systematic Review. *Plos One, 11*(10), e0163477. doi:10.1371/journal.pone.0163477

65. Yoon, J., Choi, S., & Kim, K. (2010). Invention property-function network analysis of patents: a case of silicon-based thin film solar cells. *Scientometrics, 86*(3), 687-703. doi:10.1007/s11192-010-0303-8

66. Yoon, J., & Kim, K. (2012a). Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. *Scientometrics, 90*(2), 445-461. doi:10.1007/s11192-011-0543-2

67. Yoon, J., & Kim, K. (2012b). TrendPerceptor: A property-function based technology intelligence system for identifying technology trends from patents. *Expert Systems with Applications, 39*(3), 2927-2938. doi:10.1016/j.eswa.2011.08.154

68. Yue, X., Wang, H. J., Jin, D. W., Li, M. Q., & Jiang, W. (2016). Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control. *Journal of Medical Systems, 40*(10). doi:10.1007/s10916-016-0574-6

69. Zhang, L., Li, L., & Li, T. (2015). Patent Mining: A Survey. *SIGKDD Explor. Newsl., 16*(2), 1-19. doi:10.1145/2783702.2783704

70. Zhao, W., Chen, J., & Zou, W. (2015). *Best practices in building topic models with LDA for mining regulatory textual documents*. Retrieved from

71. Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics, 16 Suppl 13*, S8. doi:10.1186/1471-2105-16-S13-S8

72. Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017, 25-30 June 2017). *An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends.* Paper presented at the 2017 IEEE International Congress on Big Data (BigData Congress).

73. Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2018). Blockchain challenges and opportunities: a survey. *International journal of web and grid services, 14*(4), 352-375.

Appendix

A. Topic-term matrix

Table A.1: Top 10 ranked phrases and corresponding probability

Topic1		Topic2		Topic3	
Phrase	Probability	Phrase	Probability	Phrase	Probability
Document	0.1160	Database	0.1317	Digital asset	0.0739
Signature	0.0661	Profile	0.0484	Consensus	0.0681
Payment	0.0570	Network device	0.0409	Blockchain node	0.0476
Verification	0.0564	Customer	0.0235	Service data	0.0388
Internet_of_thing	0.0441	Patient	0.0235	Sequence	0.0366
Iot device	0.0305	Host	0.0212	Business process	0.0271
Share	0.0246	Content	0.0197	Sample	0.0198
Layer	0.0227	Index	0.0189	Congress	0.0190
Encryption key	0.0214	Cluster	0.0167	Constraint	0.0168
Seller	0.0214	Data element	0.0151	Field value	0.0154
Topic4		Topic5		Topic6	
Phrase	Probability	Phrase	Probability	Phrase	Probability
Data source	0.0549	Client	0.0844	Vehicle	0.0478
member	0.0456	Platform	0.0505	Sensor	0.0456
Digital certificate	0.0394	Algorithm	0.0485	Certification	0.0304
Verification	0.0325	Authentication	0.0432	Identification	0.0290
Target	0.0325	Financial institution	0.0319	Container	0.0270
Article	0.0270	Game	0.0306	Digital content	0.0256
Private data	0.0263	Transaction information	0.0299	Blockchain ledger	0.0194
Node device	0.0255	Data block	0.0299	Gateway	0.0194
Transaction request	0.0255	Database	0.0226	Measurement	0.0180

Table A.1: Top 10 ranked phrases and corresponding probability (continued)

Blockchain node	0.0247	Peer	0.0219	User data	0.0180
Topic7		Topic8		Topic9	
Phrase	Probability	Phrase	Probability	Phrase	Probability
Identity	0.1024	Identifier	0.1078	Validation	0.0454
Transaction amount	0.0334	Block header	0.0430	Client	0.0426
Identity information	0.0282	Mobile device	0.0362	Dataset_	0.0302
Cryptlet	0.0208	Verificatio n	0.0293	Healthcare	0.0289
Data element	0.0208	Digital signature	0.0259	Electronic currency	0.0261
Authorization	0.0193	Biometric data	0.0225	Signal	0.0247
Reputation	0.0193	Computing system	0.0211	Activity	0.0206
Session	0.0193	Data value	0.0211	Feature	0.0206
Access control	0.0171	Transaction value	0.0211	Evaluation	0.0186
Validator node	0.0171	Data file	0.0205	Subscriber	0.0179
Topic10		Topic11		Topic12	
Phrase	Probability	Phrase	Probability	Phrase	Probability
Content	0.0673	Token_	0.1113	Smart contract	0.2118
Electronic device	0.0422	Cryptocurrency_	0.0576	Asset	0.1765
Transaction data	0.0342	Peer to peer	0.0441	Owner	0.0371
Data record	0.0312	Payment	0.0294	Segment	0.0237
License	0.0294	Merchant	0.0256	Ownership	0.0207
Authentication	0.0251	Platform	0.0256	Credit	0.0195
Media	0.0251	Customer	0.0250	Crypto	0.0164
Unique identifier	0.0245	Rights	0.0218	Authorizatio n	0.0152
Vote	0.0239	Energy	0.0211	Loan	0.0146
Authentication service	0.0202	Digital currency	0.0205	Entitlement	0.0128

