Information Systems and e-Business Management
https://doi.org/10.1007/s10257-020-00461-9

ORIGINAL ARTICLE

A scientometric review of blockchain research

Liang Zhou[1] · Lin Zhang[2] · Ying Zhao[1] · Ruoshu Zheng[3] · Kaiwen Song[1]

Received: 31 March 2019 / Revised: 23 November 2019 / Accepted: 20 January 2020

Abstract

As a new and rapidly developing distributed technology, blockchain has shown substantial impact on the fields of cryptocurrency and e-commerce and thus has attracted the interests of governments, enterprises, and research institutions. As such, it is important to understand the state of blockchain research in order for those institutions to plan their research effectively. Towards this goal, we first integrate two search strategies to construct a representative dataset. Then we analyze the blockchain literature between 2013 and 2018 in this dataset with a science mapping approach and Latent Dirichlet Allocation model. Based on the results of the analysis, we found the following insights: (1) the evolution of blockchain research involves many disciplines while two major disciplines—computer science and business—lead the way. (2) The current research can be divided into four research areas: underlying technology architecture, privacy and security, financial application, and smart scene applications. (3) The evolution of blockchain research has gone through three stages: basic blockchain technology, various business applications, and integration with advanced technologies such as fog computing, Internet of things, and artificial intelligence. (4) We also discovered a close coupling of two research issues that have transformative power, i.e., the potential by the application of cryptocurrency in the real world and the improvement of blockchain technology based on the application requirements. Overall, our study uses multiple complementary scientometric methods to generate a panoramic view of the recent developments in blockchain research. Future research directions and some limitations are discussed.

Keywords Blockchain research · Literature review · Scientometrics · Latent Dirichlet Allocation · CiteSpace

✉ Ying Zhao
zhaoying@scu.edu.cn

[1] Department of Information Management and Technology, School of Public Administration, Sichuan University, No. 24 South Section 1, Yihuan Road, Chengdu 610065, China

[2] School of Information, Renmin University of China, No. 59 Zhongguancun Street, Haidian District, Beijing 100872, China

[3] Department of Information Systems, College of Business, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong, China

Published online: 01 February 2020

1 Introduction

Blockchain was originally conceptualized by Nakamoto (2008) as a core component to support transactions conducted in Bitcoin. A blockchain is literally a chain of blocks of information that registers Bitcoin transactions; there is a set of stringent rules that govern the verification of the block and that ensure immutability and persistence of the block. The algorithms and the computational infrastructure of creating, inserting, and using the blocks are known as the blockchain technology (Zhao et al. 2016). As the interdisciplinary knowledge basis and field application of blockchain are conducive to the construction of Internet of Value and the building of credit society, blockchain has attracted, in recent years, increasing attention from governments, enterprises and academia. At the same time, with the rise of big data and cloud computing, blockchain research has gradually become a hot global research topic.

Previous review studies related to blockchain are mainly divided into two categories, namely, overall and the local situation research. The research regarding the overall situation has included scholars' qualitative assessments of the key technologies needed to realize blockchain and the breadth of applications and various challenges faced by the realization that have provided a basis for its wide use (Gao et al. 2018). There are also a few quantitative studies. Cheng et al. (2018) used CiteSpace to carry out visualization analysis of blockchain research studies with the theme of "blockchain" in 233 articles in the core database of Web of Science (as of October 19, 2017), clustered the nine research hotspots, and analyzed three future research directions. The overview of local conditions mainly refers to either the study of a specific application of blockchain technology or a specific feature of blockchain technology, such as the sharing economy (Hawlitschek et al. 2018), the Internet of things (Fernandez-Carames and Fraga-Lamas 2018; Conoscenti et al. 2016) and consensus algorithm (Du et al. 2017).

However, the existing studies are limited in literature coverage and generally based on single research methods, thus limiting the integrity of the research and diminishing the reliability of the conclusions. Blockchain is an emerging issue. At the time of collection of data for this paper, the core database of the Web of Science had collected more than 1900 articles, which far exceeds the number of articles used in review studies prior to this paper. Moreover, most of the existing research is carried out by either qualitative or quantitative methods, without considering the limitations of a single method. Therefore, in this paper, multiple complementary scientometric methods, such as literature co-citation analysis, literature burst analysis, Latent Dirichlet Allocation (LDA) topic model, and citation trajectory analysis, are utilized together to conduct research for the purpose of achieving mutual supplemental evidence validating or correcting previous results and helping researchers to identify research topics and emerging trends in the field of blockchain

We introduce the data sources and processes used in the second section of this paper and analyze the overall research of blockchain in terms of time distribution, subject contribution and core authors in the third section. In the fourth section, based on the comprehensive analysis of the external characteristics and full content of the studies, we identify four research topics in the field. In the fifth section, we identify the underlying trends in the field from two perspective of development, and finally, we conclude with a summary of our findings, future research directions and some limitations of this study in section six.

2 Data sources and analysis methods

2.1 Data collection

Most existing literature directly put "blockchain" as the key search term (Cheng et al. 2018; Xu et al. 2019). To ensure that all topics related to the blockchain are covered, we adopted a general-purpose search strategy proposed by Chen Chen 2017. Using the Web of Science core database as the data source, our input data is generated from Web of Science results from multiple topic search queries including "blockchain". In terms of the choice of search terms, this paper selects blockchain concepts, technologies and platforms that are both representative and less ambiguous according to the introduction of blockchain in Wikipedia (https://en.m.wikipedia.org/wiki/Blockchain). Terms such as "blockchain", "distributed ledger" and "smart contract", to some extent, represent the directly related concept of blockchain, while "bitcoin" and "Ethereum" are typical public chain platforms. In addition, in order to consider the integrity of the platform, this paper additionally adds the "Hyperledger Fabric", a commonly used non-public chain platform, to the search term, so that the final search term is: TS = ("blockchain" OR "distributed ledger" OR "smart contract" OR "bitcoin" OR "Ethereum" OR "hyperledger fabric"). A total of 1951 search results has been obtained as of December 28, 2018.

We believe that the accuracy and the completeness of the data are equally important bases of data reliability. Although retrieval topics have been explored as much as possible on the accuracy of the data, there is still no guarantee for zero omission. Therefore, we use a citation index-based expansion to construct our dataset, which is more robust than defining a rapidly growing field with a list of predefined keywords (Chen et al. 2012). The specific approach is that if an article cites at least one of the 1951 records, then the article will be included in the expanded dataset based on the assumption that citing a blockchain article makes the citing article relevant to the topic. The citation index-based expansion resulted in 1981 records. According to Chen (2013), using a citation expansion set has the advantage of obtaining a self-contained dataset than the more conventional practice which often uses the result set of a topic search directly.

Finally, a total of 2870 records were obtained after the data were merged and deduplicated. We deferred the assessment of relevance until the analysis stage.

2.2 Methods

We visualize and analyze the dataset with a new version of CiteSpace (5.3.R4 SE). Developed as a measurement tool, CiteSpace takes a set of bibliographic records as its input and model the intellectual structure in terms of a synthesized network based on a time series of networks derived from each year's publications (Chen 2017), which is used for the systematic evaluation of the knowledge structures. In this article, we conducted bibliometric studies on the aspects of time distribution, subject contribution, author co-citation and co-citation clusters to visually present the research history, current development, hot topics and emerging trends in the field of blockchain. We used CiteSpace to form an overview network for the systematic review of the relevant literature.

CiteSpace uses spectral clustering algorithms which based on spectral graph theory to automatically identify co-citation clustering (Luxburg 2007). It has a natural advantage for co-citation networks clustering based on link relationships rather than node attributes. Compared to traditional algorithms, such as K-means algorithm and EM algorithm, which are based on the convex spherical sample space, spectral clustering is more flexible and robust and will not fall into local optimum. In addition, in order to test the robustness of the clustering effect, we also use LDA to select the appropriate feature vocabulary of the full text and extract topics model. LDA is a Probabilistic Generative Model based on the content of text. It is a three-layer Bayesian probability model (Blei et al. 2003). By calculating the collinear frequency of all the words in the vocabulary, it can effectively cluster words with similar themes into same category. To a certain extent, LDA can make up for the cluster analysis of CiteSpace which is based on textual external features. The specific processing steps are as follows: First of all, we use Endnote and manual download to get the full text of 2792 records. The full text of these records is extracted and segmented, and the text after the word segmentation is used as a corpus for training the LDA model. In order to achieve a better clustering effect, we constructed a list of stop words before the word segmentation based on the language characteristics of English and the characteristics of the blockchain domain. Finally, we use the LDA model in the gensim library in Python to extract the topics from the above records.

3 Overall situation

3.1 Time distribution analysis

The change in the output of literature over time reflects, to some extent, the development trend in the field. Figure 1 presents the publication of papers for each year. It is noted that the research on blockchain began to expand in 2013. From 2013 to 2015, the number of published papers was small, with the majority of the early studies focusing on Bitcoin. The research on blockchain at this time was still in the initial stage of development and exploration. In 2014, however, the growth rate increased significantly, and it has been increasing almost exponentially ever since. Given that the amount of research has maintained its upward trend, this paper argues that 2014 is the first year of academic study of global blockchain, and this period of rapid development is still ongoing.

