




Article

The Sustainable Development of Financial Topic Detection and Trend Prediction by Data Mining

Yunmei Liu ^{1,2} , Shuai Zhang ^{1,2} , Min Chen ^{3,*}, Yenchun Wu ^{4,5,*}  and Zhengxian Chen ⁶

¹ School of Information Management, Wuhan University, Wuhan 430072, China; emily19950904@whu.edu.cn (Y.L.); szhang@whu.edu.cn (S.Z.)

² Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China

³ Academy of Financial Research, School of Business, Wenzhou University, Wenzhou 325035, China

⁴ Graduate Institute of Global Business and Strategy, National Taiwan Normal University, Taipei 106, Taiwan

⁵ Leisure & Recreation Administration Department, Ming Chuan University, Taipei 111, Taiwan

⁶ School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA; chen3869@purdue.edu

* Correspondence: minchen@wzu.edu.cn (M.C.); ycwu@ntnu.edu.tw (Y.W.)

Abstract: Blockchain technology is the most cutting-edge technology in the field of financial technology, which has attracted extensive attention from governments, financial institutions and investors of various countries. Blockchain and finance, as an interdisciplinary, cross-technology and cross-field topic, has certain limitations in both theory and application. Based on the bibliometrics data of Web of Science, this paper conducts data mining on 759 papers related to blockchain technology in the financial field by means of co-word analysis, bi-clustering algorithm and strategic coordinate analysis, so as to explore hot topics in this field and predict the future development trend. The experimental results found ten research topics in the field of blockchain combined with finance, including blockchain crowdfunding, Fintech, encryption currency, consensus mechanism, the Internet of Things, digital financial, medical insurance, supply chain finance, intelligent contract and financial innovation. Among them, blockchain crowdfunding, Fintech, encryption currency and supply chain finance are the key research directions in this research field. Finally, this paper also analyzes the opportunities and risks of blockchain development in the financial field and puts forward targeted suggestions for the government and financial institutions.

Keywords: blockchain technology; financial sector; topic detection; artificial intelligence; bionic algorithm



Citation: Liu, Y.; Zhang, S.; Chen, M.; Wu, Y.; Chen, Z. The Sustainable Development of Financial Topic Detection and Trend Prediction by Data Mining. *Sustainability* **2021**, *13*, 7585. <https://doi.org/10.3390/su13147585>

Academic Editor: Cheolho Yoon

Received: 11 June 2021

Accepted: 5 July 2021

Published: 7 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As an innovative application model of computer technology and cryptography technology, blockchain is expected to transform from the current information Internet to the value Internet. At the same time, it will cause a new technological innovation and industrial reform on a global scale. Blockchain is a new application form of various computer technologies at the technical level, such as database, distributed ledger, smart contract, peer-to-peer transmission, consensus mechanism, distributed data storage, encryption algorithm, etc. At the physical level, the blockchain is a structure of chained blocks of data generated by cryptography and connected in chronological order. Together, blockchain is defined as a new distributed infrastructure and computing paradigm, which is used to analyze the data validation and storage; it uses the distributed node consensus algorithm for data generation and regeneration, cryptography technology to ensure data security and access process, and automation script code composed of smart contracts for programming and data operation [1]. Among them, as a distributed database ledger, blockchain has the following characteristics compared with traditional centralized databases: decentralization, trustlessness, information cannot be tamper with or deleted, traceability and collective

maintenance, which mainly solve the problems of information authenticity and system security [2,3]; openness and transparency, terminal privacy and complete autonomy can improve the availability of information, privacy and transaction certainty for the end-users [4,5]. Based on the underlying technology of blockchain, developers can define the content of the exchange, the conditions for the generation of new blocks and the corresponding authorization, etc., by themselves according to the application scenarios. For users, blockchain technology can reduce transaction costs, improve work efficiency and at the same time, have a better user experience.

Therefore, with its various characteristics, performance and advantages, blockchain technology has dramatically impacted all walks of life, especially the financial industry [6]. At first, with the development of technology, traditional finance and banking shortcomings gradually emerged, especially the lack and asymmetry of information. In this situation, how to accurately grasp the market opportunities and challenges of blockchain technology has become a hot topic in the current global financial field. Some economists have also made predictions: digital currency may become the primary monetary payment tool in the future, and blockchain technology, as the basic technology to support the development of bitcoin, has become a new direction after VR and AR have attracted attention [1,7]. Blockchain technology can completely change the primary technology in the financial field and solve the problems existing in traditional monetary payment, such as improving business processing speed, realizing open-source, significantly reducing the cost of verification and realizing further innovation of original value and realizing risk management [8,9].

With the technology development and application of blockchain, in January 2015, The Bretton Woods White Paper 2015 was released at the Bitcoin Conference in Miami [10], dividing the development of blockchain into three stages: Blockchain 1.0 programmable currency stage, mainly reflected in the application of bitcoin, thereby building a secure, uncentered and free new digital payment system. Blockchain 2.0 programmable finance phase, smart contracts and blockchain technology are applied in financial or economic markets, extending to the contracts of stocks, bonds, futures, loans, mortgages, property rights, smart assets, etc. With Blockchain 3.0 programmable social stage and the financial industry, the application of blockchain is broader and will also improve the operational efficiency and social productivity level in retail, Internet of Things, legal and other industries. Since 2016, the United Nations, the International Monetary Fund and the governments of many developed countries have released a series of reports on blockchain, such as the UN's "How Can Cryptocurrency and Blockchain Technology Play a Role in Building Social and Solidarity Finance?", the British government's chief scientific adviser report "Distributed Ledger Technology: Beyond Blockchain" and actively promote the application of blockchain technology in finance and other industries. The international financial giants have also seized the opportunity to increase the research and application of blockchain technology. For example, R3, the world's largest private chain alliance, has attracted 42 giants such as Wells Fargo Bank, Bank of America, Bank of New York Mellon and Citibank to participate in the application research of distributed ledger. Nasdaq launched its blockchain-based securities trading platform Linq in 2015 to manage private equity shares in a more transparent and traceable way. Therefore, blockchain technology in the financial field is pervasive and critical, which will significantly impact the financial industry.

In addition to financial organizations actively trying to promote the application and practice of blockchain in this field, the academic community has also conducted an extensive and in-depth exploration of the application of blockchain technology in the financial field and published a large number of relevant studies. These scientific papers contain rich and important research work and information resources. On the one hand, through the literature information, we can objectively understand the latest situation and development status of blockchain technology in the financial field and grasp the development of blockchain technology and its practice in the economic. On the other hand, we can also predict the future research trends and hotspots in this field and analyze the potential value of applying blockchain technology in the financial field and the risks and challenges.

In addition, blockchain technology and finance is an interdisciplinary, cross-technology and cross-field topic. This paper's literature review and topic analysis can provide a comprehensive and reasonable topic selection and research reference for researchers in related fields.

At present, only a few researchers conduct standardized bibliometric research and comprehensive data analysis on existing resources in the blockchain and financial field [11–13]. Therefore, this paper tries to use a variety of bibliometric analysis methods such as co-word analysis, bi-clustering algorithm, analysis of strategic coordinates, visual analysis and so on to retrieve, sort out, topic mining and trend prediction of international literature related to the application of blockchain in the financial field. The gCLUTO and BICOMB software are used in this study to generate bi-clusters and analyze co-words. Among them, co-word analysis is based on the literature's content characteristics, taking the subject words of the literature as the analysis object and counting the occurrence times of a group of words in the same literature by pairwise [14]. The more frequently the two keywords appear, the closer the relationship between the two topics will be. Then the topic structure and research hotspots represented by these words will be analyzed. Secondly, the bi-clustering algorithm can cluster the rows and columns of the matrix at the same time to cluster the global information and analyze the high-dimensional data [15]. Through the analysis of co-word clustering in a certain field, the research topics and hotspots in this field can be found. Finally, analysis of strategic coordinates is one of the visualization methods of co-word analysis. It takes the centripetal and density as parameters to draw two-dimensional coordinates, which is mainly used to describe the internal relations and interactions between fields of a specific research field [16]. The greater the number and intensity of connections between a subject area and other subject areas, the more central the subject area will become in the overall research effort.