B. Document-topic matrix

Table B.1: Top 5 ranked patents

Topic1	
Pub. No	Title
US20170279783A1	Secure 3d model sharing using distributed ledger
US20170103167A1	Blockchain system for natural language processing
WO2017207717A1	Validating blockchain transactions regarding real money
US20180219683A1	Possession and alteration of documents
US09855785B1	Digitally encoded seal for document verification
Topic2	
Pub. No	Title
US20170034197A1	Mitigating blockchain attack
US20180241551A1	Permission information management system, user terminal, right-holder terminal, permission information management method, and permission information management program
US20180107958A1	Blockchain expense and resource utilization optimization
US20180068130A1	System and method for maintaining a segregated database in a multiple distributed ledger system
EP3346633A1	Permission information management system, user terminal, proprietor terminal, permission information management method, and permission information management program
Topic3	
Pub. No	Title
US20170295021A1	Method to assure correct data packet traversal through a particular path of a network
US20170109640A1	Generation of candidate sequences using crowd-based seeds of commonly-performed steps of a business process
US20170109639A1	General model for linking between nonconsecutively performed steps in business processes
US20170109637A1	Crowd-based model for identifying nonconsecutive executions of a business process
US20170109636A1	Crowd-based model for identifying executions of a business process
Topic4	
Pub. No	Title
US20170324738A1	Internet security
WO2017192475A1	Securing transactions for allocation of internet resources with blockchain
US20180276668A1	Method and apparatus for consensus verification
US20180165612A1	Method for providing commerce-related, blockchain-associated cognitive insights using blockchains
US20180165611A1	Providing commerce-related, blockchain-associated cognitive insights using blockchains

Table B.1: Top 5 ranked patents (continued)

Topic5	
Pub. No	Title
US20170228371A1	Blockchain-enhanced database
WO2017146333A1	Forgery/tampering verification system and method for financial institution certificates based on blockchain
WO2017059498A1	Managing technical process data
US20180225651A1	Aerospace commerce exchange
US20180157688A1	Ledger-chained distributed information handling systems and methods
Topic6	
Pub. No	Title
US20180227627A1	System and method for secure appliance operation
US20180227130A1	Electronic identification verification methods and systems
US20180173917A1	Reader device for reading a marking comprising a physical unclonable function
US20180144292A1	Apparatus and method for tracking consumer premises inventory
US20180137306A1	Container update system
Topic7	
Pub. No	Title
US20170353311A1	Systems and methods for providing identity scores
US20170140394A1	Consensus-based reputation tracking in online marketplaces
WO2016170538A1	A method of distributed management of electronic documents of title (edt) and system thereof
US20180253745A1	Blockchain data
US20180248845A1	Autonomous decentralization of centralized stateful security services with systematic tamper resistance
Topic8	
Pub. No	Title
US20170346693A1	Method and system for efficient distribution of configuration data utilizing permissioned blockchain technology
US20170344987A1	Method and system for an efficient consensus mechanism for permissioned blockchains using bloom filters and audit guarantees
US20170344435A1	Method and system for desynchronization recovery for permissioned blockchains using bloom filters
US20180255055A1	Systems and methods for providing block chain-based multifactor personal identity verification
WO2017204943A1	Method and system for an efficient consensus mechanism for permissioned blockchains using audit guarantees

Table B.1: Top 5 ranked patents (continued)