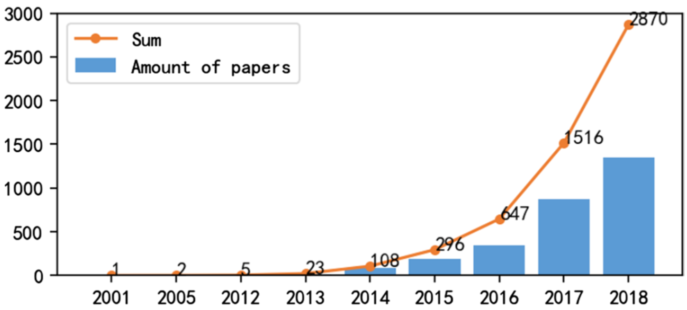

Fig. 1 Literature published in blockchain field from 2001 to 2018

3.2 Subject contribution analysis

After stacking the panoramic scientific basic maps of more than 10,000 journals based on the index of Web of Science, a dual-map overlay of citing literatures and cited literatures was created. Figure 2 shows the academic activities of various disciplines in the field of blockchain thus far. Each periodical in the figure is represented by a node, and individual journal subjects are gathered in close range in Fig. 2. Each arc represents a reference instance of a citation between studies, and the colors of the cite link distinguish the different sources (Chen and Leydesdorff 2014). On the left side is the distribution of journals in which the citing literature is located, with the most prominent ones being computers, mathematics, and economics, which are the domain applications of blockchain. The right side indicates the wide distribution of journals corresponding to the cited documents and the cited subjects. This reflects that the research foundation of blockchain is quite large and involves many disciplines, such as computer systems, economics, environmental sciences, psychology, etc. It also indicates that blockchain is an interdisciplinary and cross-domain composite frontier technology.

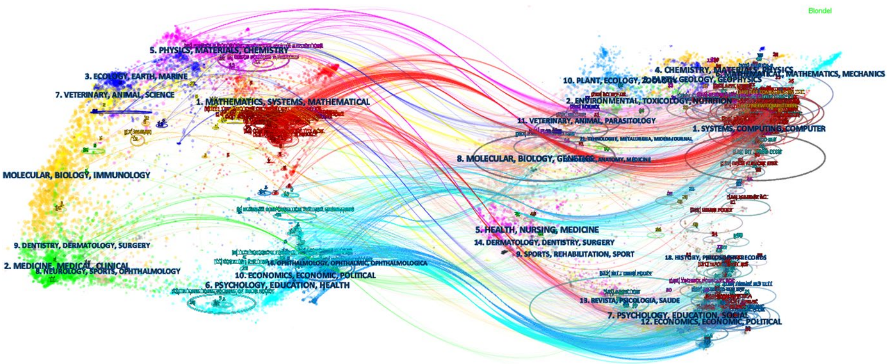

Fig. 2 Dual-map overlay of citing literature and cited literature

Table 1 Subject category whose betweenness centrality of blockchain research > 0.1

Betweenness centrality	Average year	Subject category
0.51	2013	Computer Science
0.48	2014	Business and Economics
0.26	2013	Engineering
0.2	2017	Environmental Science
0.2	2016	Psychology
0.18	2015	Mathematics
0.18	2014	Social Science

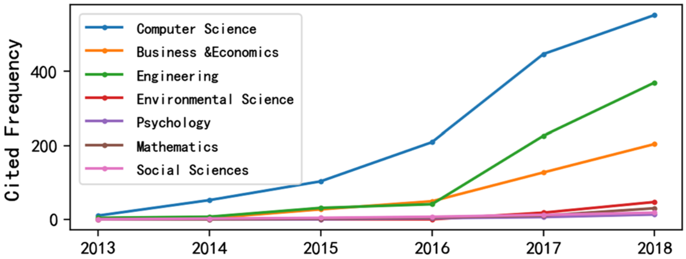

Fig. 3 The co-citation frequency distribution of the main contribution subjects from 2013 to 2018

The major subject categories in the blockchain domain are identified based on the size of node betweenness centrality. Node betweenness centrality refers to the number of shortest paths through the node in the network, and it is the measure of the connecting role of the node in the overall network (Freeman 1978). Nodes with high betweenness centrality easily become the key nodes in the network. As presented in Table 1, by merging the various research directions of the same discipline, it is determined that the existing research in the blockchain field is focused on computer Science, business and economics, engineering, environmental sciences, psychology, mathematics, and social sciences. Among these areas, two specific fields, computer science and business and economics, have made a significant contribution to the development of blockchain.

When analyzing the subject co-citations for each year, Fig. 3 shows that since 2013, computer scientists have paid increasing attention to research in this field, leading the development of blockchain technology. Their studies cover a variety of topics including interdisciplinary application, artificial intelligence theory and method, information system cybernetics, etc. Since 2015, Business and Economics and Engineering have become popular disciplines in the field of blockchain because the open and non-tampering of blockchain technology provides for the possibility of a decentralized trust mechanism. Furthermore, traditional financial assets can be integrated into the blockchain ledger and, hence, can become digital assets for blockchain. In fact, these digital assets can be stored, transferred and traded on blockchain. For example, cross-border payments, insurance claims, securities trading, bills and other aspects in the financial field have been implemented. In addition to blockchain technology, blockchain applications in finance, energy and fuel, environmental science, health, education and other research areas are receiving increasingly more attention. Although the various disciplines contribute differently to the development of blockchain, the cross-disciplinary cooperation between these disciplines better aligns the network cooperation of blockchain research, thus reflecting the application diversity of blockchain technology.

3.3 Core author analysis

Core authors are the backbone needed to promote discipline development and academic innovation. By mining and analyzing the academic activities of these authors, researchers can effectively identify the research status and development context of different fields, which is conducive to understanding the direction of the development in the field of blockchain and the latest trends in blockchain, thus promoting research on blockchain. We use the number and the citation frequencies of published articles and co-citation network analysis to help us understand the authors' academic influence. The more articles are published, the higher the authors' yields become. Since the number of published articles does not fully describe the influence of the authors, we use citation frequencies of the published articles as another index. In CiteSpace, a landmark node in merged networks has extraordinary attribute values. For example, a highly cited article tends to provide an important landmark (Chen et al. 2012). It means that a widely cited article always tends to provide important information and its authors are good candidates for significant intellectual contributions. The citation frequency is an important sign recognized by peers. Besides, as citation frequencies of journals of different disciplines vary greatly, we use co-citation analysis as a supplement. The key of co-citation analysis is to show the relationship between articles through measuring the frequency of an article cited together by other articles. The more two articles are cited together, the closer the relationship between them, which means the more similar the subject background of the two articles (Chen et al. 2014). Author co-citation analysis is developed on the basis of co-citation analysis which uses the authors as units of analysis to establish co-citation relationship. The author's citation network denotes the similarities of cognitions among the authors in the existing field and the closeness of academic relations, thus reflecting the trust of academic groups in scholars and the influence of scholars on the development of disciplines (Fu and Zhang 2018).

After dividing the studies according to affiliation, it can be seen from Table 2 that there are minor differences in the number of publications between high-yielding authors, all of whom account for between 8 and 12 articles. Among them, Choo Kim-Kwang Raymond (University of Texas System at Austin) has published 12 papers and has made outstanding contributions in research related to network security, cloud forensics and cloud technology. Among the high-yielding authors, Shi Elaine, Eyal Ittay and Juels Ari are all from Cornell University, a leader in the field of blockchain. The content of their research includes the security and scalability of Bitcoin and the smart contract system.

Table 2 High-yield authors of the blockchain field

Amount	Author	Affiliation
12	Kim-Kwang Raymond Choo	University of Texas System at Austin
10	Xiwei Xu	CSIRO
10	Elaine Shi	Cornell University
9	David Roubaud	Montpellier Business School
9	Aggelos Kiayias	University of Edinburgh
9	Elie Bouri	Holy Spirit Univ Kaslik
8	Stefan Dziembowski	University of Warsaw
8	Ittay Eyal	Cornell University
8	Ari Juels	Cornell University
8	Morgen E Peck	A journalist based in New York City

With respect to the 79,363 references in this dataset, the 50 references with the highest citation frequency in each year were selected to generate the author co-citation relationship chart. In Fig. 4, the node represents the author, and the node radius represents the citation frequency of all references by that author. The larger the radius is, the higher the citation frequency, which suggests that the author is more concerned with the specific subject. The color of the node center indicates when the author was first referenced, specifically, the darker the color of green is, the further back in time the date of the first reference of the article. In other words, a lighter green indicates that the article has been referenced more recently (Chen 2006). In addition, the nodes highlighted in red indicate a sudden increase in citations of the author's literature.