In this paper, the above methods are combined to analyze the literature on the application of blockchain technology in the financial field. This research can objectively and scientifically obtain information on the current research status and development trends in this field to guide researchers and funding agencies in the selection of research topics, project selection, research design and project evaluation. This study mainly includes four parts. Firstly, in the method part, this paper's research methods, such as co-word analysis, bi-clustering algorithm, and strategic coordinate analysis, are introduced in detail. Secondly, in the data analysis part, we systematically explain the process of literature collection, cluster analysis and trend analysis. Third, in the result part, we cluster 10 topics in this field and make trend prediction for the related topics. Finally, in the discussion part, blockchain technology's status quo in the financial field is analyzed, and targeted suggestions are put forward to the government and financial institutions.

2. Method

2.1. Bi-Clustering Algorithm

The growing availability of bibliometric methods and tools enables collection and analysis of appropriate literature resources to judge the development status of a discipline and predict its development prospects. Since researchers publish more articles about topics that many find essential, analysis can reveal research hotspots and trends [17]. Co-word analysis is an important bibliometric analysis method that can identify the trends and hot topics of a subject. In a single article, if two words co-occur, then these two words may have a potential relationship. If two words frequently co-occur in the same paper, it means that they are closely related. Using the "relationships" between terms as measured by co-occurrence, statistical methods such as cluster analysis and factor analysis can then be applied. Keywords that meet preset thresholds can be considered a research hotspot based on the field and content [18]. Paule-Vianez et al. used co-word analysis to analyze relevant research literature on irrational investors' intervention in the financial market and found 13 main research topics in this field [19]. Topalli et al. took the relevant literature on the impact of economic transformation on enterprises in Central and Eastern Europe

from 1989 to 2013 as the object and analyzed the existing literature by using co-words to find the main influencing factors and their evolution process in the process of enterprise transformation [20]. Clustering analysis can be used to obtain the semantic relationships of the research topic. Compared with traditional clustering methods, the bi-clustering algorithm can cluster the rows and columns of a matrix simultaneously, can easily cluster global information and can be used to analyze high-dimensional data [21]. Using this method, Zhu et al. explored the current hotspots and potential topics in the field of stent implantation in the treatment of pancreatic diseases and obtained a total of 8 topic clustering results, which provide a reference for future scientific research in this field [22]. Wei et al. took human neural stem cells as an example and used bi-clustering analysis to statistically quantify the popular topics in relevant research literature in this field and obtained five kinds of results [23].

The topics clustered by the bi-clustering algorithm are defined. Given n vectors d_1 to d_n and an integer k , K clusters S_1 to S_k and their centroids C_1 to C_k are required to be found to minimize the following formula:

$$\max I_{\cos} = \sum_{r=1}^k \sum_{d_i \in S_r} \cos(d_i, c_r) \quad (1)$$

This function uses the cosine function to measure the similarity between the point and the centroid. The goal is to maximize the similarity between the points in the same cluster and the centroid. Substitute the formula for calculating the angle between vectors, and the criterion function is transformed into:

$$I_{\cos} = \sum_{r=1}^k \sum_{d_i \in S_r} \frac{d_i \cdot c_r}{\|c_r\|} \quad (2)$$

For clustering rows, repeated bisection was performed first. In this study, a randomized incremental optimization method was used to compute clustering solutions, which were greedy, had a low computational requirement and were produced in a high-quality fashion [24]. To obtain the k -way solution, the whole set was split first into two clusters by bisecting. In addition, one of the two clusters was further bisected, creating three clusters. k clusters were obtained after the process was completed. We computed clustering solutions based on each of these bisections, each optimizing a different criterion function [25]. Selecting the next cluster to be bisected was a key step in the algorithm. In all experiments by Steinbach and Karypis [26], using this method, we were able to find a fairly balanced clustering solution due to the largest cluster. Additionally, partitional clusters were clustered via agglomeration within each partition. Based on the partitional algorithm, a single hierarchical tree was constructed. Hierarchical trees show that objects in an aggregation process are linked by showing their orders of merging together. In the merge, since objects were compared according to their pairwise similarity, objects that were close in proximity were more similar than objects that were far away. Clusters found in the clustering process are shown at the top of the matrix. Each row of objects in a cluster was continuously arranged. On top of that, the data matrix was transposed into a hierarchical tree by using agglomerative clustering.

In this study, we conducted a bibliometric analysis by co-word analysis and visualization on the topic of blockchain technology in the financial sector. This analysis allowed assessment of the current research status of blockchain technology in the financial industry.

2.2. Strategic Coordinates Analysis

We established the hotspot strategic coordinates of blockchain application in the financial field to predict its research trends. The strategic coordinate chart was proposed

by Law et al., in 1998, which is mainly used to describe the internal relations and mutual influences of a specific research field [15]. In the strategic coordinates, X-axis is the centripetal degree, representing the strength of the interaction between domains. Y-axis is the density, representing the strength of the internal connection within a certain field. The strategic coordinates are the two-dimensional coordinates drawn with centripetal and density parameters, which can generalize the structure of a subdomain within a domain. The typical structure is that the horizontal axis represents the centripetal degree, the vertical axis represents the density and the origin of the coordinates is at the median or average of the two axes.

The centrality calculation formula is as follows:

$$C = \frac{1}{n} \times \sum_{j=1}^{n \neq i} N(A_i \rightarrow B_j) \quad (3)$$

The density calculation formula is as follows:

$$D = \frac{1}{n} \times \sum_{i=1}^n N(A_i) \quad (4)$$

This map divides the subject area of each two-dimensional space into four quadrants, which can be used to describe the research development status of each topic. By analyzing the coordinate graph in Figure 1, along the direction of the X-axis arrow, the more to the right the position of the class cluster in the coordinate, the greater the centripetal degree of the class cluster, which also indicates that the class cluster is more closely related to other class clusters. That is to say, members of the class cluster are related to internal members and closely related to members of other class clusters. Because this kind of subject words and many subject words can form a collocation in the same literature, it shows that these subject words occupy a relatively important position in the subject field. Therefore, this kind of theme is not easy to disappear, has a strong vitality and can maintain a high word frequency in a long period of time. Along the Y-axis arrow direction, the higher the cluster's position in the coordinate is, the higher the density of the cluster is, indicating that the density of the connection between the members of the cluster is strong. In this way, the value of the association between class clusters and the value of the association between the members of the class cluster constitutes the X-axis and Y-axis of the coordinate axis. The origin where the two axes intersect is the centripetal and density of all clusters, dividing all clusters into four quadrants. Lu et al. conducted a hotspot analysis of relevant literature on cancer immunotherapy, determined strategic coordinates by using co-word matrix and cluster analysis and analyzed the distribution of various organs or diseases and the subcategories of tumor immunotherapy, to determine important fields for future scientific research [27].

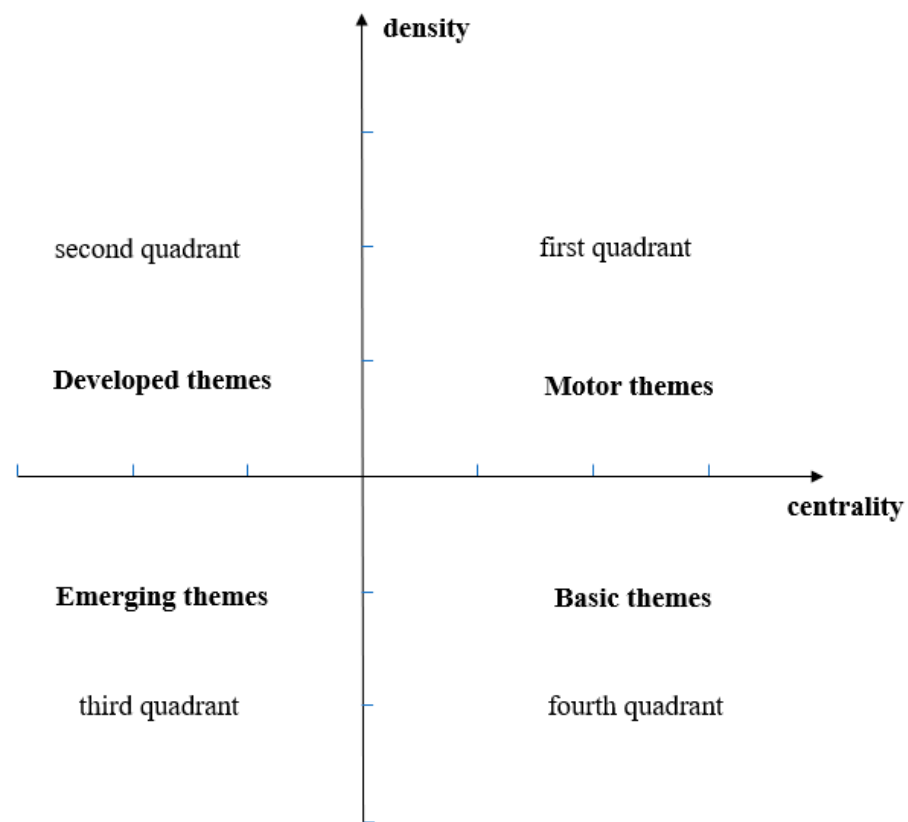


Figure 1. Example diagram of strategic coordinates.

