# Identification of Breakthrough Innovation Topics from a Dynamic Topic Network Perspective: A Case Study of the Blockchain Domain (Postprint)

**Authors:** Chen Hongshu, Song Yahui, Jin Qianqian, Wang Xuefeng

**Date:** 2023-04-01T00:00:00+00:00

## Abstract

[目的/意义] Breakthrough innovation plays a critical role in scientific and technological development. In the big data environment, the inherent characteristics of scientific and technological development—such as complexity, multidimensionality, and continuous evolution—are becoming increasingly prominent. Identifying breakthrough innovation topics from a dynamic perspective holds significant importance for delineating breakthrough innovation domains, rationally allocating innovation resources, and providing solutions for innovation upgrading for nations, enterprises, and universities. [方法/过程] This study comprehensively employs methods including topic models, word embedding algorithms, and complex network analysis to construct dynamic topic networks, fully considering the structural characteristics of topics within time windows and their evolutionary states across time windows, and based on this foundation, identifies breakthrough innovation topics by incorporating the novelty, disruptiveness, impact, and interdisciplinary characteristics of breakthrough innovation. [结果/结论] An empirical study was conducted in the blockchain domain, identifying that the topics of Neural Network and Edge Computing exhibit the most significant breakthrough innovation characteristics. By combining existing blockchain research with the Critical and Emerging Technologies list released by the U.S. National Science and Technology Council, the feasibility and effectiveness of the proposed method were validated. However, quantitative validation of the results and breakthrough innovation topic identification integrating multi-source data require further research.

## Full Text

# Dynamic Topic Network Perspective for Radical Innovation Topic Identification: A Case Study of the Blockchain Field

*Chen Hongshu, Song Yahui, Jin Qianqian, Wang Xuefeng*

School of Management and Economics, Beijing Institute of Technology, Beijing 100081

**Abstract:** Radical innovation plays a critical role in scientific and technological development. In the big data environment, the complex, multidimensional, and continuously evolving characteristics of science and technology development are becoming increasingly prominent. Identifying radical innovation topics from a dynamic perspective is of great significance for countries, enterprises, and universities to analyze radical innovation fields in detail, allocate innovation resources rationally, and provide solutions for innovation upgrades. This paper integrates topic modeling, word embedding algorithms, and complex network analysis to construct dynamic topic networks, comprehensively considering both the structural characteristics of topics within time windows and their evolutionary states across time windows. Based on this foundation, combined with the novelty, mutability, impact, and interdisciplinary characteristics of radical innovation, it identifies radical innovation topics. An empirical study in the blockchain field identifies Neural Network and Edge Computing as the two topics with the most significant radical innovation characteristics. By combining existing blockchain research with the list of critical and emerging technologies issued by the U.S. National Science and Technology Council, the feasibility and effectiveness of the proposed method are verified. However, quantitative verification of the results and radical innovation topic identification that integrates multi-source data require further research.

**Keywords:** radical innovation; topic network; topic identification; LDA; Word2vec model; blockchain

**Classification Number:** G250.2

**DOI:** 10.13266/j.issn.0252-3116.2022.10.004

---

## Introduction

The world is poised for a new round of scientific and technological revolution, and China has entered a critical period of its "14th Five-Year Plan" development. General Secretary Xi Jinping has repeatedly emphasized that "innovation is an important force driving a country and a nation forward." Radical Innovation, as a highly revolutionary innovative activity, represents a key element for enterprises to transform industrial chains and enhance competitiveness, as well as an important guarantee for seizing opportunities in increasingly fierce international competition in the new era [1-3]. Against the backdrop of "improving the efficiency of the innovation system," timely and accurate identification of radical innovation is a critical link in providing decision-making support for national policy formulation, corporate strategic layout, and academic research planning, and has become an important research issue of common concern to both academia and industry [1, 6-8].

Specifically, previous studies on radical innovation identification have employed citation analysis and co-word analysis, constructing relevant indicators from perspectives such as citation counts, citation novelty, citation keywords, co-word networks, and word frequency changes [9-10]. However, citation analysis suffers from temporal lag issues, while co-word analysis has limitations in exploring textual semantics and feature expression. To address these problems, some scholars have combined text mining and network analysis methods for radical innovation identification [11-12]. Nevertheless, there remains a lack of systematic work on fully considering the dynamics of technological evolution from a network perspective and comprehensively measuring the multiple characteristics of radical innovation. Based on existing research, this paper focuses on two research questions: (1) How to reasonably extract and vectorize topics to construct dynamic topic networks that reflect the evolution process and state of topics in target fields? (2) How to measure the multiple characteristics of radical innovation based on question (1) to more systematically identify radical innovation topics within fields?

The concept of radical innovation is based on Schumpeter' s "creative destruction" [4]. W. J. Abernathy et al. defined it as innovation that uses technological innovation to enhance corporate status and reshape market patterns, laying the foundation for subsequent radical innovation research [5]. As a non-incremental innovation activity, radical innovation possesses multiple characteristics including mutability, novelty, and interdisciplinary nature. Currently, a large number of studies use bibliometrics, text mining, and network analysis methods to conduct research on radical innovation identification and have achieved certain results [1, 6-8]. However, existing review studies point out that current radical innovation identification methods still have limitations in considering topic evolution [13].