Influential authors in the blockchain field include Nakamoto (2008), who first came up with the idea of Bitcoin and its digital currency algorithm. Melanie Swan

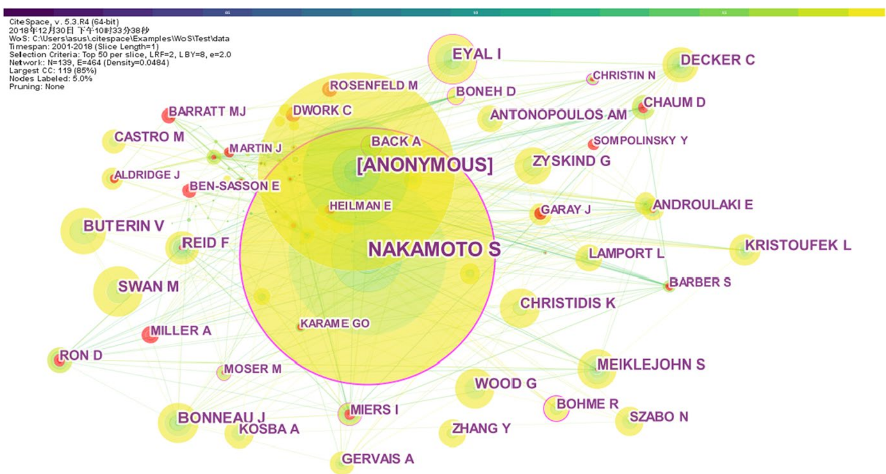

Fig. 4 The co-citation network of the most influential authors (total reference = 79,363, Node N = 139)

is the founder of the Institute for Blockchain Studies and author of the best-selling book *Blockchain: Blueprint for a New Economy*. Other than being the co-founder of *Bitcoin Magazine*, Vitalik Buterin is also involved in the creation of Ethereum, a Bitcoin network dedicated to distributed applications. Gavin Wood, as the ethereum co-founder, wrote *Ethereum: A Secure Decentralised Generalised Transaction Ledger—Byzantium Version* (i.e., *Ethereum Yellow Book*), laying the technical cornerstone of ethereum. Christidis and Devetsikiotis (2016) began to explore the possibility of blockchain integration in the field of Internet of things (IoT) and proposed two combinations of blockchain and the Internet of things given that the combination of blockchain and the Internet of things can cause significant changes in multiple industries and thus pave the way for new business models and new distributed applications.

Although there is no direct positive relationship between the author's influence and the number of author posts in the field of blockchain, the pioneering research results determine the influence of the author. These influential authors are pioneers in a certain field of research and have been widely recognized by other authors, thereby continuously promoting the development of blockchain.

4 Major specialties

4.1 Citespace clustering results

Based on the papers published between 2001 and 2018, a landscape view is generated and presented in Fig. 5. There are 70,363 references in this dataset. The top 50 citations with the highest citation frequency in each year were selected to construct the co-citation network for that year, and the co-citation network for each year was

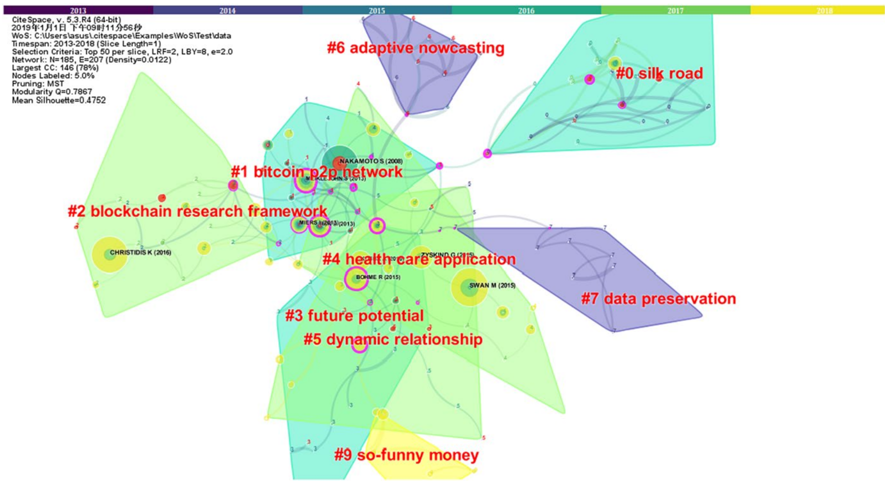

Fig. 5 A landscape view of the co-citation network, generated by top 50 per slice between 2013 and 2018

then synthesized. The network contains 8 clusters, and the largest cluster includes 25 nodes, accounting for 13.5% of the whole network. The quality of the clusters is judged from two indicators: (1) the modularity Q, which measures the degree to which the network is divided into independent modules. The larger the value is, the smaller the coupling between classes, and the better the network structure. (2) Mean silhouette can infer the degree of homogeneity in the network. The value range is [− 1, 1] whereby the larger the value is, the higher the degree of homogeneity will be, and the clearer the theme will be (Chen et al. 2012). In this paper, the modularity Q = 0.7867, which suggests that the co-cited clusters clearly define each subdomain of the blockchain. The second indicator of cluster validity, the mean silhouette, = 0.4752. Due to the existence of many small clusters, this value is relatively low. However, considering the large clusters on which we focus, the average silhouette value is actually very high. The tag of each cluster is marked by the subject words in the title of the cited document of the cluster citation. The tag extracted by CiteSpace only shows the first word with the highest probability in each cluster, which has poor readability. Therefore, the tag is manually marked in Table 3. The color of the cluster block indicates the time of the occurrence of a co-citation in the cluster where the dark blocks indicate an earlier time than the light ones.

The timeline view is another characteristic of the clusters. The cited references for each cluster are visualized horizontally according to the published timeline, with the left and right representing the time from previous to recent, respectively. Clusters are vertically arranged according to the descending order of the number of literatures, with the largest clustering appearing at the top. It is further noted, however, that the sustainability of the different clusters varies (Chen et al. 2010). While the larger nodes in Fig. 6 represent the highly cited literature, there are, again, two special nodes. The nodes with magenta outer rings represent the studies that develop across different stages of the field, and the thickness of the magenta ring indicates the betweenness centrality of the studies, which is a measure related to the potential that the scientific contribution of the study will change. The red reference ring indicates that the node has detected a reference outbreak, i.e., a reference burst. Four large clusters are described in detail.

Table 3 Details of clusters

Cluster	Mean year	Silhouette	Accumulated % of the network	From	To	Theme
#0	2013	0.986	0.14	2010	2016	Silk Road (website)
#1	2013	0.88	0.26	2011	2016	Bitcoin
#2	2014	0.887	0.37	2011	2017	Framework
#3	2013	0.85	0.47	2011	2016	Economy
#4	2014	0.851	0.56	2008	2016	Application
#5	2014	0.828	0.64	2011	2017	Fintech
#6	2011	0.989	0.70	2009	2014	Social governance
#7	2012	0.955	0.75	2008	2014	Data preservation
#9	2016	0.98	0.78	2016	2017	Income

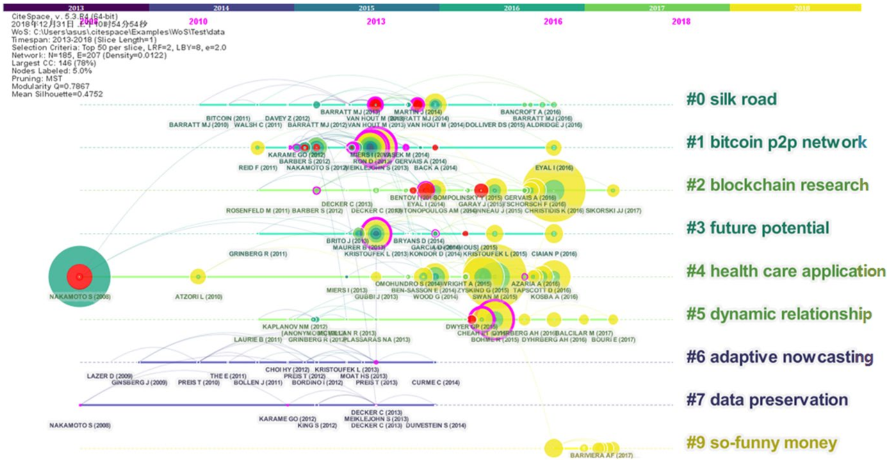

Fig. 6 Timeline view of the largest clusters

4.1.1 Cluster #1-Bitcoin

Cluster #1 is one of the largest clusters, containing 76 references from 2011 to 2016. The median year of all references in this cluster, which includes 22 highly cited papers, is 2013. This period is in the 1.0 period of the development of blockchain. As the underlying supporting technology of Bitcoin, blockchain has gradually attracted attention with the circulation of Bitcoin. Bitcoin is a peer-to-peer electronic currency system whose network protocol allows all nodes to work together to maintain a P2P network for the exchange of transaction data and also allows for download and verification. As the recognition of digital currencies, such as Bitcoin, increases, some individuals and institutions have increasingly used Bitcoin for payments and transactions. As a consequence, Bitcoin users have experienced increasing economic losses caused by security incidents. This cluster discusses the security problems of Bitcoin from the aspects of denial-of-service attacks (Vasek et al. 2014), double-spend attacks (Karame et al. 2015), and public-key-replacement attacks, and it proposes, as well, many solutions to these problems, such as the design of a network topology with more anonymity and security and the proposal of an extended and upgraded version of the Bitcoin system (Miers et al. 2013).