Topic9	
Pub. No	**Title**
US20170364655A1	Monitoring adherence to a healthcare plan
US20160358253A1	Electronic currency management method and electronic currency system
WO2017100864A1	A mobile earth station
US20180219945A1	System and method for evaluating the feasibility of introducing a new node in a blockchain infrastructure
US20180198818A1	Cryptographic network protocol escalation path
Topic10	
Pub. No	**Title**
US20180234496A1	Systems and methods for synchronizing content and information on multiple computing devices
US20180174097A1	Tracking shipments with a local and remote blockchain
WO2018124857A1	Blockchain database-based method and terminal for authenticating user non-face-to-face by utilizing mobile id, and server utilizing method and terminal
WO2018109598A1	System, method, and computer program for implementing a license ledger in a network function virtualization (nfv) based communication network
WO2018065441A1	Method for electronically documenting license information
Topic11	
Pub. No	**Title**
US20170374049A1	Distributed key secret for rewritable blockchain
US20170338957A1	Rewritable blockchain
US20170286951A1	Dynamic delivery authorization for cryptographic payments
US20170169363A1	Integrated system of search, commerce and analytics engines supported by beacons, mobile consumer and merchant applications which discover, connect to, report on, communicate and transact with places, people and objects based on their proximal, ephemeral and analytical attributes on a symmetric basis
US20170132630A1	Block chain alias for person-to-person payments
Topic12	
Pub. No	**Title**
US20170140408A1	Transparent self-managing rewards program using blockchain and smart contracts
WO2017145007A1	System and method for controlling asset-related actions via a blockchain
US20180241546A1	Splitting digital promises recorded in a blockchain
US20180218343A1	Smart contract execution on a blockchain
US20180218176A1	System and method of creating an asset based automated secure agreement

C. Questionnaire

區塊鏈專利主題命名

親愛的填答者您好：

此研究的題目為「利用主題式專利探勘模型探索區塊鏈新興技術趨勢」，其中我們利用區塊鏈相關專利資料進行主題分群。為提升主題命名以及後續分析之可信度，此問卷將收集相關產業專家之建議，對主題名稱進行修改。

在此感謝您寶貴的時間以及意見，若有任何相關問題，歡迎透過下方聯絡資訊聯繫。您所提供之資訊將不會用於商業用途，且您的資料將不會被洩露，請放心填答。

國立中興大學科技管理研究所 劉嘉鴻 敬上
信箱：jus8447321@gmail.com

此研究使用LDA主題模型，依據專利文件中詞與詞的共同出現的頻率進行主題分群；且每個文件將混合不同的主題比率。

在此問卷中，將提供每個主題前十關鍵字與前五專利之標題，填答者可根據所提供的資訊，對研究者的主題之命名進行評分以及給予修改之建議。

此研究將專利分為十二個主題，以下為主題一：

Keyword	
Document	Encryption key
Signature	Seller
Payment	Buyer
Verification	Digital signature
Internet_of_thing	Electronic document
Title of patent	
Digitally encoded seal for document verification	
Method for paying cost of iot device based on blockchain and merkle tree structure related thereto, and server, service providing terminal, and digital wallet using the same	
System and method for tracking of provenance and flows of goods, services, and payments in responsible supply chains	
Validating blockchain transactions regarding real money	
Threshold digital signature method and system	

主題一之命名：

"Document and payment verification related technologies on Blockchain”

・請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

・您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題二：

Keyword	
Database	Content
Profile	Permission request
Customer	Transaction information
Patient	Index
Host	Data type
Title of patent	
Permission information management system, user terminal, right-holder terminal, permission information management method, and permission information management program	
System and method for maintaining a segregated database in a multiple distributed ledger system	
Blockchain cluster processing system and method, computer device and storage medium	
Method and system for data conversion and data model optimization	
Enterprise-accessible customer locker	

主題二之命名：

"Database management and data storage based on Blockchain related technologies"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題三：

Keyword	
Digital asset	Business process
Consensus	Member
Blockchain node	Congress
Service data	Constraint
Sequence	Validation
Title of patent	
Method to assure correct data packet traversal through a particular path of a network	
Digital asset traceability and assurance using a distributed ledger	
Genetic algorithms in blockchain space	
Enhancing processing efficiency of blockchain technologies using parallel service data processing	
Generation of candidate sequences using crowd-based seeds of commonly-performed steps of a business process	

主題三之命名：

"Consensus building and digital asset management related technologies on Blockchain”

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題四：

Keyword	
Data source	Article
Member	Private data
Digital certificate	Node device
Target	Transaction request
Verification	Blockchain node
Title of patent	
Securing transactions for allocation of internet resources with blockchain	
Blockchain network, article transaction method and apparatus, and node device	
Method and apparatus for processing transaction requests	
Hash values for the bidirectionally linked blockchain	
Blockchain system and data storage method and apparatus	