Topic extraction is one of the key foundations for radical innovation topic identification [13]. The quality of technical topic extraction affects the identification of various topic properties. The transition from macro-level statistical research to specific and in-depth knowledge discovery is the research trend of scientometrics and scientific text mining methods revealing disciplinary knowledge structures in the big data context. Among existing topic extraction studies, keyword-based methods provide the most detailed expression of technical concepts but often require multi-layered, large-scale, supervised screening, where screening principles and clustering granularity directly affect the generation of technical topics, posing more challenges for subsequent topic semantic expression [14]. The Latent Dirichlet Allocation (LDA) topic model, which can deeply mine implicit semantics in large texts, has attracted widespread attention in fields such as topic identification [15-16], technology forecasting [17], and scientific mapping [18] in recent years. However, existing research has not yet reached a consensus on how to reasonably preset the number of topics. At the same time, existing review studies indicate that current radical innovation identification methods still have limitations in considering topic evolution [13].

To reveal the generation, development, evolution, and demise of topics in the process of scientific and technological innovation, and thereby better identify radical innovation topics, it is necessary to reasonably calculate topic similarity within single or multiple time windows. However, topic models such as LDA have inherent problems in measuring the "distance" between topics and calculating topic similarity, lacking systematic connection between topic identification and subsequent technology evolution and feature analysis [19]. Word embedding algorithms can discover potential semantics in large-scale text data while considering content context [20]. In recent years, due to their excellent ability to map words to vector space, they have attracted widespread attention. Word vectors can be used to replace traditional word representations in scientific text mining, bringing a new perspective to topic extraction and topic similarity calculation [21-22].

Based on the above background, facing the new challenges of topic extraction, relationship representation, and indicator system construction for radical innovation identification in the big data environment, this paper integrates topic models, word embedding, and complex network analysis methods to construct dynamic topic networks that simultaneously reveal topic evolution processes. Based on this foundation, combined with multiple characteristics of radical innovation, it identifies radical innovation topics.

---

## 1 Research Status

### 1.1 Connotation and Characteristics of Radical Innovation

To date, numerous scholars have defined and studied radical innovation from different perspectives and aspects, with research dimensions mainly including micro and macro levels. The micro level focuses on the breakthrough brought by the technology itself, believing that radical innovation differs from the minor changes and adjustments of existing technologies in incremental innovation, but integrates new disciplinary knowledge, is based on different scientific and technological principles, combines scientific frontiers, and creates revolutionary technological changes [2, 23-24]. The macro level defines it from the substantial impact generated by innovation activities [25], mainly including two aspects: its impact on market or industry patterns [24, 26], and its academic influence in scientific research [27].

Existing studies have different entry points and emphases in summarizing the connotation of radical innovation. Through systematic review, this paper comprehensively summarizes the characteristics of radical innovation in current research, including frontier nature, mutability, high impact, interdisciplinary nature, discontinuity and nonlinearity, long-term nature, uncertainty and unpredictability, divergence, and randomness and contingency. The specific characteristic explanations are shown in Table 1 . Although existing research has summarized multiple characteristics of radical innovation, some characteristics such as uncertainty, divergence, and contingency are difficult to quantify directly. In existing quantitative research, novelty [10, 28], interdisciplinary nature [29-31], mutability [32-33], and impact [34] are the main features used for radical innovation topic identification. Therefore, drawing on the understanding of radical innovation connotation in existing research and experience in selecting main features, this paper builds a hierarchical indicator system based on the most commonly used features in mainstream research, namely novelty, mutability (i.e., major breakthrough), high impact, and interdisciplinary nature, to conduct quantitative research.

### 1.2 Topic Extraction and Evolution Analysis

Topic extraction, as a specific application of text mining, is currently mainly based on keyword (subject term) clustering, SAO semantic structure identification [45], and probabilistic topic models. Overall, these three methods have their own advantages and disadvantages in extracting and expressing core technical content: (1) Traditional keyword-based methods provide the most detailed expression of technical concepts but have limited semantic expression, and require multi-layered, large-scale, supervised screening, where screening principles and clustering granularity directly affect the generation of technical topics [14]; (2) Compared with keywords, SAO semantic structures can identify context and improve semantic expression, but in big data environments, methods centered on SAO structures have difficulties in dimensionality reduction [46]; (3) Topic models represented by LDA can mine implicit semantics in large texts, and expressing concepts in the form of word distributions (which can be regarded as word clusters) can avoid ambiguity caused by synonyms, thus being widely used in topic extraction research in the past decade. However, since the total number of topics in LDA needs to be set in advance as a parameter [47], and too large or too small values will affect the accuracy and readability of topic capture and expression, although existing research has formed solutions to determine the number of topics through perplexity [48], in actual research, scholars often need to continue evaluating or screening topics to balance readability [48].