In Fig. 7, we can see the development trend of the research in the field of the Bitcoin network. In cluster #1, a number of contributing articles appears every year. From these articles, we can trace the development path of Bitcoin from proposal to popularity to stability. As indicated in Table 4, Reid and Harrigan (2011) assumed the lead in discussing the security of one-to-many mapping between Bitcoin users and public keys of Bitcoin and proposed the concept of a topological structure of the two networks based on the public transaction history of Bitcoin to ensure account security. Then, in 2013, Bitcoin began to expand rapidly, and the number of Bitcoin trading platforms increased substantially. Using the characteristics of Bitcoin anonymity and mobility, Meiklejohn et al. (2013) classified

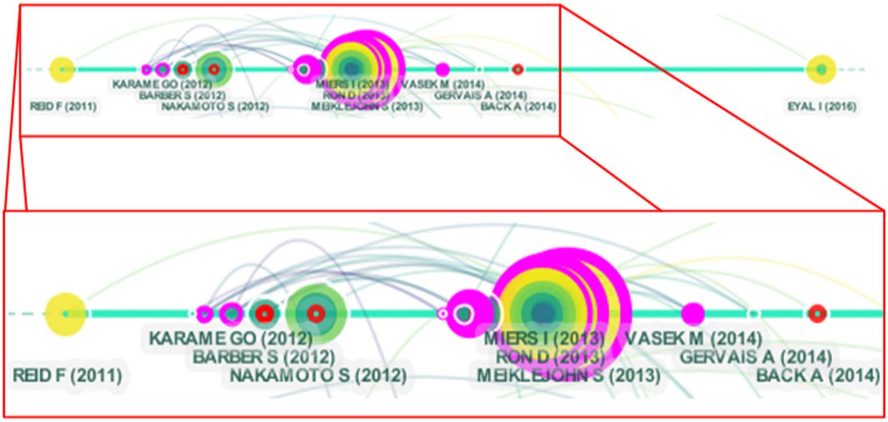

Fig. 7 Timeline view of cluster #1

Table 4 Most-cited papers of cluster #1

Cited freq	Author	Title	Year	Source
88	Sarah Meiklejohn	A fistful of Bitcoins: characterizing payments among men with no names	2013	P 2013 C INT MEAS C
34	Fergal Reid	An analysis of anonymity in the Bitcoin system (Book Chapter)	2013	SECURITY PRIVACY SOC
36	Ittay Eyal	Bitcoin-NG: A Scalable Block-chain Protocol	2016	NSDI, 2016:45–59

Bitcoin wallets by heuristic clustering. The reidentification attack method was then used to further analyze the owners of these wallets and reveal what type of activities are occurring in the public account. From 2015 to 2016, the research topics presented a deeper understanding of the latest developments in the field of the Bitcoin network. For example, due to the scalability limitations of the Bitcoin-derived blockchain protocol, simultaneous delays cannot be guaranteed. Eyal et al. (2016) introduced a new Byzantine fault tolerance blockchain protocol, Bitcoin-NG, that aims to be extended, along with other new indicators used to quantify the security and efficiency of Bitcoin-like blockchain protocol.

The main research content of the blockchain 1.0 stage is Bitcoin, while the research field with respect to technology has remained relatively stable. At present, research is primarily engaged in application transaction, while safety issues are those technical problems that must be considered throughout any practical application process. Although blockchain technology with tamper-resistant features is considered innately "safe", the results of the above research show that there are still security issues to be addressed, studied and discussed throughout the development of blockchain technology.

4.1.2 Cluster #2 framework

The mean silhouette of cluster #2 is 0.887, which indicates that the cluster has high homogeneity. Cluster #2 contains 182 references over a 6-year period, from 2011 to 2017, and includes 21 highly cited papers. As a new technology architecture that uses decentralization and non-tampering mechanisms, blockchain is an innovation in the underlying architecture of the Internet. The research framework of blockchain includes distributed ledgers, data storage mechanisms, incentive mechanisms, smart contracts, and other technical factors.

In Fig. 8, we can see that the development trend of the framework can be roughly divided into two stages. In the first stage, from 2011 to 2014, the field was relatively inactive from the perspective of burst detection and citations. As presented in Table 5, the article, “How to Use Bitcoin to Design Fair Protocols,” was published with a burst value of 5.4 in 2014. The major contribution of this paper was to propose the fairness model of security computing (Bentov and Kumaresan 2014). In the second stage, from 2014 to 2016, with respect to the number of burst detections and citations, the field of framework experienced productive research results and contributions. For example, Eyal and Sirer (2014) conducted an in-depth analysis of the Bitcoin mining protocol and concluded that the Bitcoin mining protocol is not incentive-compatible. Additionally, Bonneau et al. (2015) systematically introduced Bitcoin and other cryptocurrencies, analyzed three key components associated with the Bitcoin design, studied the consensus mechanism, currency distribution mechanism, computing puzzle, and management tools, and offered suggestions for improvement. Garay et al. (2015) analyzed and substantiated the core of the Bitcoin protocol while also improving the Byzantine protocol. In 2016, Christidis and Devetsikiotis (2016) began to focus on the possibility of applying blockchain and smart contracts to the Internet of things.

The literature in cluster #2 analyzes the transition from Bitcoin to blockchain. On the one hand, scholars actively explore the technical characteristics of Bitcoin and

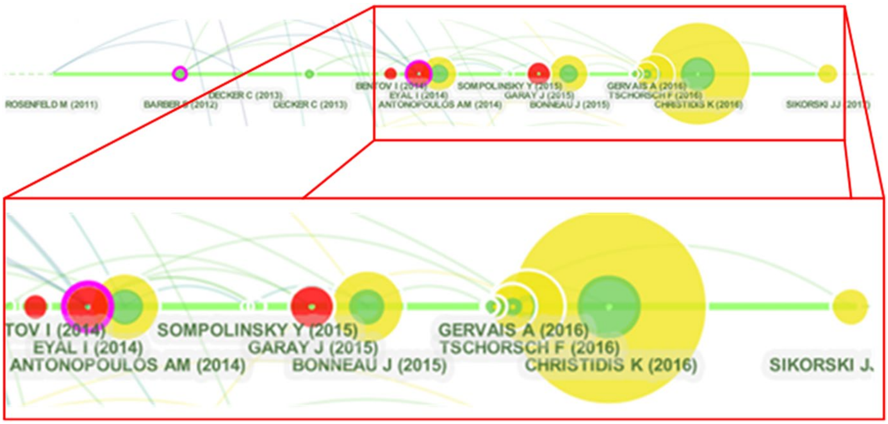

Fig. 8 Timeline view of cluster #2

Table 5 Most-cited papers of cluster #2

Cited freq	Author	Title	Year	Source
141	Konstantinos Christidis	Blockchains and Smart Contracts for the Internet of Things	2016	IEEE ACCESS
58	Joseph Bonneau	SoK: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies	2015	P IEEE S SECUR PRIV
53	Florian Tschorsch	Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies	2016	IEEE COMMUN SURV TUT
51	Andreas M Antonopoulos	MASTERING BITCOIN	2014	BOOK

put forward suggestions to solve existing problems, while on the other hand, they begin to explore the application scenarios of blockchain next to Bitcoin, thereby paving the way for the application of Bitcoin in the Internet of things.

4.1.3 Cluster #3-economy

Cluster #3 contains 76 references across a 6-year (2011 to 2016) and includes 18 of the most cited papers. The median year for all references in this cluster is 2013. This cluster's mean silhouette value of 0.85 is close to 1, indicating that the cluster has a relatively high level of homogeneity. Furthermore, the high frequency tags in cluster #3 are all related to economy, such as cryptocurrency, alternative monetary exchange, digital economy, Bitcoin economy, socioeconomic signal, digital trace, etc. Therefore, cluster #3 mainly discusses the interaction between digital currency (Bitcoin, etc.) and the social economy.

Figure 9 reveals that there are several influential nodes in cluster #3, while the research activity in this field is at a moderate level. Cluster #3 pays particular attention to the financial characteristics of digital currencies and their role in the economy. As shown in Table 6, Kristoufek (2013) analyzed the price of digital currency with Google Trends and Wikipedia's search queries and found the relationship between user search and the price of digital currency. In 2015, he continued to investigate the potential drivers of Bitcoin prices, especially in the Chinese market, through empirical analysis. Ciaian et al. (2016) also studied the factors affecting the price of digital currencies.

In addition to price factors, Maurer et al. (2013) explored the relationship between Bitcoin and privacy, labor and value. Even though it experienced a high citation burst (5.66), subsequent research has not continued.

Bitcoin is expected to have currency attributes since its inception. Therefore, the impact of economic attributes brought by blockchain technology on the social economy is a popular topic in blockchain research. The studies in cluster #3 are not very active, perhaps because the price fluctuations in digital currency are relatively large. The sharp drop in the price of Bitcoin significantly affects people's confidence in

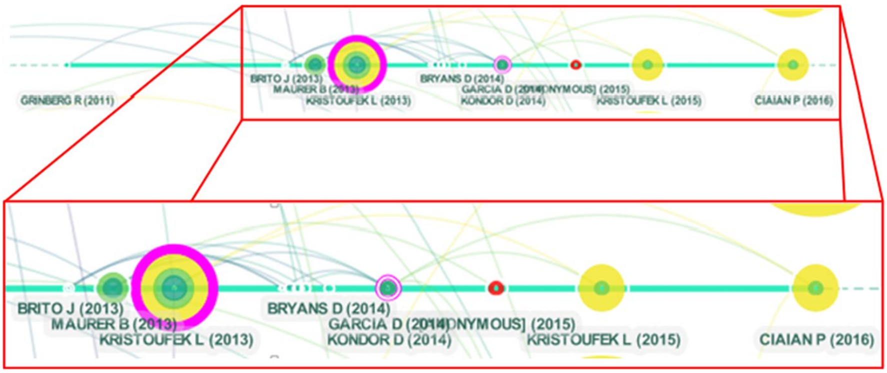

Fig. 9 Timeline view of cluster #3

Table 6 Most-cited papers of cluster #3

Cited freq	Author	Title	Year	Source
64	Ladislav Kristoufek	Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era	2013	SCI REP-UK
43	Pavel Ciaian	The economics of Bitcoin price formation	2016	APPL ECON
41	Ladislav Kristoufek	What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis	2015	PLOS ONE
29	Bill Maurer	"When perhaps the real problem is money itself!": the practical materiality of Bitcoin	2013	SOC SEMIOT
18	Dániel Kondor	Do the Rich Get Richer? An Empirical Analysis of the Bitcoin Transaction Network	2014	PLOS ONE

digital currency. Furthermore, each country has adopted different policies on digital currency, resulting in an incomplete realization of the impact of Bitcoin on the social economy.