主題四之命名：

"Cross-chain transaction related technologies on Blockchain”

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

____

以下為主題五：

Keyword	
Client	Game
Platform	Data block
Algorithm	Transaction information
Authentication	Database
Financial institution	Peer
Title of patent	
Blockchain-enhanced database	
Forgery/tampering verification system and method for financial institution certificates based on blockchain	
Ledger-chained distributed information handling systems and methods	
Distributed blockchain data management in a satellite environment	
Processing of a user device game-playing transaction based on location	

主題五之命名：

"Database information processing based on Blockchain related technology"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題六：

Keyword	
Vehicle	Digital content
Sensor	Blockchain ledger
Certification	Gateway
Identification	Measurement
Container	User data
Title of patent	
Vehicle control for reducing road wear	
Apparatus and method for tracking consumer premises inventory	
Using ledger sensors to enable contextual contracts across various enterprise blockchain applications	
Blockchain-based shadow images to facilitate copyright protection of digital content	
System and method utilizing facial recognition with online (social) network to access casualty health information in an emergency situation	

主題六之命名：

"Integrate information across various resources to Blockchain related technologies"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題七：

Keyword	
Identity	Authorization
Transaction amount	Reputation
Identity information	Session
Cryptlet	Access control
Data element	Validator node
Title of patent	
Systems and methods for providing identity scores	
Registration and authorization method, device and system	
Method and system for electronic vouchers via blockchain	
Industrial network using a blockchain for access control, and access control method	
System and method to dynamically setup a private sub-blockchain based on agility of transaction processing	

主題七之命名：

"Identity management related technology on Blockchain"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題八：

Keyword	
Identifier	Biometric data
Block header	Computing system
Mobile device	Data value
Verification	Transaction value
Digital signature	Data file
Title of patent	
Method and system for efficient distribution of configuration data utilizing permissioned blockchain technology	
Systems and methods for providing block chain-based multifactor personal identity verification	
Verifying authenticity of computer readable information using the blockchain	
Block chain generation device, block chain generation method, block chain verification device, block chain verification method and program	
Method and system for storage and transfer of verified data via blockchain	

主題八之命名：

"Identity verification on Blockchain related technologies"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題九：

Keyword	
Validation	Signal
Client device	Activity
Dataset_	Feature
Healthcare	Evaluation
Electronic currency	Subscriber
Title of patent	
Monitoring adherence to a healthcare plan	
Detecting medical fraud and medical misuse using a shared virtual ledger	
Method and apparatus for blockchain verification of healthcare prescriptions	
Dental health tracking via blockchain	
Electronic currency management method and electronic currency system	

主題九之命名：

"Healthcare applications using Blockchain and electronic currency management related technologies"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題十：

Keyword	
Content	Authentication
Electronic device	Media
Transaction data	Unique identifier
Data record	Vote
License	Authentication server
Title of patent	
Systems and methods for synchronizing content and information on multiple computing devices	
Tracking shipments with a local and remote blockchain	
Electronic voting system and control method	
Methods and apparatuses for controlling electronic voting	
Blockchain database-based method and terminal for authenticating user non-face-to-face by utilizing mobile id, and server utilizing method and terminal	

主題十之命名：

"Access authentication and data synchronization related technology on Blockchain"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題十一：

Keyword	
Token_	Platform
Cryptocurrency_	Customer
Peer to peer	Rights
Payment	Energy
Merchant	Digital currency
Title of patent	
Distributed key secret for rewritable blockchain	
Dynamic delivery authorization for cryptographic payments	
Block chain alias for person-to-person payments	
Computer-implemented methods and systems for validating tokens for blockchain-based cryptocurrencies	
Determining a common secret for the secure exchange of information and hierarchical, deterministic cryptographic keys	

主題十一之命名：

"Cryptocurrency payment, token distribution and Blockchain security related technologies"

- 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

  ☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

- 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

以下為主題十二：

Keyword	
Smart contract	Credit
Asset	Crypto
Owner	Authorization
Segment	Loan
Ownership	Entitlement
Title of patent	
Transparent self-managing rewards program using blockchain and smart contracts	
System and method for controlling asset-related actions via a blockchain	
Smart contract execution on a blockchain	
System and method of creating an asset based automated secure agreement	
Blockchain for program code credit and programmer contribution in a collective	

主題十二之命名：

"Digital asset management using smart contract related technologies"