After topic extraction, revealing the evolution process, patterns, and trends of scientific research topics is of great significance for grasping field development trends and detecting radical innovation topics. More than 20 years ago, R. Watts and A. Porter [49] proposed exploring the evolution of technical topics by statistically tracking keyword changes, which, although not considering deeper semantic relationships between word pairs, laid the foundation for subsequent topic evolution analysis. To address the issue that keywords cannot reveal the association relationships between technical topics, evolution analysis methods based on citations use mutual citation information between measurement objects to detect technical topics and trends in fields [50]. However, citation-centered analysis methods cannot truly delve into the semantic content from the perspective of content semantics. In recent years, topic evolution research based on scientific text mining has attracted increasing attention from researchers.

Text mining and network analysis methods have also received widespread attention in the field of radical innovation topic identification in recent years. Researchers use natural language processing technology to mine and analyze keywords or topics in scientific literature and patent literature, while relying on citation networks to leverage the advantages of complex network theory and methods to achieve identification and detection of radical innovation topics [53]. For example, J. Yoon et al. identified outlier patents based on SAO-calculated patent text similarity to characterize breakthrough technological innovation [11]. N. Shibata et al. treated paper citation networks as complex networks, compared and analyzed the gallium nitride (GaN) and complex network fields, and identified incremental and bifurcation innovations (a type of radical innovation) in the two fields through node intra-module degree and participation coefficient based on citation network clustering [12].

Overall, although bibliometric methods are efficient and direct, they have problems such as citation lag, failure to delve into semantic levels, and remaining at the descriptive research level of static bibliometric characteristics of topics. Text mining has certain advantages in exploring implicit semantic relationships, but both keyword-centered and topic model-based topic extraction methods have inherent problems in calculating topic similarity, making it difficult to directly measure technology evolution and conduct evolution-related feature analysis. Finally, the introduction of complex network-related indicators has expanded a new perspective of topological property measurement for radical innovation identification research. Whether citation networks or semantic networks, they can better identify radical innovation by showing knowledge structures.

Based on existing research, this paper integrates text mining and network analysis perspectives, incorporates topic evolution into the methodological system, considers time factors to show dynamic changes of topics, and comprehensively identifies topics based on multiple characteristics of radical innovation on the basis of in-depth mining of knowledge structure characteristics.

---

## 2 Research Methods

This paper simultaneously employs topic models, word embedding, and complex network analysis methods. Based on topic evolution and knowledge structure changes in dynamic topic networks, it constructs hierarchical indicator systems for measuring topic "novelty," "mutability," "impact," and "interdisciplinary nature" to identify radical innovation topics. The overall research framework is shown in Figure 1 [Figure 1: see original paper]. Specifically, this paper uses scientific paper data from multiple time windows as the data source. After data preprocessing, it first extracts topic sets in different time windows through topic models and uses Word2vec to map them into a unified vector space, generating topic vector matrices for different time windows. Then, from a network perspective, this paper defines the evolution states of topics in different time windows to reflect the processes of topics being newly born, evolving, merging, and dying out over time. Finally, based on analyzing the structural characteristic changes of dynamic topic networks and knowledge flow, this paper constructs hierarchical indicator systems for "novelty," "mutability," "impact," and "interdisciplinary nature" of radical innovation to identify radical innovation topics.

### 2.1 Topic Extraction and Vectorization

**2.1.1 LDA-Based Topic Extraction** LDA is currently one of the most widely used topic models [47]. It extracts the distribution of texts in topic space through a random generation process and expresses topic concepts in the form of word distributions. Generally, the generation process of LDA can be represented by the joint distribution of random variables [54], as shown in Equation (1):

$$p(\vec{w}_d, \vec{z}_d, \vec{\theta}_d, \Phi|\vec{\alpha}, \vec{\beta}) = \prod_{n=1}^{N_d} p(w_{d,n}|\vec{\phi}_{z_{d,n}})p(z_{d,n}|\vec{\theta}_d)p(\vec{\theta}_d|\vec{\alpha})p(\Phi|\vec{\beta})$$

where $\vec{Z}_d$ is the topic distribution of document $d$, $\vec{\theta}_d$ is the corresponding topic proportion, $Z_{d,n}$ represents the topic distribution of the $n$-th word in the $d$-th document, $\vec{\phi}_{1:K}$ represents topics, each $\vec{\phi}_k$ is a word distribution, with a total of $K$ topics, and $\alpha$ and $\beta$ are two hyperparameters, usually set to default values [55]. This paper divides the dataset by year $T$, trains LDA models on multiple time windows respectively, obtains $T$ document-topic probability distribution matrices and topic-word probability distribution matrices. After topic extraction, all extracted topics are named using word distribution probability ranking and manual verification to generate topic labels, laying the foundation for subsequent evolution state evaluation.

Extracting topics through the LDA topic model requires presetting the number of topics $K$ in advance. Based on existing research, this paper adopts a method combining perplexity [48] and manual parsing complexity [22] to set the $K$ value, as shown in Equation (2), where $Perplexity(D)$ represents model perplexity, calculated as shown in Equation (3); $Complexity$ represents the parsing complexity of model results, calculated as shown in Equation (4). A smaller perplexity value indicates better model fit to data, while smaller manual parsing complexity indicates relatively lower complexity in parsing topic content.