4.1.4 Cluster #4 application

Cluster #4 is one of the largest clusters, with 237 references over a 9-year period, from 2008 to 2016, and it includes 17 most-cited papers. The median year of all references in this cluster is 2014. Cluster #4 contains the most cited article, that of Nakamoto (2008), who first proposed the Bitcoin system. The successful operation of Bitcoin greatly encouraged the development of blockchain. Cluster #4 describes the application scenarios of blockchain in various industries, including the Internet of things, data management systems, biomedicine, and healthcare.

Figure 10 reveals that the most important papers were published in 2015 and 2016. Although these articles describe different technical factors and application scenarios, they all reflect one theme, specifically, the application scenario of blockchain has been expanded. As presented in Table 7, with respect to security incidents where third parties collect large amounts of personal data and infringe on user privacy, Zyskind et al. (2015) proposed a protocol that converts blockchain into an automatic access control manager that does not require a trusted a third-party and ensures that users have control over their personal data. Their data provide a comprehensive solution to the problem of trusted computing in society. Kosba et al. (2016) designed a decentralized system, Hawk, which uses "Zero-knowledge proof" to interact and validate transactions. This system is implemented in the form of a smart contract to protect the privacy of the transaction data.

The extension of the application scenario is inseparable from the support of the underlying technology. Wood (2014) released the Ethereum Yellow Book, which explains the implementation of the blockchain paradigm and transaction encryption

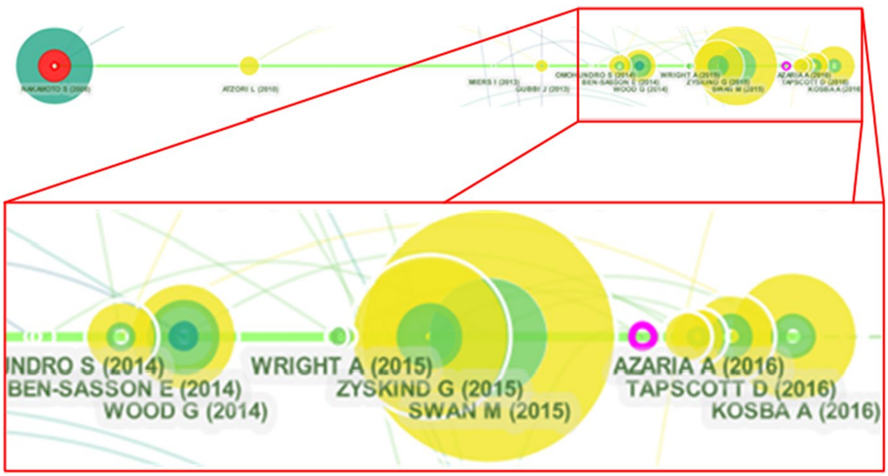

Fig. 10 Timeline view of cluster #4

Table 7 Most-cited papers of cluster #4

Cited freq	Author	Title	Year	Source
146	Melanie Swan	Blockchain: Blueprint for a New Economy	2015	O'Reilly Media, Inc.
92	Guy Zyskind	Decentralizing Privacy: Using Blockchain to Protect Personal Data	2015	IEEE Security & Privacy Workshops
75	Ahmed Kosba	Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts	2016	2016 IEEE Symposium on Security and Privacy, SP 2016
67	Gavin Wood	Ethereum: a secure decentralised generalised transaction ledger	2014	Ethereum project yellow paper

combination from a technical perspective, proposes a programmable smart contract, and discusses the design, implementation, opportunities and future obstacles of blockchain technology.

Nakamoto (2008) and Atzori et al. (2010), respectively, describe the application scenarios of blockchain and IoT in different areas, which establishing the theoretical foundation for the rapid development and application expansion of blockchain. Thus far, a large number of blockchain application platforms have been implemented, and the application solutions of blockchain technology in various industries will continue to be studied. Based on blockchain technology, any digital transaction-based costs and tracking costs are reduced and security is improved. We believe that the ultimate cost reduction will be a critical factor in the in-depth application of the technology.

4.1.5 Remaining clusters

As the remaining clusters are either relatively small in size or short in terms of the length of their duration, we omit detailed discussions of these clusters. However, we have outlined a few additional relevant clusters as follows.

Cluster #0 contains 55 references with a median year of 2013. This cluster focuses on studies of the Bitcoin trading platform. Since Bitcoin is a peer-to-peer payment network, it is neither tied to any legal currency nor is it supported by any bank, which is a perfect trading medium for those who want to hide their true identities. In the darknet (Silk Road, etc.), it is fast and easy to trade and circulate. Cluster #0 discusses the technical factors of the encryption mechanism, anonymity and trustworthiness of the digital trading platform in detail.

Cluster #5 focuses on digital asset research in the financial sector, including an empirical analysis of Bitcoin's value, investment risks, regulations, etc. It has been active since 2015, a factor that perhaps suggests the rationality that there is profit in places where there is interest. Cluster #9 focuses on the empirical analysis of Bitcoin's revenue and market efficiency. Bitcoin has been in operation for nearly a decade, and its long-running public trading data set provides a good opportunity for researchers to explore.

Cluster #7 focuses on the study of data preservation. In a narrow sense, as blockchain is essentially a decentralized distributed database, the research in this field originated first. However, due to the limited influence of research results, the research in this field has not continued.

In summary, we believe that the research content of blockchain can be roughly divided into four research directions. A first direction is privacy and security, which includes cluster #0 and cluster #1. Silk road (website) has played a significant role in the development of Bitcoin. As the representative of the blockchain 1.0 period, the technical factors of anonymity, data encryption mechanisms and trustworthiness of Bitcoin have been widely discussed and improved. With the advent of smart contract, blockchain technology has developed issues related to security, trust, and access control during the process of building a credit society. How to maintain the security and stability of the blockchain system while ensuring user privacy is also a research hotspot. Second, the underlying technology architecture includes cluster #2 and cluster #7. In addition to security and privacy issues, the infrastructure research of blockchain, such as distributed ledger technology, consortiums, consensus mechanisms, and scalability, provide underlying technical support for the decentralized application of blockchain. Third, the financial applications include cluster #3, cluster #5 and cluster #9. The research content in these clusters focuses on the impact of digital currency on the social economy. The long-term stable operations of Bitcoin have produced a substantial amount of valuable data that haves led to empirical research on Bitcoin's revenues, risks, investments, market efficiency, etc. Beginning with the blockchain 2.0 period, blockchain technology has gradually separated from the innovation of digital currency, and its application range has been extended to commercial fields, such as financial transactions and stock clearing. This period also acknowledges the value and development potential of blockchain technology in the financial field. Fourth, the smart scene application includes cluster #4 and cluster #6. The application of blockchain technology extends from the financial field to new applications, such as the Internet of things, data management systems, identity verification, supply chains, and biomedical and social governances.

4.2 Comparison of LDA and spectral clustering

Keywords are those words that are extracted from the title and the text of the literature. As such, they play a key role in the expression of the theme of the work and have retrieval significance. Keywords express the theme clearly and intuitively, allowing the reader to know the theme of the academic paper before making a decision on whether to read the paper. Although, the collection of keywords of a large number of academic research achievements in an academic field over a long period of time can reveal the overall content characteristics of the research results, the internal relationship between the research contents, the development context of the research, the direction of academic research, etc. However, at present, keywords are nonstandard natural words manually marked by the author or editor, and therefore, there are many problems with the use of keywords, as they may result in unclear scope, inconsistent marking forms, extensive meanings, omissions, incomplete meanings, etc. Considering the limitations of this thematic analysis, which uses keywords as representative of the theme, we decided to use LDA model to test the robustness of the clustering effect.

The results of topic extraction of the full text with LDA are shown in Table 8. The same clusters of LDA and CiteSpace include Bitcoin, economy, fintech, and social governance. These tags represent the three stages of blockchain. First, the research on "Bitcoin" represents the research related to decentralized digital currency in blockchain 1.0. The research on "economy" and "fintech" represents the digital asset finance applications in blockchain 2.0. Finally, the research on "social governance" represents the research content related to the blockchain 3.0 stage, which is no longer limited to the financial and economic industries but rather is applied to other fields to create a trust society.

The comparisons of the different research hotspots obtained by LDA and CiteSpace are shown in Table 9, which reveals that the differences are mainly in the research of

Table 8 LDA topic extraction results

Topic	Tags
Consensus mechanism	Drug, protocol, scheme, signature, key
Smart contract	Contract, smart, data, transaction, payment
Fintech	Financial, bank, payment, transaction, market
Smart grid	Energy, smart, system, power, contract
Bitcoin	Transaction, block, network, node, attack
Security	Data, system, network, security, user
Social governance	Technology, law, data, digital, system
Economy	Market, price, data, analysis, distribution
Cryptocurrency	Currency, money, transaction, exchange, system
Supply chain	System, chain, industry, management, technology

Table 9 Comparison of different tags between clusters of LDA and CiteSpace

Method	Different tags of clusters
CiteSpace	Silk Road, framework, application, data preservation, income
LDA	Consensus mechanism, smart contract, smart grid, security, cryptocurrency, supply chain

privacy and security. Security, privacy, trust, access control and other issues emerged during the process of building a credit society. The main topics of LDA extraction, such as smart contract, encryption, consensus mechanisms and other specific technologies, come from a microscopic perspective, while the tags of CiteSpace, such as infrastructure, applications, data protection, and cost benefits, are mainly from a macro perspective. This is primarily because CiteSpace extracts tags from the titles of the articles, whereas LDA extracts tags from content of the articles. The titles of articles usually contain more macro terms that summarize the contents of the articles, whereas technical terms generally appear within the text of the articles. Therefore, although the tags of LDA and CiteSpace vary somewhat according to the clusters' content, the research content is still about privacy and security. In addition, LDA's tags reveal that the hot areas of blockchain applications include smart grids, cryptocurrency, and supply chain industries.