3. Data Analysis

3.1. Literature Collection

To collect the literature systematically and comprehensively, this paper adopts the following three steps to collect the literature. (1) Constructional retrieval. First, free word search was carried out in Google Scholar, and a preliminary search formula was constructed by browsing the titles and keywords of literatures with high citation frequency. Several rounds of search and repeated adjustment were performed, until no new search terms and search relations were found. Second, the search query was further referred to Cai's review (Cai, 2018). Finally, the search query of this study was determined as follows: blockchain AND (finance OR "financial markets" OR "investment" OR "asset pricing" OR "financial intermediation" OR "capital markets") [28]. (2) Database retrieval. The foreign language databases retrieved include Web of Science, Elsevier, ScienceDirect, Wiley Online Library and EBSCO. Literature types include journal papers and conference papers. The literature retrieval time was from 7 January to 10 January 2021. (3) Extended retrieval. We also screened all articles published in the leading 35 journals in the finance field, ranking 4*, 4 or 3 in the 2018 Academic Journal Guide (AJG) by the Chartered Association of Business Schools. A graph was created to illustrate the criteria for article selection and justify them (see Figure 2). A total of 759 research documents were obtained, and the final sample literature list was formed (see Figure 3). The geographic dispersion of the sample of studies and the number of citations per study each year were created (see Figures 4 and 5).

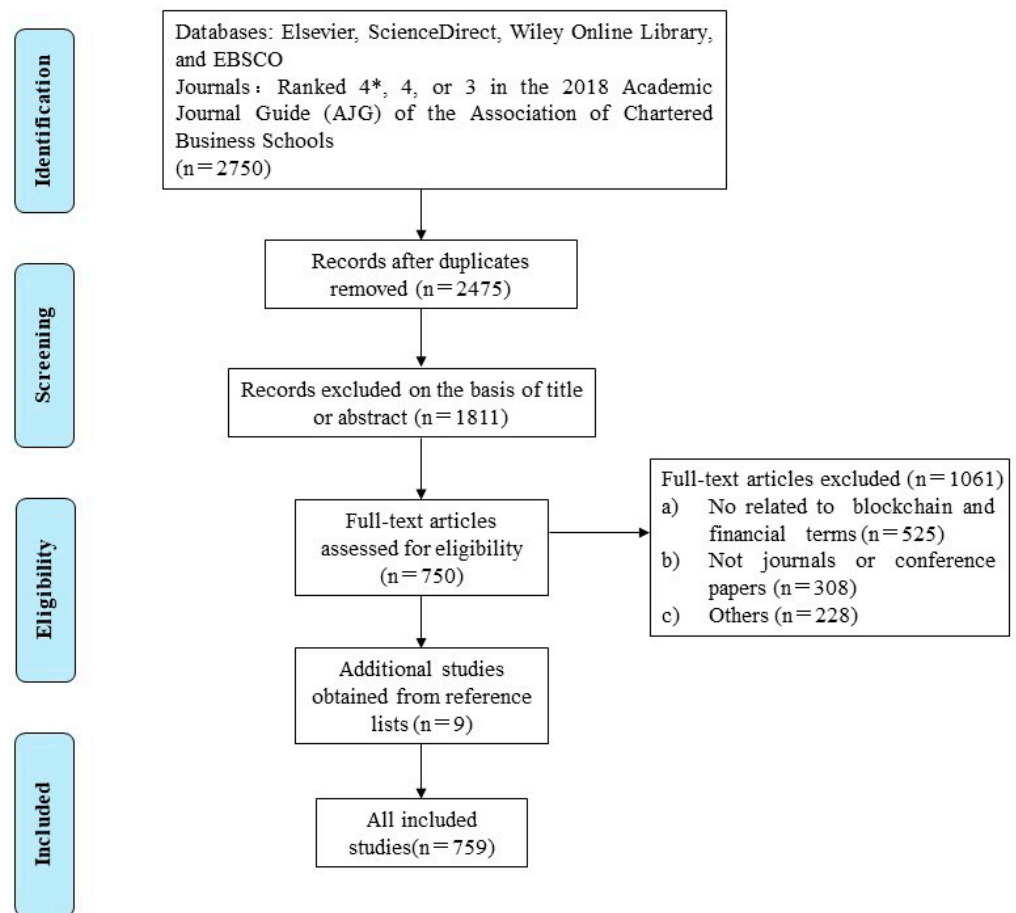


Figure 2. Flow diagram for identifying blockchain and financial studies.

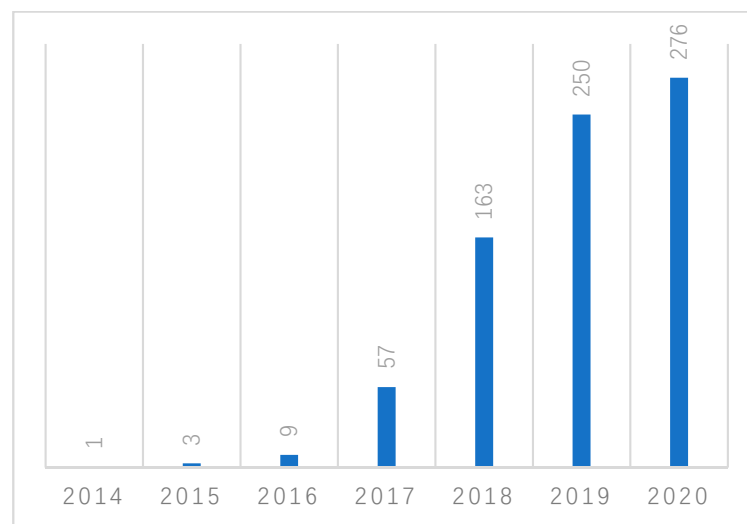


Figure 3. Literature statistics of blockchain technology applied to the financial field.

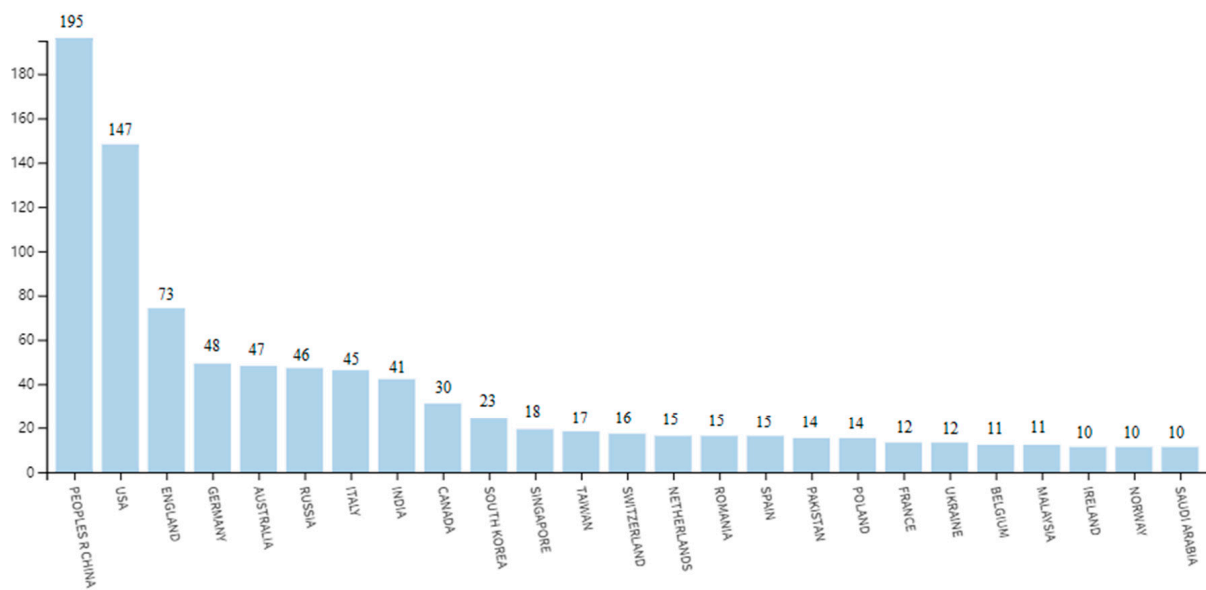


Figure 4. Geographical distribution of studies.