This paper comprehensively considers model effectiveness and manual parsing complexity to determine the optimal number of topics.

$$\arg\min_{K} f(K) = \frac{perplexity(K) - \min perplexity(K)}{\max perplexity(K) - \min perplexity(K)} + \frac{complexity(K) - \min complexity(K)}{\max complexity(K) - \min complexity(K)}$$

$$Perplexity(D) = \exp\left(-\frac{\sum_{d=1}^{D} \log(p(w))}{K - \min(K)}\right)$$

$$Complexity = \exp\left(\frac{\sum_{d=1}^{D} \log(p(w))}{K - \min(K)}\right)$$

In the above perplexity calculation formula, $\sum \log(p(w))$ represents the likelihood of the given training model on the test corpus.

**2.1.2 Word2vec-Based Topic Vectorization** To more effectively calculate topic similarity and construct dynamic topic networks, this paper uses the Word2vec algorithm to vectorize topics. As an efficient word embedding technology, Word2vec can capture contextual information of words in texts and transform words into low-dimensional dense real-valued word vectors containing semantic relationships [20]. The specific implementation models of Word2vec include Continuous Bag-of-words and skip-gram models. According to existing research, there is no significant difference in training effectiveness between the two [56]. This paper uses the skip-gram model to map vocabulary to vectors, combines topic discovery results to achieve semantic information extraction and vectorization of text data, and lays a solid foundation for constructing topic networks and indicator identification models.

Specifically, given a document collection sequence $D$ containing $N$ words and $N'$ non-repeating words, the skip-gram model is used to train the text set, generating a word vector set $V$ with dimension $\gamma$. Since each extracted topic is specifically represented as the $n$ non-repeating words with the highest probability on $\vec{\phi}_k$, using the probability of each word as weight, the $\gamma$-dimensional word vectors of the non-repeating words belonging to the topic are weighted and averaged to generate the vector of the topic in the unified vector space $v(T_{t,i})$, calculated as shown in Equation (5):

$$v(T_{t,i}) = \sum_{j=1}^{s} P(term_{t,i,j}) \cdot v(term_{t,i,j})$$

where $v(T_{t,i})$ is the vector representation of the $i$-th topic in time window $t$, $s$ is the number of top-ranked words under the topic, $P(term_{t,i,j})$ represents the probability value of the word, and $v(term_{t,i,j})$ represents the word vector corresponding to the word.

### 2.2 Dynamic Topic Network Construction and Topic Evolution Analysis

**2.2.1 Topic Network Construction** Based on the previously generated topic vectors, this paper uses cosine similarity to calculate the semantic correlation between topics within the same time window and retains edges based on the mean of inter-topic similarity to construct topic networks for each time window, as specifically calculated in Equation (6):

$$Similarity(T_{t,i}, T_{t,j}) = \cos(v(T_{t,i}), v(T_{t,j})) = \frac{v(T_{t,i}) \cdot v(T_{t,j})}{\|v(T_{t,i})\| \cdot \|v(T_{t,j})\|}$$

where $T_{t,i}, T_{t,j}$ represent topics, $v(T_{t,i}), v(T_{t,j})$ represent the vector forms of topics $T_{t,i}, T_{t,j}$, and $Similarity(T_{t,i}, T_{t,j})$ represents the similarity between topics $v(T_{t,i}), v(T_{t,j})$, with values between 0 and 1. The nodes of this network are topics extracted by LDA, and each topic is represented by a word cluster with probability distribution.

**2.2.2 Topic Evolution State Definition** Overall, identifying radical innovation topics requires understanding the knowledge state before and after their emergence [13], and topic evolution analysis can reveal the macro processes of topics being newly born, merging, evolving, and dying out in scientific and technological innovation, thus providing a dynamic perspective for radical innovation topic identification. Scientific literature published in a certain field over a period of time can be regarded as a dynamic dataset that develops over time. In this dataset, topic content evolution relationships are usually manifested as whether topics in a field have appeared, when they appeared, which other topics they are associated with, and how relationships develop—that is, whether they are newly emerged, merged with other topics, or have disappeared. Based on the research of Y. Zhang et al. [57], this paper sets the evolution states of topics over time windows into five categories: new birth, identical, derivative, fusion, and demise. The specific definitions of each state are as follows:

1. **New-born topic**: A newly emerged topic with no predecessor topics, having only low or zero correlation with topics in previous time windows.
2. **Identical topic**: Existing topics and subsequent topics have extremely high correlation, with similarity reaching above the threshold, and are considered the same topic.
3. **Derivative topic**: A new topic derived from an existing topic, having high correlation with the current topic but not being very similar, not belonging to the same topic, and may have a one-to-many relationship.
4. **Fusion topic**: A fusion topic has certain correlation with multiple predecessor topics, is the result of the fusion of multiple topics, but is not very similar to each topic and does not belong to the same topic.
5. **Demise topic**: If subsequent topics generated in the time window have no correlation or extremely low correlation with existing topics, the existing

topic can be considered a demise topic.

The schematic diagram of the above five topic evolution states is shown in Figure 2 [Figure 2: see original paper].