To summarize, we clustered texts to mine the research hotspots based on CiteSpace and LDA. Whether it is from external features or content, the four main research hotspots of blockchain can be identified, namely, privacy and security, underlying technology architecture, financial applications and smart scene application, a result that validates the three stages of blockchain development proposed by Melanie Swan.

5 Research trends

5.1 Development tendency

CiteSpace was used to collect the keywords with high frequency from 2013 to 2018 within the same time zone, i.e., the time when they first appeared. These are arranged chronologically (Li and Chen 2016). The time zone view reveals the time evolution trend of the blockchain research. In Fig. 11, the triangular nodes represent the keywords such that the larger the node, the more frequent the corresponding keyword appears and, hence, the more attention that keyword receives. The thickness of the curves between the nodes are proportional to the frequency of co-occurrence of the corresponding two keywords. In Fig. 11, the high-frequency keywords from 2013 to 2018 are shown from the bottom left to the upper right, respectively. It is further evident that the connection density between the keywords is relatively high, thus indicating that foreign blockchain research exhibits a strong correlation and inheritance.

The main research contents of blockchain are divided into three levels. (1) From 2013 to 2014, the currency network and its trading platform, mining mechanism, and credit problems became the focus of academic attention. Then, with the deepening of the focus on the underlying technology of Bitcoin, i.e., blockchain, blockchain technology began to be applied in other digital currencies and cryptocurrencies. At this point, the security and privacy of transactions using blockchain technology began to attract great attention as a way to solve the decentralization of currency and payment means. (2) From 2015 to 2016, the emergence of the public chain platform Ethereum made the implementation of smart contract possible. Accordingly, smart contract was written into blockchain in digitalized form, and the whole process of storage, reading and execution was ensured due to the transparent, traceable and non-tampering characteristics of blockchain technology. Blockchain technology

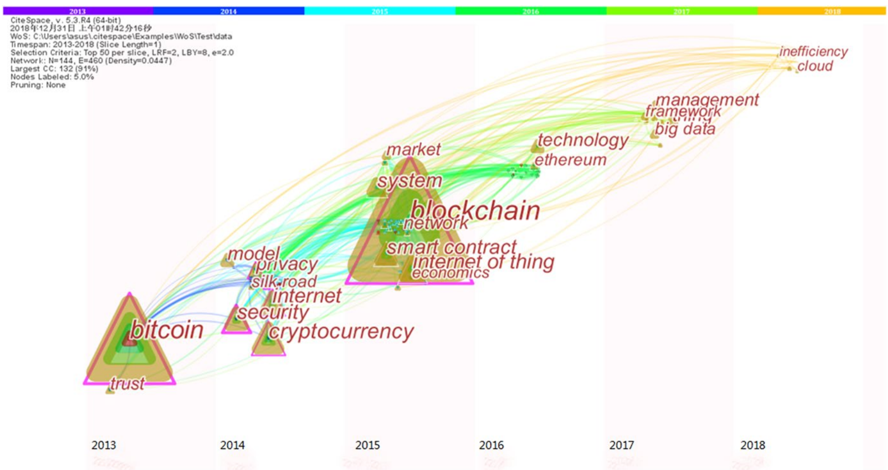

Fig. 11 Timezone view of keywords from 2013 to 2018

was used to create and transform different types of digital assets from a decentralized means of payment to a decentralized market. (3) In 2017 and 2018, the application field of blockchain technology was broadened, and blockchain technology was combined with big data, cloud computing, fog computing, edge computing, artificial intelligence and other technologies to explore the practical application of blockchain technology in the field of big data, the Internet of things, and supply chains.

Combing the development trend of blockchain, we find that its early development focused on the exploration of underlying technology. By 2016 and 2017, the technology had become relatively mature, and thus, numerous application scenarios and practical applications appeared. However, by 2018, the research on blockchain returned to the technical level, with almost all recent studies being a combination of blockchain technology and other technologies. Hence, it is speculated that there are still some unsolved technical problems in the application process of blockchain. This technology-application-technology research route makes blockchain a subject of rapid development and a topic of heated discussion.

5.2 Research frontier

5.2.1 Literature detection and analysis

The research frontier refers not only to the research problems with high density and those that are frequently cited over a short period of time but also to the emerging and sudden theoretical trends and new topics, and this research problem is likely to be deeply studied and mined by researchers during this period or during a certain period in the future (Yang 2009). CiteSpace's burst function is to detect whether a given object has statistically significant fluctuations over the whole time period and to describe the change rate in citation frequency or keyword occurrence frequency. If the citation frequency of a certain document changes significantly in a short time, then the document is said to have a citation burst (Chen 2013).

Figure 12 lists the first 20 references that exhibited a citation burst based on the previous clustering data. *Year* represents the year in which the original text was published. A citation burst has two attributes, namely, the intensity of a burst and the duration of a sudden state, where the strongest indicates the strength of a sudden state, and the begin-end is the time period during which the document is cited. From the time the literature burst occurred, articles with a high intensity are in the early research of the study of blockchain (from 2008 to 2014). During this period, researchers cited a large number of studies that focused on the technical characteristics of the Bitcoin network and payment system, such as security, privacy, incentive mechanism, etc., thereby resulting in the in-depth exploration of blockchain as the underlying support technology of Bitcoin and other digital currencies. At this stage, there were also studies on cloud services, but the literature highlighted in Fig. 12 does not intentionally combine cloud computing with blockchain. Therefore, it is determined that the research frontier from 2013 to 2016 focuses primarily on the exploration and improvement of the underlying technological framework of blockchain.

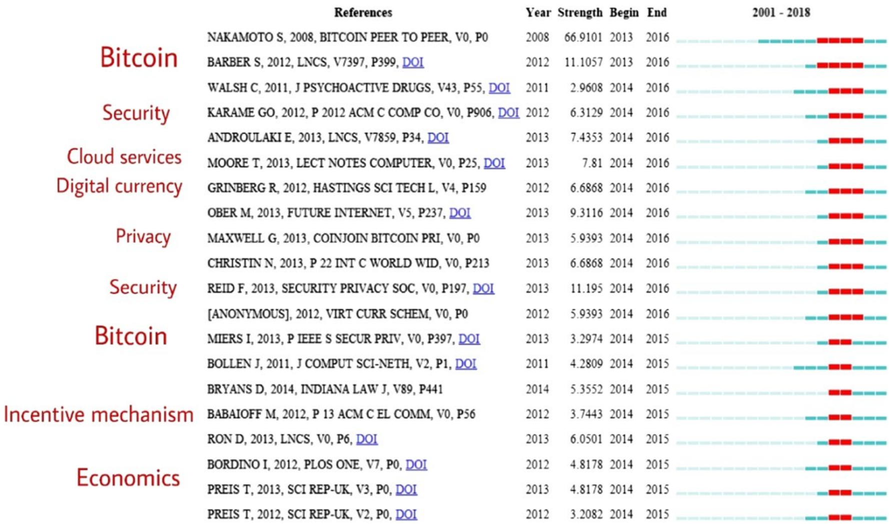

Fig. 12 The top 20 articles in the cluster with citation burst. The font size of the cluster represents the amount of references in this field within these 20 articles

Based on the emergent results from 2013 to 2016, the above viewpoints are verified. Thus, it is concluded that the time and intensity of the emergence of the research can be used to detect the trend in the future. Therefore, we analyzed the literature in the citation burst from 2016 to 2018 to obtain the latest research direction, as presented in Fig. 13.

Between 2016 and 2018, the literature research with high emergence value included two areas. On the one hand, it improved the underlying protocol of blockchain. For example, Eyal and Sirer (2014) and Garay et al. (2015), respectively, analyzed the core content of the Bitcoin protocol and proposed an improvement plan for the Bitcoin protocol. On the other hand, the research discusses the implementation of decentralization, encryption algorithms, consensus mechanisms, and other technologies related to blockchain. After 2016, the research frontier of blockchain field focused on the research of blockchain protocol, blockchain technology improvement, and blockchain application practice.

Fig. 13 The citation burst literature from 2016 to 2018. The font size of the cluster represents the amount of references in this field

5.2.2 Structural variation analysis

The citation-based metrics mentioned above rely on long-term citations, a factor that limits the prediction of the potential for new published literature. One solution to this limitation is to focus on the extent to which a newly published article affects the conceptual structure of the knowledge domain of interest. The idea is to determine the potential of an article to establish unusual or unexpected connections between/among different clusters (Chen 2012). The measurement does not involve the use of data. The influence of literature on the conceptual structure of existing knowledge is measured by the three structural variation metrics based on citation trajectory analysis: modularity change rate, inter-cluster linkage, and centrality divergence. According to the theory of scientific discovery, many significant contributions come from crossover ideas, and as such, transformative links can comprehensively reflect the three structural variation metrics generated by CiteSpace. Table 10 lists the 10 studies in the literature with the highest transformation link values.