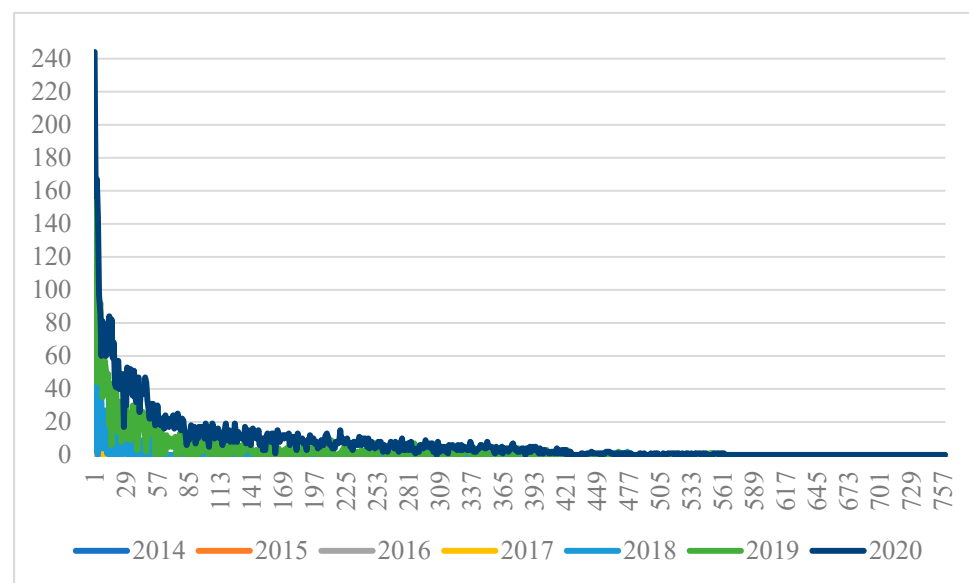


Figure 5. The number of citations per study each year.

3.2. Bi-Clustering Analysis

3.2.1. High-Frequency Keyword Determination

Through the word frequency statistics of the keywords in the sample literature of blockchain technology applied in the financial field, the hotspots and research directions of academic research in this field can be reflected. The higher the occurrence frequency of a keyword, the higher the concentration of the research content related to this keyword, and the more likely it is to be the critical research direction in this field. The keyword frequency statistics of 759 sample documents were found, with a total of 3731 keywords. Data cleaning for 3731 keywords, including conversion of synonyms or synonyms, unified capitalization of English words (e.g., blockchain & Blockchain), singular and plural (e.g., cryptocurrency & cryptocurrencies), spelling (e.g., hyper-ledger & hyperledger), etc. On this basis, concerning the low-frequency word boundary formula of high-frequency words [29], this study puts forward the keywords with word frequency greater than five, and a total of 116 high-frequency keywords are obtained. Some high-frequency keywords are shown in Table 1.

Table 1. High-frequency keywords (partly).

No.	High-Frequency Keyword	Word Frequency	No.	High-Frequency Keyword	Word Frequency
1	blockchain	459	21	innovation	20
2	smart contracts	126	22	trust	19
3	cryptocurrency	113	23	payments	17
4	bitcoin	92	24	transaction	16
5	distributed ledger	62	25	decentralization	15
6	fintech	54	26	hyperledger	15
7	securities	43	27	big data	14
8	supply chain	38	28	industry 4.0	14
9	ethereum	38	29	artificial intelligence	14
10	initial coin offering	37	30	bionic algorithm	13
11	finance	37	31	cloud computing	13
12	privacy	35	32	entrepreneurial finance	13
13	Internet of Things	34	33	electronic health record	13
14	insurance	33	34	cybersecurity	13
15	token	29	35	globalization	12
16	investment	28	36	machine learning	12
17	consensus algorithms	27	37	digitalization	12
18	digital currency	25	38	banking	12
19	healthcare	24	39	financial markets	11
20	crowdfunding	22	40	decentralized applications	11

3.2.2. Literature-Keyword Matrix Construction

On the basis of the statistics of high-frequency keywords, the occurrence times of each high-frequency keyword in the sample literature were counted, and the biblio-keyword matrix (part) was established, as shown in Table 2. In the first line, the Arabic numerals 001, 002, 003, etc. respectively represented sample literature 1, sample literature 2, sample literature 3, etc. The first is the high-frequency keywords. In the literature—keyword matrix, “1” represents the high-frequency keyword’s occurrence in the sample literature, and “0” represents the absence of the high-frequency keyword in the sample literature. For example, the keyword “smart contracts” had appeared in sample literature 2 (“1”), and it had not appeared in sample literature 1 (“0”).

Table 2. Literature-keyword matrix (partly).

Keyword	001	002	003	004	005	006	007	008	009	010
blockchain	1	1	1	1	1	1	1	1	1	1
smart contracts	0	1	0	0	0	0	1	0	0	1
cryptocurrency	1	0	0	0	1	0	0	0	0	0
bitcoin	0	0	0	0	1	0	0	0	1	0
distributed ledger	0	0	1	0	0	0	0	1	0	0
fintech	0	0	0	0	0	0	0	1	0	0
securities	0	0	0	1	0	0	0	0	0	0
supply chain	0	0	1	0	0	0	0	0	0	0
ethereum	0	0	0	0	0	0	0	0	0	1
initial coin offering	0	0	0	0	0	1	0	0	0	0

3.2.3. Bi-Clustering Matrix Atlas

Next, the bi-clustering analysis was performed using gCLUTO software [30]. The bi-clustering options were as follows: algorithm: repeated bisection; similarity function: cosine; criterion function: I2. Each cell’s colour in the bi-clustering matrix represents the relative occurrence frequency of the sample literature corresponding to this row and the high-frequency keywords corresponding to this column. The darker the color, the higher the relative frequency. White indicates zero relative frequency. Horizontal lines in the

figure separate color squares, and the areas separated by the horizontal lines represent the categories of clustering. To determine the best number of clusters, we repeated the bi-clustering several times by selecting different numbers of clusters. The lowest average similarity between classes (ESim) and the highest similarity within class (ISim) values can be used as the optimization results (Table 3). The bi-clustering results are shown in Table 3. Matrix visualization of bi-clustering is shown in Figure 6 and mountain visualization of bi-clustering is shown in Figure 7.

3.3. Trend Analysis

Co-word analysis can be used as a tool to understand and describe the relationship between scientific topics. Co-word analysis can help distinguish the local environment and each research topic [31]. By using Excel, the co-word matrix composed of high-frequency words was used to calculate the intra-class link averages and the inter-class link average (Table 4), allowing calculation of centrality and density, respectively (Table 5). Using two-dimensional coordinates with centrality and density as parameters, a graph was constructed to describe certain topics' internal integrality and the effects of their interactions. In a strategic diagram, the intensity of the interaction of topics is expressed with the centrality's X-axis. The greater the number and intensity of the links between one subject area and other disciplines, the more central the subject area are to the overall research. The centrality of a category is calculated by the strength of the links between the category's main items and other categories. The Y-axis represents the density, indicating the strength of the internal integrality within a given category, and the level of each category can maintain and develop itself.