**2.2.3 Evolution State Measurement** Reasonable calculation of topic similarity is the foundation for identifying changes in topic evolution states. To track the dynamic changes and development of topics across different time windows, it is necessary to capture the topic sets corresponding to each time window of the dynamic dataset and calculate the similarity between topic sets in adjacent windows, thereby understanding the "origin" and "destination" of each topic. Therefore, after identifying topics and their corresponding topic vectors in each time window, this paper adopts cosine similarity to calculate the semantic relevance between topics in adjacent time windows, as shown in Equation (7):

$$Similarity(T_{t,i}, T_{t+1,j}) = \cos(v(T_{t,i}), v(T_{t+1,j})) = \frac{v(T_{t,i}) \cdot v(T_{t+1,j})}{\|v(T_{t,i})\| \cdot \|v(T_{t+1,j})\|}$$

where $T_{t,i}$ represents a topic at time $t$, $T_{t+1,j}$ represents a topic at time $t+1$, $v(T_{t,i}), v(T_{t+1,j})$ represent the vectors of topics $T_{t,i}, T_{t+1,j}$ respectively, and $Similarity(T_{t,i}, T_{t+1,j})$ represents the similarity between topics $v(T_{t,i}), v(T_{t+1,j})$, with values between 0 and 1.

This paper combines semantic relevance and topic labels generated in Section 2.1.1 to quantitatively measure topic evolution states. As shown in Figure 3 [Figure 3: see original paper], it first calculates the upper quartile (Q1) and median (Q2) of the topic similarity matrix (hereinafter referred to as "similarity matrix" ) between every two adjacent time windows as critical points for topic states: when inter-topic similarity reaches above Q1 and topic labels are the same, the topics in different time windows are considered "identical" topics; if the Q1 threshold condition is met but the topic label requirement is not, there is a strong correlation between the two, considered as derivative or fusion states; if the similarity value is between Q1 and Q2, it is also considered a derivative or fusion state; if it is less than Q2, it is considered a new-born or demise state.

Through quantitative evaluation of evolution states, this paper connects topic networks on adjacent time windows according to evolution relationships, that is, uses topic evolution relationships to link multiple topic networks across time windows, forming a dynamic topic network $G$. $G$ can be expressed as: $G = (G_1, G_2, \ldots, G_t, \ldots, G_T)$, where $G_t$ is the topic network within time period $t$, and $T$ is the number of divided time windows.

## 2.3 Indicator Construction

This paper constructs a hierarchical indicator system for measuring topic "novelty,""mutability,""impact,"and "interdisciplinary nature"by combing through the connotation and characteristics of radical innovation. Based on the established dynamic topic network, it comprehensively considers the structural characteristics of topics within time windows and evolution characteristics across time windows to identify radical innovation topics. The specific hierarchical indicator system is shown in Table 2 . This paper does not consider demise topics in the evolution process when identifying radical innovation topics.

**Table 2 Hierarchical Indicator System**

| Indicator | Connotation or Representation |
|---|---|
| **Novelty** | Topic word discrete age difference: The discrete difference between topic age and average age of all words; Topic word weighted age: The weighted average age of words contained in the topic, i.e., the age difference from the earliest appearance year of words in the entire dataset to the current moment, with weights being word distribution probability values |
| **Mutability** | The degree of change in topic content between adjacent moments; the greater the change, the higher the topic mutation degree |
| **Impact** | Degree centrality: The number of edges or sum of edge values associated with the node, indicating the degree of connection between the node and other nodes in the network; Closeness centrality: The degree to which most direct paths connecting a node to other nodes in the network are short (rather than long), reflecting the important position of the node in the network; Betweenness centrality: The proportion of shortest paths passing through a node in the network, reflecting the node' s influence in the network; Structural holes: Non-redundant connections between two nodes; structural holes can provide opportunities for their occupants to obtain "information benefits" and "control benefits," thus being more advantageous than other network members |
| **Interdisciplinary Nature** | Subject diversity: Different disciplinary intersections often breed new scientific frontiers and major scientific breakthroughs; the more subject types a topic contains, the higher its interdisciplinary degree |

**2.3.1 Novelty Measurement Based on Word Age** When measuring the novelty of technical topics, this paper focuses on the age of words in the topic —the later a word appears, the younger its age, and the higher its novelty. Specifically, combining the established dynamic topic network, it uses two indicators, word discrete age difference [58] and word weighted age, for quantitative judgment of topic novelty, and uses the entropy weight method to fit them to obtain a comprehensive novelty value ($TopicNovelty$). The specific calculation of word discrete age difference $TopicDiscreteAgeDifference(TDA)$ [58] is shown in Equation (8):

$$TDA_i = \sum_{i=n} Y_i - \sum_{j=n} Y$$

where, since topic $i$ changes dynamically across multiple time windows, $T_i$ represents the total number of years from its birth to the last time window, $N$ is the total number of words in the collection, $n$ represents the number of top-ranked words under topic $i$, and $Y_j$ is the earliest year the word appears in the entire dataset. $TDA$ has positive and negative values; the larger the value, the later the main words in the topic appear, and the higher the topic novelty.