These higher ranked articles represent three types of research that can be used as predictors, i.e., the review paper on blockchain (Seebacher and Schuritz 2017; Risius and Spohrer 2017), the research related to the application of cryptocurrencies, such as Bitcoin (Polasik et al. 2016; Peters et al. 2016; Matta et al. 2016; Bashir et al. 2016; Parino et al. 2018; Li and Wang 2017), and the research on the reliability of cryptocurrencies such as Bitcoin (Khalilov and Levi 2018; Giechaskiel et al. 2016). While the latter two are related studies on cryptocurrencies such as Bitcoin, the former of these two goes beyond cryptocurrencies to study issues related to their application scenarios. For example, Polasik et al. (2016) conducted a comprehensive empirical study on the payment and investment characteristics of Bitcoin in e-commerce scenarios and their impact on e-commerce behaviors. Accordingly, the study examined how the specific characteristics of countries, customers and companies interact with a proportion of Bitcoin sales. The latter of these two is based more on questions about the reliability of Bitcoin. For example, Khalilov and Levi (2018) conducted an overview and a detailed survey on the anonymity and privacy of the Bitcoin-based digital cash system before proposing guidelines for improving anonymity and privacy and discussing future research directions.

By combining the results of the two frontier probes, it is evident that some studies have focused on the problems caused by the application of cryptocurrencies in real life and the improvement of blockchain technology based on the specific limitations of these real-world applications. This study believes that these two types of research may well serve as the cutting edge in the field of blockchain.

6 Conclusions

Blockchain is a research field that attracts great attention and has great potential impact. Although the specific times of transformative applications of blockchain technology cannot be determined yet, blockchain provides researchers with a broad and interesting research space as demonstrated in this study. The histories of economy, science, and technology suggest that any new technology research and

Table 10 Top 10 articles with transformative potentials

Year	ΔM	ΔC–CL	ΔCen	TL	Title	References
2016	2.35	0.13	0.26	17	Price fluctuations and the use of Bitcoin: An empirical inquiry	Polasik et al. (2016)
2016	2.11	0.13	0	15	Opening discussion on banking sector risk exposures and vulnerabilities from virtual currencies: An operational risk perspective	Peters et al. (2016)
2015	2.46	0.22	0.08	11	Is Bitcoin's Market Predictable? Analysis of Web Search and Social Media	Matta et al. (2016)
2018	1.16	0.16	0	8	A Survey on Anonymity and Privacy in Bitcoin-like Digital Cash Systems	Khalilov and Levi (2018)
2016	0.98	0.08	0.03	8	What Motivates People to Use Bitcoin?	Bashir et al. (2016)
2018	0.8	0.34	0	6	Analysis of the Bitcoin blockchain: socio-economic factors behind the adoption	Parino et al. (2018)
2017	0.96	0.12	0	5	Blockchain Technology as an Enabler of Service Systems: A Structured Literature Review	Seebacher and Schuritz (2017)
2017	0.86	0.04	0	5	A Blockchain Research Framework	Risius and Spohrer (2017)
2017	0.8	0.08	0	5	The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin	Li and Wang (2017)
2016	0.53	0.06	0	5	On Bitcoin security in the presence of broken cryptographic primitives	Giechaskiel et al. (2016)

ΔM, ΔModularity; ΔC–CL, ΔC–C Linkage; ΔCen, ΔCentrality; TL, transformative links application have an overall influence on the real society only after several cycles of development. Accordingly, the impact of blockchain technology on the economy and society is also expected to emerge over a long evolutionary process, as the development of ledger technology also requires governmental regulation and business rules to be developed step by step. Therefore, it is important for the researchers to keep abreast of research trends so as to gain insight on the research directions at any time.

To the best of our knowledge, we are among the first to combine multiple complementary scientometric methods, such as literature co-citation analysis and literature burst analysis, with the LDA model to incorporate evolutionary stages into a visual and analytical study in the blockchain field. This mixed method increases the validity of our study.

This study provides a systematic view on how the blockchain field evolves, captures the state-of-the-art advances in this domain, and recognizes the emerging trends in this field. Based on our analysis in this study, the blockchain research mainly focuses on four aspects: underlying technology architecture, privacy and security, blockchain application in finance and blockchain application in new intelligent scenarios. While various disciplines contribute to the development of blockchain research, two specific subjects, computer science and business, have made the most significant contribution. Following the technology-application-new technology route, the research of blockchain now gradually turns to how to combine some new technologies like fog computing with blockchain technology to solve concrete issues in specific scenarios. By combining the results of literature detection and structural variation analysis, we believe two questions: how to solve the problems caused by the application of cryptocurrencies in the real world and how to improve blockchain technology based on the specific application requirements, may serve as the emerging trends in blockchain research. Specifically, we anticipate that future research on blockchain may be conducted from two topics:

1. How to apply digital currency in other areas? The influence research of the application of digital currency in other areas mostly stands from quantitative perspective to focus on the possible interrelationship between digital currency and people as well as organizations during the application process. As a relatively mature blockchain application, digital currency has been used in reality (real world), but relevant empirical research is little. With the wide application of digital currency, we can collect enough data to support our studies to transfer from theoretical exploration to empirical test. Considering the application of digital currency, the following directions may be potential interests of research:
   - Empirical research on the impact of digital currency application on E-commerce
   - Risk research of digital currency applicated in banking and financial services
   - Research on the relationship between digital currency and Internet public opinions
2. How to optimize technology in blockchain application? Research on technology optimization of blockchain applications mainly focuses on how to optimize the technology to solve the application limitations of blockchain. Blockchain technol-

ogy has not yet produced a disruptive application other than digital currency. Part of the reason may be its technical capabilities are yet far from meeting the needs of its application scenarios. The existing research on technology optimization has not fully take into account on the requirements of the application scenario, and furthermore, the combination of existing technology and new technology is also limited, and solely deployed new technology cannot meet the needs of certain application scenarios. For example, the authenticity of data before it is saved to the blockchain needs to be proved by some biometric techniques, and data privacy also requires technical support such as Edge computing and Zero-knowledge proof. Therefore, this kind of issues could be addressed by the followings:

- Research on the improvement of blockchain technology architecture (i.e. how to optimize performance metrics such as throughput and security based on application requirements).
- Research on the integration of blockchain with other new technologies (e.g. Zero-knowledge proof, Edge computing)

Accordingly, we want to emphasize two aspects which we believe are important for conducting blockchain research. One critical theme is to expand the application area of blockchain technology. Though computer science and business contribute a lot to blockchain research, it seems that there are still a few articles in some blockchain-related categories like surgery. As one of the two important questions in blockchain field is to rebuild blockchain technology based on application scenario, we believe that expanding the application area of block chain technology is very valuable. Besides, Fig. 4 suggests that the pioneering research results determine the influence of the author. The expanding application area is also an important way to improve the influence of researchers. Secondly, to solve the specific problems caused by the application of blockchain technology, we also recommend that researchers may pay attention to cross-disciplinary research. As the problems in application of blockchain technology require the combination of blockchain and some other new technologies, cross-disciplinary research can provide some novel insights. We believe that our study is especially helpful to those scholars who are going to initiate research in the blockchain field.

We also want to emphasize our research's contributions on industrial development. First, we summarize a panoramic view of blockchain, which helps the industry to quickly understand the development of blockchain and facilitate future business innovation. In fact, many people in the industry are curious about blockchain and are keen to explore potential business opportunities but have limited overall understanding of blockchain knowledge in the field. Therefore, the authors, literature, hotspots, and trends identified in our research can greatly reduce the time cost of exploring new areas and new business opportunities, thus taking a first-mover advantage for the industry practitioners. Secondly, our frontier research on blockchain will provide the possibility to solve industrial problems and business bottlenecks. Many companies have faced business bottlenecks, and they are keen for emerging technologies to empower their businesses and address these bottlenecks, some of which can certainly be solved by the frontier research of blockchain. For example, the combination of blockchain and edge computing technology is expected to solve problems regarding privacy paradox. Lastly, our identification of core research institutions and researchers provide a reference for people in the industry to seek cooperation, especially for companies that are keen to explore and research with research institutions.

Since the results of this study are based on the specific dataset, the data scope is limited by the given data sources and the query strategies employed. Future researchers can use more datasets and adopt new query strategies to improve the quality and quantity of the data. Furthermore, a more in-depth analysis in a few more years will further enhance our results in this paper since the blockchain field is expected to develop very quickly. Finally, due to the limitations of the Bibliometrics method, the influence of the newly published literature can only be identified as time goes by, and some newly published articles with high potential of influence cannot be recognized at this stage. However, we hope that in the near future, there will be new bibliometric methods to discover the potential of these articles.

Acknowledgements This work is supported by Sichuan Soft Science Research Program of China (Grant No. 2018ZR0366), Sichuan University Fundamental Research Funds for the Central Universities (Grant No. 2018ggzy-02) and Shenzhen Special Fund for Strategic Emerging Industries Development (Grant No. JCYJ20170818100156260).