Table 3. Result parameters of bi-clustering.

Cluster	ISim	ISdev	ESim	ESdev
0	0.235	0.046	0.015	0.005
1	0.163	0.020	0.014	0.010
2	0.165	0.034	0.016	0.017
3	0.155	0.034	0.011	0.005
4	0.155	0.026	0.011	0.005
5	0.150	0.024	0.015	0.009
6	0.145	0.025	0.012	0.007
7	0.138	0.018	0.009	0.004
8	0.144	0.036	0.022	0.027
9	0.127	0.016	0.010	0.007

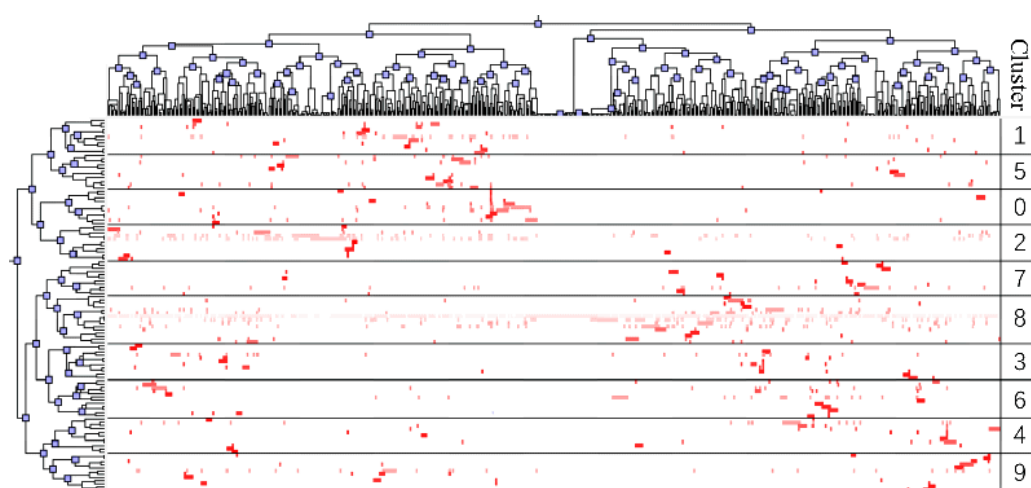


Figure 6. Matrix visualization of bi-clustering.

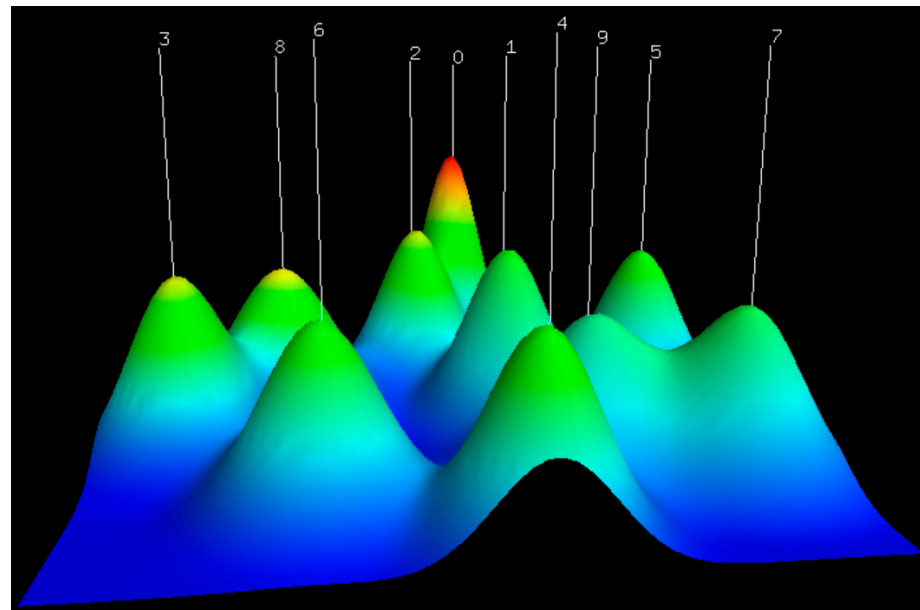


Figure 7. Mountain visualization of bi-clustering.

Table 4. A co-word matrix of high-frequency keywords (localized).

Keywords	Blockchain	Smart Contracts	Cryptocurrency	Bitcoin
Blockchain	521	119	84	74
Smart contracts	119	126	7	9
Cryptocurrency	84	7	113	48
Bitcoin	74	9	48	92
Distributed ledger	59	18	6	11
Fintech	40	2	12	6
Securities	40	10	5	4
Supply chain	31	9	1	2
Ethereum	32	27	11	10
Initial coin offering	28	2	19	5

Table 5. The centrality and density of the 10 clusters.

Cluster	Intra-Class Link Averages	Density-Y	Inter-Class Link Average	Centrality-X
0	44.10	5.07	26.60	3.44
1	55.92	0.93	22.46	15.27
2	20.83	11.87	4.67	21.82
3	26.80	−17.93	3.60	−13.86
4	38.83	−6.20	15.33	−1.82
5	26.70	−12.33	9.20	−13.96
6	25.14	−16.68	4.86	−15.51
7	63.12	31.67	53.20	22.46
8	29.24	−9.65	11.88	−11.42
9	27.73	−16.81	4.73	−12.93
Average	40.66		21.53	

4. Result

4.1. Topic Analysis Result

After the bi-clustering analysis, the application of blockchain technology in the financial field is grouped into ten main themes (Table 6). Combined with the high-frequency keywords in each category of the bi-clustering matrix graph, each category's names can

be given to the maximum extent to include its meaning. Class 0 contains high-frequency keywords such as “initial coin offering”, “token”, “crowdfunding”, and “entrepreneurial finance”, so Class 0 is defined as “blockchain crowdfunding”. Class 1 contains high-frequency keywords such as “fintech”, “big data”, “artificial intelligence”, “bionic algorithm”, so Class 1 is defined as “financial technology”. Class 2 contains high-frequency keywords such as “cryptocurrency”, “bitcoin”, “digital currency”, “virtual currency”, so Class 2 is defined as “encryption currency”. Class 3 contains high-frequency keywords such as “consensus algorithms”, “proof-of-work”, “access control”, so Class 3 is defined as “consensus mechanism”. Class 4 contains high-frequency keywords such as “internet of things”, “industry 4.0”, “cloud computing”, so Class 4 is defined as “intelligent manufacture”. Class 5 contains high-frequency keywords such as “investment”, “payments”, “banking”, “financial markets”, so Class 5 is defined as “digital finance”. Class 6 contains high-frequency keywords such as “insurance”, “healthcare”, “electronic health record”, so Class 6 is defined as “medical insurance”. Class 7 contains high-frequency keywords such as “supply chain”, “smart city”, “integration”, so Class 7 is defined as “supply chain finance”. Class 8 contains high-frequency keywords such as “blockchain”, “smart contracts”, “distributed ledger”, so Class 8 is defined as “smart contracts”. Class 9 contains high-frequency keywords such as “finance”, “globalization”, “financial inclusion”, “innovation”, so Class 9 is defined as “financial innovation”.

Table 6. Theme labels and keywords.

Cluster	Keywords	Theme Labels
0	Initial coin offering/token/crowdfunding/entrepreneurial finance/sustainability/crypto assets/mining/venture capital/smes finance/ scalability/institutional economics	Blockchain crowdfunding
1	Fintech/ big data/artificial intelligence/bionic algorithm/machine learning/markets/algorithmic trading/China/ financial services	Financial technology
2	Cryptocurrency/bitcoin/digital currency/virtual currency/money laundering/ trade/cryptoeconomics/e-money/databases/time bank token	Encryption currency
3	Consensus algorithms /proof-of-work/access control/services/p2p computing/zero-knowledge proof/cryptography/hash function	Consensus mechanism
4	Internet of things/ industry 4.0/cloud computing/cybersecurity/business/ smart grid/ecosystems/government/future accounting/ charity/cloud services	Intelligent manufacture
5	Investment/payments/banking/financial markets/financial risks/digital finance/ asset pricing/systemic risk/portfolio diversification/digital economy/ digitalization/business models/public finance/taxation/ auditing/stock exchange	Digital finance
6	Insurance/healthcare/electronic health record/decentralized applications/medical services/vehicles/internet finance/game theory/companies/software/health records/distributed systems/healthcare services/covid-19	Medical insurance

Table 6. Cont.