Word weighted age $TopicWeightedAge(TWA)$ measures the age of main words in the topic based on weighted sum of word probabilities, calculating the age difference of main vocabulary in the topic from the earliest appearance year to the current moment, as specifically calculated in Equation (9):

$$TWA_i = \sum_{j=1}^{n} P(w_n) Y_j - 1$$

where $T_i$ represents the total number of time windows experienced by topic $i$ in its evolution, $n$ is the number of top-ranked words under topic $i$, $P(w_n)$ is the probability value of the corresponding word, and $Y_j$ is the age of word $j$, i.e., the time difference from the earliest appearance year of the word in the entire dataset to the current moment. The larger the $TWA$ value, the higher the topic novelty.

**2.3.2 Mutability Measurement Based on Topic Similarity** Based on the previously calculated topic vectors, this section sets the topic mutability indicator ($TopicMutation$) based on catastrophe theory. The higher the topic mutability degree, the greater the topic change, and vice versa, as specifically calculated in Equation (10):

$$TM_i = \sum_{t=1}^{T_i} \left(1 - \frac{v(T_{t,i}) \cdot v(T_{t+1,i})}{\|v(T_{t,i})\| \cdot \|v(T_{t+1,i})\|}\right)$$

where $T_i$ still represents the total number of time windows experienced by topic $i$ in its evolution, $v(T_{t,i}), v(T_{t+1,i})$ represent the topic vectors of topic $i$ in adjacent time windows. The larger this value, the greater the degree of topic change.

**2.3.3 Impact Measurement Based on Network Indicators** In network analysis, network centrality is often used to measure the influence of nodes in the network. Related indicators include closeness centrality, betweenness centrality, and degree centrality [59], while structural holes are often used to measure the key position of nodes. This paper selects centrality and structural holes to measure topic impact respectively, and uses the entropy weight method to fit them to obtain a comprehensive topic impact value ($TopicInfluence$). This section calculates indicators based on the topic network constructed in Section 2.2.1, as follows:

1. **Degree centrality** can be calculated by the ratio of the number of edges connected to topic node $i$ in the topic network to the maximum possible edges, as shown in Equation (11) [60]:

$$C_D(i) = \frac{k_i}{N-1}$$

   where $T_i$ represents the total number of time windows experienced by topic $i$ in its evolution, $k_i$ represents the number of topics connected to topic $i$ in time window $t$. The larger this value, the higher the topic' s impact.

2. **Closeness centrality** can be calculated by the average length of the shortest paths from topic node $i$ to all other topic nodes in the topic network, as shown in Equation (12) [60]:

$$CC(i) = \frac{N-1}{\sum_{j\neq i} d_{ij}}$$

   where $d_{ij}$ represents the shortest distance from topic $i$ to topic $j$, $T_i$ represents the total number of time windows experienced by topic $i$. The larger this value, the more central the node' s position in the network, indicating higher topic impact.

3. **Betweenness centrality** represents the proportion of shortest paths passing through a topic node in the network, as shown in Equation (13) [60]:

$$CB(i) = \sum_{t=1}^{T_i} \sum_{j\neq i\neq k\in V, j<k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

   where $T_i$ represents the total number of time windows experienced by topic $i$ in its evolution, $\sigma_{jk}(i)$ is the number of shortest paths between nodes $j$ and $k$ that pass through node $i$ in time window $t$, and $\sigma_{jk}$ is the total number of shortest paths between nodes $j$ and $k$. This indicator reflects the node' s influence in the network; the larger the value, the higher the topic' s impact.

4. In structural hole-related research, the network constraint coefficient is usually used to calculate the positional advantages occupied by each node, depicting the closeness of a node' s direct or indirect connections with

other nodes. The smaller the value, the more structural holes, the more important the position, and the more capable the node is of obtaining diverse knowledge, making it a potential innovation node [61]. This paper uses UCINET software to calculate the structural hole constraint coefficient of topic $i$ in the topic network for each time window during its evolution, and uses the total number of time windows experienced by $i$ in its evolution for weighted averaging, finally obtaining the structural hole constraint coefficient of topic $i$ across multiple windows.

#### 2.3.4 Interdisciplinary Measurement Based on Subject Classification

The intersection of different disciplines is often the growth point of new science and new scientific frontiers, and is most likely to produce major scientific breakthroughs [62]. According to existing research, this section calculates the interdisciplinary degree of topics on dynamic networks based on the subject classification of Web of Science (Web of Science Category). Since each paper belongs to one or more subjects and simultaneously covers several topics, i.e., the contribution degree of different subjects under each topic varies, this paper proposes the subject diversity indicator ($TopicSubjectDiversity$) to characterize the interdisciplinary nature of topics. The larger this value, the more subject types the topic contains, and the higher its interdisciplinary degree, calculated as shown in Equation (14):

$$TST_i = \sum_{t=1}^{T_i} \frac{\sum_{m=1}^{h} P(d_{t,i,m}) S(d_{t,i,m})}{\sum_{m=1}^{h} P(d_{t,i,m})} / T_i$$

where $T_i$ represents the total number of time windows experienced by topic $i$ in its evolution, $h$ represents the number of documents in time window $t$, $P(d_{t,i,m})$ represents the probability that the $m$-th document belongs to topic $i$ in time window $t$, $S(d_{t,i,m})$ represents the number of subjects contained in the $m$-th document in time window $t$, and $S$ represents the total number of subject classifications in Web of Science.