• 請問您認為上述名稱是否符合關鍵字以及專利標題所描述？

☐非常不符合 ☐ 不符合 ☐ 普通 ☐ 符合 ☐ 非常符合

• 您認為如何命名將更精準傳達關鍵字以及專利意涵？（選填）

____

D. Vital applicants in topic-applicant network

Table D.1: Vital applicants in topic-applicant network

Applicant	Category	Main Country	Strongest linked topic	Website
Accenture	Consulting	IE	11	www.accenture.com
Alibaba	E-commerce	CN	3	www.alibaba.com
Bank of America	Finance	US	7	www.bankofamerica.com
China Mobile	Telecommunication	CN	4	www.chinamobileltd.com
CloudMinds	Technology	CN	4	www.en.cloudminds.com
CognitiveScale	Technology	US	4	www.cognitivescale.com
Coinplug	Blockchain	KR	1	www.coinplug.com
Dais Technology	Technology	US	12	www.daistechnologies.com
Factom	Blockchain	US	5	www.factom.com
FMR	Finance	US	12	www.fidelity.com
Guardtime	Blockchain	VG	1	guardtime.com
IBM	Technology	US	12	www.ibm.com
Intel	Technology	US	6	www.intel.com
Mastercard	Finance	US	8	www.mastercard.com
Microsoft	Technology	US	3	www.microsoft.com
Nasdaq	Finance	US	8	www.nasdaq.com
nChain	Blockchain	AG	3	nchain.com
Nokia	Telecommunication	FI	1	www.nokia.com
Northern Trust	Finance	US	1	www.northerntrust.com
Panasonic	Home appliances	US	10	www.panasonic.com
Panaya	Technology	IL	3	www.panaya.com
Red Hat	Technology	US	8	www.redhat.com
ShapeShift	Blockchain	CH	12	shapeshift.io
ShoCard	Blockchain	US	6	shocard.com
Sony	Diversification	JP	6	www.sony.net
TD Bank	Finance	CA	12	www.td.com
Visa	Finance	US	4	www.visa.com
Walmart	Retail	US	6	www.walmart.com

E. Patent valuation

Table E.1: Top 20 ranked patent by technology value

Publication number	Forward citation	Claims	Patent family	Market value	Topic
US20160342978A1	27	20	12	2.0161	7
US20150332283A1	118	41	2	1.8256	9
US20170132630A1	29	20	4	1.8163	11
US20160342994A1	24	16	7	1.7649	7
US20180006826A1	7	21	2	1.6995	8
US20160212146A1	18	20	6	1.5567	1
US20190052454A1	0	19	245	1.5547	12
US20170180134A1	15	20	12	1.5333	4
US20180255055A1	0	80	45	1.5175	8
US20180096175A1	5	20	3	1.4783	10
US20180262493A1	1	30	45	1.4756	8
US20170046664A1	12	13	24	1.4747	12
US20180247302A1	1	78	16	1.4676	11
US20180227275A1	4	20	6	1.4615	3
US20170046693A1	10	17	24	1.4563	12
US20180075527A1	5	20	2	1.4046	12
US20180108024A1	5	33	1	1.3730	1
US20170132626A1	14	20	6	1.3728	7
US20160203572A1	14	32	3	1.3711	12
US09935772B1	5	30	1	1.3550	1

F. Cosine similarity matrix

Table F.1: Cosine similarity between each topic

	Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9	Topic10	Topic11	Topic12
Topic1	1											
Topic2	0.0033	1										
Topic3	0.0354	0.0020	1									
Topic4	0.0759	0.0024	0.1199	1								
Topic5	0.0415	0.1321	0.0048	0.0207	1							
Topic6	0.0081	0.0632	0.0063	0.0017	0.0148	1						
Topic7	0.0120	0.0208	0.0528	0.0194	0.0314	0.0264	1					
Topic8	0.0793	0.0045	0.0371	0.0518	0.0358	0.0024	0.0467	1				
Topic9	0.0475	0.0165	0.0528	0.0453	0.0572	0.0051	0.0286	0.0558	1			
Topic10	0.0443	0.1333	0.0093	0.0636	0.1260	0.0755	0.0597	0.1037	0.0626	1		
Topic11	0.0652	0.0274	0.0051	0.0039	0.0631	0.0067	0.0288	0.0119	0.0144	0.0379	1	
Topic12	0.0106	0.0010	0.0019	0.0123	0.0024	0.0025	0.0157	0.0010	0.0069	0.0046	0.0033	1