Cluster	Keywords	Theme Labels
7	Supply chain/smart city/integration/intelligence/traceability/consortium blockchain/project bank accounts/deep learning/risk diversification/agricultural products	Supply chain finance
8	Blockchain/smart contracts/distributed ledger/securities/ethereum/privacy /trust/transaction/decentralization/hyperledger/e-voting/performance/secure elements/authentication service	Smart contracts
9	Finance/globalization/financial inclusion/innovation/credit markets/networks/knowledge/financial applications/economics/international finance/e-commerce/ education	Financial innovation

4.2. Trend Result Analysis

The horizontal axis of the strategic coordinate indicates the centrality, the vertical axis represents the density and the first quadrant is the upper-right corner, then moving clockwise, the second quadrant, the third quadrant and the fourth quadrant. As we can see from Figure 8, clusters 0, 1, 2 and 7 are in the first quadrant, representing the corresponding category in the central and core field. The research maturity of these four subject categories is relatively high and will continue to be the mainstream direction and hot issue in the field of blockchain research in the future. Clusters 3, 4, 5, 6, 8 and 9 are in the third quadrant, indicating that their corresponding categories are in relatively peripheral cold fields (Figure 8). These six types of topics are emerging topics in the blockchain field. Although the current research enthusiasm is not high, these topics are expected to become an important exploration direction in the future.

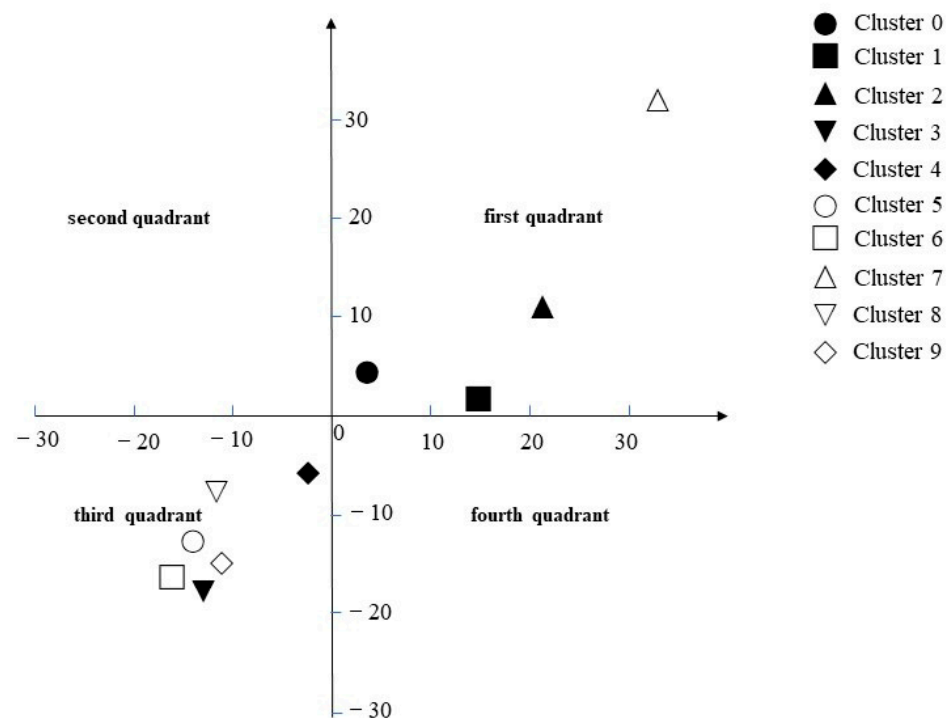


Figure 8. Strategic diagram of clusters.

In the current era when computer technology continues to promote social changes, important changes in the financial field are always the combination of original technology and computer technology. Blockchain, as one of the latest innovations in the computer field, although there are still some technical shortcomings, with the continuous improvement in the future, its model will have an important impact on the financial field. It will also lead to major changes in the traditional financial system and will also have an impact on the existing trust mechanisms in the financial sector. It is believed that in the future, the existing bilateral trust or central trust mechanism will be replaced by a new type of credit mechanism of social common credit. Financial institutions can make full use of the technical advantages of blockchain technology to innovate new application models.

This paper discovers ten research topics of blockchain in the financial field through topic detection and data mining of related research literature. According to the trend prediction results, blockchain crowdfunding, financial technology, encryption currency and supply chain finance are distributed in the first quadrant. These four types of research topics will be the motor research themes. Blockchain technology is a distributed ledger technology with great development prospects in the financial field. In the next few years, this technology will play a more important role in the digital currency technology architecture, optimizing the financial credit system and influencing the financial technology architecture. The key role fully reflects the application value of blockchain technology in the financial market.

Secondly, the remaining six research topics are all distributed in the third quadrant and belong to emerging themes. For example, the third category of topics is the issue of consensus mechanisms. The consensus mechanism of the blockchain will generate an intelligent trust store that records various reliable information. The information cannot be tampered with through the consensus mechanism, which largely solves the trust problem in the financial industry. Internet trust has now become an important part of the entire social and economic trust system. Moreover, with the in-depth development of the digital economy, the share of consensus mechanisms will become higher and higher. Therefore, in the future, it is necessary to add Internet trust in addition to the traditional trust system and to add institutional arrangements that adapt to Internet trust, especially an algorithm-based consensus mechanism that is highly compatible with the Internet economy.

The fourth category of emerging themes is the field of intelligent manufacturing. Intelligent manufacturing is a human-machine integrated intelligent system composed of intelligent machines and human experts. Blockchain technology can solve the problems of information asymmetry and resource non-sharing in the traditional manufacturing industry. It can monitor all aspects of production and manufacturing for a long time, improve the safety and reliability of production and manufacturing and can also help the company's internal operations. Production brings value. Therefore, in future research, blockchain technology will have a profound impact on the global economic structure and will have profound changes in corporate architecture, Internet industry ecology, social order and even production relations, and it will change the industry ecology of modern manufacturing.

The fifth category of emerging themes is the field of digital finance. Digital finance refers to the use of digital technology by traditional financial institutions and Internet companies to realize financing, payment, investment and other new financial business models. Digital finance breaks the time and space constraints, cost constraints, information barriers and customer exclusion of financial services, making finance better serve the real economy. As a cutting-edge digital technology, blockchain technology can provide powerful technical support for digital finance, making the economic development of digital financial services more inclusive and effective. Blockchain technology is still facing certain difficulties in the legal system, regulatory system, talent construction, technical level and other aspects of the digital finance field. In future research, it is necessary to continuously improve the blockchain technology and actively explore the depth of "blockchain + finance". The integrated innovation model continuously improves the supply of digital financial products to better promote the high-quality development of digital finance.

In the sixth category of medical insurance topics, insurance business is affected by Internet information technology; secondly, the insurance industry is affected to a certain extent by blockchain technology. The application of blockchain technology in the insurance industry can optimize the untrustworthy problems caused by the Internet based on the conditions of the Internet. At the same time, the technology is also a driving force for the continuous innovation and development of the insurance industry. It is not only conducive to reducing transaction costs but also improving the level of trust in it. In future research, blockchain technology can be applied to optimize the insurance process and better construct the customer's data information database, which will help business personnel efficiently find all information about customers and to realize the intelligence of business processing and improve it. Second, the sharing of insurance information is affected by the blockchain, which can not only realize the openness and transparency of information but also realize the sharing of resources, which is particularly important for the overall development of the insurance industry.

In the eighth category of smart contract topics, the emergence of blockchain technology redefines the concept of smart contracts. Therefore, in future research, smart contracts can be embedded in various activities of traditional finance, providing innovative solutions for financial development. Smart contracts have the characteristics of automatic operation, self-management, etc., without third-party intervention and supervision, which not only saves manpower and material resources but also guarantees the fairness of the contract to a large extent. The smart contract is the activator of the blockchain, which provides a programmable operating mechanism for the static underlying data; the characteristics of the smart contract can also include the complex behavior of each node in the distributed system, which helps to promote the use of blockchain technology in manual labor and various applications in intelligent systems.