---

## 3 Empirical Analysis

Blockchain, as one of the frontier fields of current information technology, continuously stimulates, empowers, and accelerates digital economic development with its foundational, leading, and innovative characteristics, forming an all-round, strategic impact on current information technology. The following section takes the blockchain field as an example, conducts comprehensive collection of relevant scientific paper data, and carries out radical innovation topic identification based on dynamic topic networks to verify the feasibility and effectiveness of the proposed method and related research work.

### 3.1 Data Acquisition

As a rapidly developing frontier field, there is no consensus on retrieval strategies for blockchain. Most existing studies use “blockchain” or “blockchain” or “bitcoin” as keywords for Chinese and English literature retrieval. This paper reviews blockchain-related literature, improves the retrieval strategy based on the strategy constructed by Shang Qi and Chen Hongmei [63], and obtains the following retrieval strategy:

$$TS = ("chainofblock*"OR"blockchain*"OR"blockchain*"OR"genesisblock*"OR"Bitcoin"OR"Ethereum"OR"$$

Using the above retrieval formula, this paper searches the SCI and SSCI databases of Web of Science for English journal and conference literature from 2011 to 2020, obtaining a total of 10,817 entries. To further improve data accuracy, a small number of entries related to chemistry, materials science, immunology, cytology, pharmacy, and other areas weakly correlated with blockchain core content are removed through manual intervention [63]. After screening, 9,805 blockchain-related data entries are retained, forming the initial corpus for radical innovation topic identification in this field. Figure 4 [Figure 4: see original paper] shows the annual number of papers in the blockchain field. It can be seen that the number of relevant literature was relatively small in the early stage with relatively flat growth, and began to increase significantly from 2016.

### 3.2 Dynamic Topic Network Construction

To retain only the text content most directly related to blockchain technology, this paper uses natural language processing to clean the title and abstract fields in the initial corpus, removing stopwords and common phrases. Then, taking year as the unit, it establishes 10 time windows from 2011 to 2020 and divides the text set according to time windows. For the text set under each window, this paper balances perplexity and manual parsing complexity, setting the number of topics $K$ for the 10 stages as: 7, 10, 12, 14, 15, 25, 25, 25, 30, and 25. Due to space limitations, this paper only shows the process of determining the total number of topics for the 2020 time window; the calculation logic is the same for other time windows, as shown in Figure 5 [Figure 5: see original paper].

Subsequently, this paper uses 2,000 Gibbs sampling iterations to infer latent variables and distributions to extract topics for each period, and uses word distribution probability ranking and manual verification to name topics in each network and generate topic labels. Simultaneously, it uses the Python Gensim toolkit to train word vectors on the entire corpus, with the dimension parameter $\gamma$ set to 150 and window size set to 5. According to the method given in Section 2.1.2, topics under each time window are transformed into 150-dimensional vectors in a unified vector space. This paper calculates topic correlation matrices using Equation (6) and constructs topic networks for each period, as shown in Figure 6 [Figure 6: see original paper] (only topic networks for time windows 1 and 10 are shown here due to space limitations).

To track the dynamic changes and development of topics across different time windows, this paper then calculates the similarity between topic sets in adjacent time windows to identify the "origin" and "destination" of each topic in the blockchain field. Based on the previously calculated topic vectors, this section uses cosine similarity to obtain 9 correlation matrices representing changes in topic sets between adjacent time windows, and quantitatively evaluates the evolution states of topics across windows by calculating the upper quartile (Q1) and median (Q2). Based on evolution state measurement results, a total of 87 non-repeating topics have emerged in the blockchain field over the past 10 years, covering all states of new birth, identical, derivative, fusion, and demise. Their evolution process is shown in Figure 7 [Figure 7: see original paper]. Among them, solid lines represent identical topics, dashed lines represent derivative or fusion states, and dark gray nodes represent demise topics. From Figure 7, it can be seen that the number of topics began to increase from 2016, more research topics evolved in 2019, most topics showed derivative or fusion states, and each topic is in dynamic change, with derivative or fusion evolution being the norm for knowledge flow in this field.

### 3.3 Radical Innovation Topic Identification

Based on the dynamic topic network and using the indicator calculation methods in Section 2.3, the "novelty," "mutability," "impact," and "interdisciplinary nature" indicator values for each topic can be obtained and standardized to map them uniformly to the [0,1] interval, as shown in Table 3 . According to the mean values of each indicator, the 25 existing topics in 2020 (see Table 4 ) are divided into 16 categories, representing the overall characteristics of each topic across the four-dimensional indicators. Among them, two topics belong to the category with high novelty, high mutability, high impact, and strong interdisciplinary nature, and are identified as radical innovation topics: Neural Network (2020-T6-NeuralNetwork) and Edge Computing (2020-T8-EdgeComputing).