References

Atzori L, Iera A, Morabito G (2010) The internet of things: a survey. Comput Netw 54:2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

Bashir M, Strickland B, Bohr J (2016) What motivates people to use bitcoin? Soc Inform Part II 10047:347–367. https://doi.org/10.1007/978-3-319-47874-6_25

Bentov I, Kumaresan R (2014) How to use bitcoin to design fair protocols advances in cryptology. Crypto Partt II 8617:421–439. https://doi.org/10.1007/978-3-662-44381-1_24

Blei DM, Ng AY, Jordan M (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022. https://doi.org/10.1162/jmlr.2003.3.4-5.993

Bonneau J, Miller A, Clark J, Narayanan A, Kroll JA, Felten EW (2015) SoK: research perspectives and challenges for bitcoin and cryptocurrencies. IEEE Symp Secur Priv 2015:104–121. https://doi.org/10.1109/Sp.2015.14

Chen CM (2006) CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci Technol 57:359–377. https://doi.org/10.1002/asi.20317

Chen CM (2012) Predictive effects of structural variation on citation counts. J Am Soc Inf Sci Technol 63:431–449. https://doi.org/10.1002/asi.21694

Chen CM (2013) Hindsight, insight, and foresight: a multi-level structural variation approach to the study of a scientific field. Technol Anal Strateg 25:619–640. https://doi.org/10.1080/09537325.2013.801949

Chen CM (2017) Science mapping: a systematic review of the literature. J Data Inf Sci 2:1–40. https://doi.org/10.1515/jdis-2017-0006

Chen CM, Leydesdorff L (2014) Patterns of connections and movements in dual-map overlays: a new method of publication portfolio analysis. J Assoc Inf Sci Technol 65:334–351. https://doi.org/10.1002/asi.22968

Chen CM, Ibekwe-SanJuan F, Hou JH (2010) The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis. J Am Soc Inf Sci Technol 61:1386–1409. https://doi.org/10.1002/asi.21309

Chen CM, Hu ZG, Liu SB, Tseng H (2012) Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. Expert Opin Biol Ther 12:593–608. https://doi.org/10.1517/14712598.2012.674507

Chen Y, Chen CM, Hu ZG et al (2014) Principles and application of analyzing a citation space. Science Press, Beijing
Cheng C, Zhang Y, Ning X, Yang Y (2018) Overseas blockchain research topics and prospects. E-Government 186(60):20–30
Christidis K, Devetsikiotis M (2016) Blockchains and smart contracts for the internet of things. IEEE Access 4:2292–2303. https://doi.org/10.1109/Access.2016.2566339
Ciaian P, Rajcaniova M, Kancs D (2016) The economics of BitCoin price formation. Appl Econ 48:1799–1815. https://doi.org/10.1080/00036846.2015.1109038
Conoscenti M, Vetro A, De Martin JC (2016) Blockchain for the internet of things: a systematic literature review. In: International conference of computer systems and applications (AICCSA). https://doi.org/10.1109/AICCSA.2016.7945805
Du MX, Ma XF, Zhang Z, Wang XW, Chen QJ (2017) A review on consensus algorithm of blockchain. In: IEEE international conference on systems, man, and cybernetics (SMC), pp 2567–2572
Eyal I, Sirer EG (2014) Majority is not enough: Bitcoin mining is vulnerable. Commun Acm 61:95–102. https://doi.org/10.1007/978-3-662-45472-5_28
Eyal I, Gencer AE, Sirer EG, van Renesse R (2016) Bitcoin-NG: a scalable blockchain protocol. In: 13th Usenix symposium on networked systems design and implementation (Nsdi '16), pp 45–59
Fernandez-Carames TM, Fraga-Lamas P (2018) A review on the use of blockchain for the internet of things. IEEE Access 6:32979–33001. https://doi.org/10.1109/Access.2018.2842685
Freeman L (1978) Centrality in social networks conceptual clarification. Soc Netw 1:281–310. https://doi.org/10.1016/0378-8733(78)90021-7
Fu YP, Zhang ZJ (2018) Bibliometric analysis of mobile learning literature based on CiteSpace. J Hainan Trop Ocean Univ 25:125–133
Gao WC, Hatcher WG, Yu W (2018) A survey of blockchain: techniques, applications, and challenges. In: 2018 27th international conference on computer communication and networks (ICCCN)
Garay J, Kiayias A, Leonardos N (2015) The Bitcoin backbone protocol: analysis and applications advances in cryptology. Eurocrypt Partt II 9057:281–310. https://doi.org/10.1007/978-3-662-46803-6_10
Giechaskiel I, Cremers C, Rasmussen KB (2016) On bitcoin security in the presence of broken cryptographic primitives. Comput Secur Esorics Part II 9879:201–222. https://doi.org/10.1007/978-3-319-45741-3_11
Hawlitschek F, Notheisen B, Teubner T (2018) The limits of trust-free systems: a literature review on blockchain technology and trust in the sharing economy. Electron Commer Res Appl 29:50–63. https://doi.org/10.1016/j.elerap.2018.03.005
Karame GO, Androulaki E, Roeschlin M, Gervais A, Capkun S (2015) Misbehavior in bitcoin: a study of double-spending and accountability. ACM Trans Inform Syst Secur 18:2. https://doi.org/10.1145/2732196
Khalilov MCK, Levi A (2018) A survey on anonymity and privacy in bitcoin-like digital cash systems. IEEE Commun Surv Tutor 20:2543–2585. https://doi.org/10.1109/Comst.2018.2818623
Kosba A, Miller A, Shi E, Wen Z, Papamanthou C (2016) Hawk: the blockchain model of cryptography and privacy-preserving smart contracts. In: 2016 IEEE symposium on security and privacy (Sp), pp 839–858. https://doi.org/10.1109/sp.2016.55
Kristoufek L (2013) BitCoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. Sci Rep 3:3415. https://doi.org/10.1038/srep03415
Li J, Chen C (2016) Citespace: text mining and visualization in scientific literature. Capital University of Economics and Business Press, Beijing
Li X, Wang CA (2017) The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin. Decis Support Syst 95:49–60. https://doi.org/10.1016/j.dss.2016.12.001
Luxburg UV (2007) A tutorial on spectral clustering. Stat Comput 17(4):395–416. https://doi.org/10.1007/s11222-007-9033-z
Matta M, Lunesu I, Marchesi M (2016) Is Bitcoin's market predictable? Analysis of web search and social media. In: Fred A, Dietz J, Aveiro D, Liu K, Filipe J (eds) Knowledge discovery, knowledge engineering and knowledge management. IC3K 2015. Communications in computer and information science, vol 631. Springer, Cham, pp 155–172. https://doi.org/10.1007/978-3-319-52758-1_10
Maurer B, Nelms TC, Swartz L (2013) "When perhaps the real problem is money itself!": the practical materiality of Bitcoin. Soc Semiot 23:261–277. https://doi.org/10.1080/10350330.2013.777594

Meiklejohn S, Pomarole M, Jordan G, Levchenko K, Mccoy D, Voelker GM, Savage S (2013) A fistful of bitcoins: characterizing payments among men with no names. Commun ACM 59:86–93. https://doi.org/10.1145/2896384

Miers I, Garman C, Green M, Rubin AD (2013) Zerocoin: anonymous distributed e-cash from bitcoin. In: IEEE symposium on security and privacy, pp 397–411 https://doi.org/10.1109/sp.2013.34

Nakamoto S (2008) Bitcoin: a peer-to-peer electronic cash system. http://bitcoin.org/bitcoin.pdf. Accessed 24 Oct 2018

Parino F, Beiro MG, Gauvin L (2018) Analysis of the Bitcoin blockchain: socio-economic factors behind the adoption. EPJ Data Sci 7:38. https://doi.org/10.1140/epjds/s13688-018-0170-8

Peters GW, Chapelle A, Panayi E (2016) Opening discussion on banking sector risk exposures and vulnerabilities from virtual currencies: an operational risk perspective. J Bank Regul 17:239–272. https://doi.org/10.1057/jbr.2015.10

Polasik M, Piotrowska AI, Wisniewski TP, Kotkowski R, Lightfoot G (2016) Price fluctuations and the use of bitcoin: an empirical inquiry. Int J Electron Commer 20:9–49. https://doi.org/10.1080/10864415.2016.1061413

Reid F, Harrigan M (2011) An analysis of anonymity in the bitcoin system. In: Security and privacy in social networks, pp 197–223. https://doi.org/10.1109/PASSAT/SocialCom.2011.7

Risius M, Spohrer K (2017) A blockchain research framework what we (don't) know, where we go from here, and how we will get there. Bus Inform Syst Eng 59:385–409. https://doi.org/10.1007/s12599-017-0506-0

Seebacher S, Schuritz R (2017) Blockchain technology as an enabler of service systems: a structured literature review. Lect Notes Bus Inf Process 279:12–23. https://doi.org/10.1007/978-3-319-56925-3_2

Vasek M, Thornton M, Moore T (2014) Empirical analysis of denial-of-service attacks in the bitcoin ecosystem. Lect Notes Comput Sci 8438:57–71. https://doi.org/10.1007/978-3-662-44774-1_5

Wood G (2014) Ethereum: a secure decentralised generalised transaction ledger. Ethereum Proj Yellow Pap 151:1–32

Xu M, Chen X, Kou G (2019) A systematic review of blockchain. Financ Innov 5:27. https://doi.org/10.1186/s40854-019-0147-z

Yang Y (2009) Knowledge measurement in domestic and foreign robotics. Dalian University of Technology

Zhao JL, Fan SK, Yan JQ (2016) Overview of business innovations and research opportunities in blockchain and introduction to the special issue. Financ Innov 2:28. https://doi.org/10.1186/s40854-017-0059-8

Zyskind G, Nathan O, Pentland A (2015) Decentralizing privacy: using blockchain to protect personal data. In: 2015 IEEE security and privacy workshops (Spw), pp 180–184. https://doi.org/10.1109/spw.2015.27