The last category of emerging topics is about financial innovation. Fintech provides new ideas for alleviating the financing problems of SMEs. Its core is to integrate information technology, big data, etc. into the decision-making process of financial institutions, so as to improve the information screening and risk control capabilities of financial institutions. As one of the representative technologies of financial technology, the importance of blockchain technology application in financial innovation is becoming increasingly prominent, especially in the field of supply chain finance. In the future of this research direction, a complete blockchain data governance system should be established, and information technologies such as the Internet of Things, big data and artificial intelligence should be used in the data chain link to ensure the quality and effectiveness of the application of blockchain in financial innovation.

5. Discussion

In summary, the integration of blockchain and the financial industry has attracted extensive attention from many scholars worldwide. Scholars have analyzed the nature of blockchain from theoretical research and studied its operating mechanism and transaction mechanism in the financial field. At the same time, practical research in this field is also active. The development and implementation of numerous application scenarios of blockchain in the financial field provide directions and cases for further research on this technology. In addition, the key trends of the research on the application of blockchain in the financial field include theoretical and applied research. One is the theoretical research on blockchain crowdfunding, bionic algorithms, cryptocurrency, consensus mechanism and the Internet of Things [32–34]. The second is the role of blockchain in equity registration, letter of credit and international exchange; applied research in the field of cross-border payment; and the construction of basic platform based on blockchain [35–37].

The research contribution of this paper, which is different from other review papers in the financial blockchain field, is: First, a standardized retrieval strategy is selected, and a systematic and diversified analysis of the more comprehensive literature data in the financial blockchain field is formed. Secondly, this paper comprehensively uses a variety of

comprehensive, accurate data mining techniques and methods, such as co-word analysis, bi-clustering algorithm, strategic coordinate map, visual analysis, etc., revealing a research map in this research field from the perspective of informatics. Finally, in addition to the analysis of the current research status of the financial blockchain field, this paper also makes a further analysis of the future research trends in this field.

Although blockchain technology has been born for nearly 12 years since 2009, its real attention and research time is still short. At present, its application is still only extended from the digital currency stage of 1.0 to the smart contract stage of 2.0 and has begun to penetrate into other non-financial fields [38]. At the research level, most of the application scenarios of blockchain technology are still in the laboratory stage and most of the applications of blockchain deal with some simple problems. However, in practical application, because the technology is not mature, its operability, application scenario design and research and development capabilities still need to be further improved [39,40]. Therefore, blockchain technology's current development and application is still the coexistence of opportunities and risks, driving force and challenges [41].

At present, the development of blockchain technology is still in its infancy and improvement stage. Many problems need to be solved, especially for the high-risk digital finance field, which involves the immediate interests of consumers, participants and other relevant issues. The application is likely to face the problem of energy consumption [42]; the block generation requires the participants to do much meaningless calculation; such calculation is very energy intensive. Then there is regulation. The decentralization, traceability and anonymity of blockchain weaken the concept of national regulation. In the absence of regulatory-coverage, the market's profit-seeking nature will lead to the application of blockchain technology in illegal fields, which may be used by criminals for money laundering, fraud and tax evasion [43,44]. Third is the understanding of different countries [45]. It is a challenge to the central banks' authority to adopt some virtual currency as the equivalent to realize the real-time global settlement. Finally, there are technical issues that are difficult to overcome, including poor correctability, low system security, long latency, redundant storage and so on.

Therefore, all sectors of society should take active measures to actively welcome the industry changes brought by the new blockchain technology. While seriously studying the technology feasibility, we should guard against the risks that may be brought by the new technology. They should be fully aware of the risks brought about by new technologies and gradually establish a softer regulatory system for central banks and regulatory authorities. On the one hand, it should continue to encourage the development of relevant enterprises and provide a good regulatory system to greatly reduce consumers' rights and interests and the risks generated in the process of controlling transactions. On the other hand, it should launch legal digital currency in a timely manner, vigorously cultivate professional talents and formulate some incentive measures to encourage more financial enterprises to use blockchain technology to carry out their own business. For financial institutions, blockchain technology should be viewed with an open mind to provide better financial services to the real economy. On the one hand, it should fully tap the potential of blockchain technology, research feasible landing schemes and comprehensively improve internal management, risk prevention and control and profitability. On the other hand, it actively participates in the formulation of new technology standards and the design of application programs such as blockchain at home and abroad, so as to better adapt to international development, carry out inter-agency cooperation and expand the global market.

In summary, new technology needs to develop its own technical details at the beginning of development, especially for the integration of blockchain, which is the underlying technology and closely integrated with the practical application. Huge space and broad prospects mean that it needs to be polished for a long time, involving many aspects such as finance, law, taxation and even morality. Secondly, a single technology cannot drive the economy. Blockchain applications must be combined with emerging technologies such as the Internet of Things, cloud computing, big data and artificial intelligence [46,47]. There-

fore, as one of the most promising emerging Internet technologies at present, blockchain technology deserves the attention of all sectors of society to study and research and realize scientific construction and continuous optimization of its scientific research system [48,49], so as to improve the application level of the technology and better promote the sustainable development of digital finance.

Author Contributions: Y.L.: formal analysis, data curation, writing—original draft preparation; S.Z.: formal analysis, data curation, writing; M.C.: writing—review and editing, visualization, supervision; Y.W.: writing—review and editing; visualization, supervision; Z.C.: formal analysis, data curation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (grant number 71420107026 and 71661167007). This research was also supported by the project of “Executives’ Dynamic Relationship Network, System Differences, and Chinese Enterprises’ Internationalization Strategy: Internal Mechanism and Empirical Test” (Grant No.: 21NDJC144YB) (2021 Philosophy and Social Science Project of Zhejiang Province).

Institutional Review Board Statement: The study was approved by the Ethics Committee of Wuhan University (13/1524).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Acknowledgments: We wish to thank every anonymous referee for important insightful comments and suggestions.

Conflicts of Interest: All Authors declare that they have no conflict of interest.

References

1. Treleaven, P.; Brown, R.G.; Yang, D. Blockchain Technology in Finance. *Computer* **2017**, *50*, 14–17. [CrossRef]
2. Workie, H.; Jain, K. Distributed ledger technology: Implications of blockchain for the securities industry. *J. Secur. Oper. Custody* **2017**, *4*, 347–355.
3. Wang, H.; Chen, K.; Xu, D. A maturity model for blockchain adoption. *Financ. Innov.* **2016**, *2*, 12. [CrossRef]
4. Kosba, A.; Miller, A.; Shi, E. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In Proceedings of the IEEE Symposium on Security and Privacy, San Jose, CA, USA, 26 May 2016.
5. Yue, X.; Wang, H.J.; Jin, D.W. Healthcare data gateways: Found healthcare intelligence on blockchain with no-vel privacy risk control. *J. Med. Syst.* **2016**, *10*, 1–8. [CrossRef] [PubMed]
6. Fanning, K.; Centers, D.P. Blockchain and its coming impact on financial services. *J. Corp. Acc. Financ.* **2016**, *5*, 53–57. [CrossRef]
7. Don, T.; Alex, T. *Blockchain Revolution: How the Technology behind Bitcoin Is Changing Money, Business and the World*; Penguin: London, UK, 2016.
8. Sun, J.; Yan, J.; Zhang, K.Z.K. Blockchain-based sharing services: What blockchain technology can contribute to smart cities. *Financ. Innov.* **2016**, *2*, 26. [CrossRef]
9. Crosby, M.; Pattanayak, P.; Verm, S. Blockchain technology: Beyond bitcoin. *Appl. Innov. Rev.* **2016**, *2*, 6–10.
10. Buterin, V. A Next-Generation Smart Contract and Decentralized Application Platform. 2014. Available online: <https://github.com/ethereum/wiki/wiki/White-Paper> (accessed on 3 June 2020).
11. Hasankhani, A.; Hakimi, S.M.; Shafie-khah, M.; Asadolahi, H. Blockchain technology in the future smart grids: A comprehensive review and frameworks. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106811. [CrossRef]
12. Yalcin, H.; Daim, T. Mining research and invention activity for innovation trends: Case of blockchain technology. *Scoentometrics* **2021**, *5*, 3775–3806. [CrossRef]
13. Hendershott, T.; Zhang, X.Q.; Zhao, J.L.; Zheng, Z.Q. FinTech as a Game Changer: Overview of Research Frontiers. *Inf. Syst. Res.* **2021**, *1*, 1–17. [CrossRef]
14. Law, J.; Bauin, S.; Courtial, J.-P.; Whittaker, J. Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Science* **1988**, *14*, 251–264. [CrossRef]
15. Pirolli, P.; Pitkow, J.; Rao, R. Sick from a sow’s ear: Extracting usable structures from the Web. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 13–18 April 1996.
16. Yang, Y.; Cui, L. Biclustering in co-citation analysis. *Libr. J. China* **2018**, *5*, 67–73.
17. Ellegaard, O.; Wallin, J.A. The bibliometric analysis of scholarly production: How great is the impact? *Science* **2015**, *105*, 1809–1831. [CrossRef] [PubMed]

18. Li, F.; Li, M.; Guan, P.; Ma, S.; Cui, L. Mapping Publication Trends and Identifying Hot Spots of Research on Internet Health Information Seeking Behavior: A Quantitative and Co-Word Biclustering Analysis. *J. Med. Internet Res.* **2015**, *17*, e81. [CrossRef] [PubMed]
19. Paule-Vianez, J.; Gómez-Martínez, R.; Prado-Román, C. A bibliometric analysis of behavioural finance with mapping analysis tools. *Eur. Res. Manag. Bus. Econ.* **2020**, *26*, 71–77. [CrossRef]
20. Topalli, M.; Ivanaj, S. Mapping the evolution of the impact of economic transition on Central and Eastern European enterprises: A co-word analysis. *J. World Bus.* **2016**, *51*, 744–759. [CrossRef]
21. Zhang, Z.; Murtagh, F.; Van Poucke, S.; Lin, S.; Lan, P. Hierarchical cluster analysis in clinical research with heterogeneous study population: Highlighting its visualization with R. *Ann. Transl. Med.* **2017**, *4*, 75. [CrossRef]
22. Zhu, X.; Qiu, J.; Xie, M.; Wang, J. A multi-objective biclustering algorithm based on fuzzy mathematics. *Neurocomputing* **2017**, *253*, 177–182. [CrossRef]
23. Liu, J.; Wei, W.-J.; Shi, B.; Guan, X.; Ma, J.-Y.; Wang, Y.-C. Mapping theme trends and knowledge structures for human neural stem cells: A quantitative and co-word biclustering analysis for the 2013–2018 period. *Neural Regen. Res.* **2019**, *14*, 1823–1832. [CrossRef]
24. Rasmussen, M.; Karypis, G. gCLUTO—An Interactive Clustering, Visualization, and Analysis System. 2004. Available online: <https://conservancy.umn.edu/handle/11299/215615> (accessed on 6 July 2021).
25. Zhao, Y.; Karypis, G. Criterion functions for document clustering: Experiments and analysis. *Mach. Learn.* **2004**, *55*, 311–331. [CrossRef]
26. Karypis, G.; Han, E. *Concept Indexing: A Fast Dimensionality Reduction Algorithm with Applications to Document Retrieval and Categorization*; Defense Technical Information Center (DTIC): Fort Belvoir, VA, USA, 2000.
27. Lu, K.; Yu, S.; Yu, M.; Sun, D.; Huang, Z.; Xing, H.; Bi, J.; Li, Z.; Li, Z.; Liu, X.; et al. Bibliometric Analysis of Tumor Immunotherapy Studies. *Med. Sci. Monit.* **2018**, *24*, 3405–3414. [CrossRef]
28. Cai, C.W. Disruption of financial intermediation by FinTech: A review on crowdfunding and blockchain. *Acc. Financ.* **2018**, *58*, 965–992. [CrossRef]
29. Donohue, J.C. *Understanding Scientific Literature: A Bibliographic Approach*; The MIT Press: Cambridge, MA, USA, 1973.
30. Liu, Y.; Cui, L. Research hotspots in current foreign and domestic medical decision support system. *Chin. J. Med. Libr. Inf. Sci.* **2016**, *11*, 49–55.
31. Bauin, S.; Michelet, B.; Schweighoffer, M.G.; Vermeulin, P. Using bibliometrics in strategic analysis: “understanding chemical reactions” at the CNRS. *Science* **1991**, *22*, 113–137. [CrossRef]
32. Yermack, D. Corporate Governance and Blockchains. *Rev. Financ.* **2017**, *1*, 7–31. [CrossRef]
33. Howell, S.T.; Marina, N.; David, Y. Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales. *Rev. Financ. Stud.* **2020**, *33*, 3925–3974. [CrossRef]
34. Raskin, M.; Yermack, D. Digital currencies, decentralized ledgers, and the future of central banking. In *Research Handbook on Central Banking*; Edward Elgar Publishing: Northampton, MA, USA, 2018.
35. Mehta, D.; Tanwar, S.; Bodkhe, U.; Shukla, A.; Kumar, N. Blockchain-based royalty contract transactions scheme for Industry 4.0 supply-chain management. *Inf. Process. Manag.* **2021**, *58*, 102586. [CrossRef]
36. Chen, F.; Wan, H.; Cai, H.; Cheng, G. Machine learning in/for blockchain: Future and challenges. *Can. J. Stat.* **2021**. [CrossRef]
37. Masilela, J.J.; van Wyk, R.B.; Marwa, N. Assessing the variability of crypto collateral assets in secured lending on the blockchain. *Dev. S. Afr.* **2021**, 1–11. [CrossRef]
38. Zheng, Z.B.; Xie, S.A.; Dai, H.N.; Chen, X.P.; Wang, H.M. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. In Proceedings of the 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, HI, USA, 25–30 June 2017; pp. 557–564. [CrossRef]
39. Li, X.; Wang, C.A. The technology and economic determinants of cryptocurrency exchange rates: The case of bitcoin. *Decis. Support Syst.* **2016**, *95*, 49–60. [CrossRef]
40. Zhu, H.; Zhou, Z.Z. Analysis and outlook of applications of blockchain technology to equity crowdfunding in China. *Financ. Innov.* **2016**, *2*, 29. [CrossRef]
41. Zhao, J.L.; Fan, S.; Yan, J. Overview of business innovations and research opportunities in blockchain and introduction to the special issue. *Financ. Innov.* **2016**, *1*, 28. [CrossRef]
42. Sikorski, J.J.; Haughton, J.; Kraft, M. Blockchain technology in the chemical industry: Machine-to-machine electricity market. *Appl. Energy* **2017**, *195*, 234–246. [CrossRef]
43. Beck, R.; Müller-Bloch, C.; King, J.L. Governance in the Blockchain Economy: A Framework and Research Agenda. *J. Assoc. Inf. Syst.* **2018**, *19*, 1020–1034. [CrossRef]
44. Cai, Y.; Zhu, D. Fraud detections for online businesses: A perspective from blockchain technology. *Financ. Innov.* **2016**, *2*, 20. [CrossRef]
45. Pieiro-Chousa, J.; López-Cabarcos, M.N.; Ribeiro-Soriano, D. The influence of financial features and country characteristics on B2B ICOs’ website traffic. *Int. J. Inf. Manag.* **2021**, *59*, 102332. [CrossRef]
46. Li, P.; Xu, H.; Ma, T. An efficient identity tracing scheme for blockchain-based systems. *Inf. Sci.* **2021**, *561*, 130–140. [CrossRef]
47. Chen, M.; Liu, Q.; Huang, S.; Dang, C. Environmental cost control system of manufacturing enterprises using artificial intelligence based on value chain of circular economy. *Enterp. Inf. Syst.* **2020**. [CrossRef]

-
48. Shen, C.-W.; Min, C.; Wang, C.-C. Analyzing the trend of O2O commerce by bilingual text mining on social media. *Comput. Human Behav.* **2019**, *101*, 474–483. [[CrossRef](#)]
 49. Liu, Y.; Chen, M. Applying Text Similarity Algorithm to Analyze the Triangular Citation Behavior of Scientists. *Appl. Soft Comput.* **2021**, *107*, 107362. [[CrossRef](#)]