**Table 3 Indicator Values of "Novelty," "Mutability," "Impact," and "Interdisciplinary Nature" for 25 Topics in 2020**

| Topic | Novelty | Mutability | Impact | Interdisciplinary |
|---|---|---|---|---|
| Distributed Ledger | 0.1518 | 0.9593 | 0.7012 | 0.7983 |
| Cloud Storage | 0.0442 | 0.5788 | 0.7164 | 0.8145 |
| Product Traceability | 0.1487 | 0.9365 | 0.4842 | 0.7883 |
| Byzantine Fault Tolerant | 0.0687 | 0.9001 | 0.5071 | 0.8077 |
| Mining | 0.1008 | 0.8805 | 0.4040 | 0.8052 |
| Communication | 0.0986 | 0.0714 | 0.6681 | 0.8678 |
| Process Management | 0.0590 | 0.7525 | 0.7830 | 0.8145 |

| Topic | Novelty | Mutability | Impact | Interdisciplinary |
|---|---|---|---|---|
| Voting | 0.0877 | 0.8516 | 0.7279 | 0.8317 |
| Query | 0.2297 | 0.5785 | 0.5905 | 0.8258 |
| Smart Contract | 0.0429 | 0.7102 | 0.5336 | 0.7955 |
| Energy | 0.0399 | 0.0557 | 0.2928 | 0.8064 |
| Healthcare | 0.1042 | 0.1231 | 0.3056 | 0.0000 |
| Token | 0.1974 | 0.6782 | 0.3515 | 0.8201 |
| Auction | 0.0841 | 0.9018 | 0.7354 | 0.8183 |
| Signature | 0.0685 | 0.2077 | 0.8503 | 0.8718 |
| Cryptocurrency | 0.0464 | 0.6961 | 0.5334 | 0.7978 |
| Software | 0.1656 | 0.6493 | 0.5674 | 0.7826 |
| Supply Chain | 0.3136 | 0.7739 | 0.1084 | 0.6064 |
| Digital Service | 0.0551 | 0.7912 | 0.5221 | 0.8431 |
| Neural Network | 0.5115 | 1.0000 | 0.6398 | 0.7990 |
| Edge Computing | 0.2734 | 0.7525 | 0.6066 | 0.6643 |
| Financial Market | 0.2691 | 0.8413 | 0.6789 | 0.8445 |
| Consensus Algorithm | 0.0475 | 0.6742 | 0.9501 | 0.8645 |
| … | … | … | … | … |

**Table 4 Labels, Main Content, and Radical Innovation Characteristic Measurement of 2020 Topics**

| Topic ID | Topic Label | Main Content | Novelty | Mutability | Impact | Interdisciplinary |
|---|---|---|---|---|---|---|
| 2020-T6 | Neural Network | ICO, neural_{network}, algorithm, neuron, computing_{power} | High | High | High | High |
| 2020-T8 | Edge Computing | edge, algorithm, edge_{computing}, mobile, AI | High | High | High | High |
| … | … | … | … | … | … | … |

Topic 6 "Neural Network" (2020-T6-NeuralNetwork) mainly covers Initial Coin Offering (ICO), neural networks, algorithms, computing power, etc., involving the fusion application of neural network algorithms and blockchain technology in the financial field. Existing blockchain-related research can verify the radical innovation attributes of this topic: blockchain + neural networks can effectively improve transaction security, identity authentication reliability, and solve information non-disclosure problems, providing technical guarantees and strong driving forces for real economy development and digital economy ecosystem realization. Its breakthrough progress has received widespread attention in recent years [64-66]. The U.S. National Science and Technology Council (NSTC) released a new version of the Critical and Emerging Technologies (CETs) list in February 2022 [67], incorporating distributed ledger technology (blockchain technology) into the financial technology category, fully reflecting its importance and the rapid development of the two fields.

Topic 8 "Edge Computing" (2020-T8-EdgeComputing) mainly involves edge computing, mobile edge computing, edge artificial intelligence computing, etc. Existing research shows that edge computing can provide resources for blockchain services, mainly including communication resources and computing resources [68], while blockchain technology is responsible for ensuring security and edge computing is responsible for improving communication efficiency. In 2020, China Mobile released the "Blockchain + Edge Computing Technology White Paper," pointing out that the integrated application of "blockchain + edge computing" as a new field of communication and information technology integration development can promote resource sharing, optimal allocation, cross-boundary collaboration and innovation, accelerate social informatization transformation, and has broad research prospects [69]. In addition, the CETs list issued by the U.S. National Science and Technology Council lists edge computing as a representative of advanced computing, which also verifies the effectiveness of the identification results to a certain extent. Overall, the "blockchain +" industry fusion model is developing rapidly, bringing profound changes to various fields while also marking that blockchain development has entered the 3.0 era [70], opening a new stage of development.

---

## 4 Conclusion and Outlook

Radical innovation is a key element for countries to grasp the initiative in industrial revolutions. Accurately identifying radical innovation topics can provide decision-making support for national policy formulation and corporate strategic layout, and point out directions for academia to focus on research priorities. Summarizing existing research, radical innovation topics need to be analyzed from a dynamic perspective, and revealing the dynamic evolution process, patterns, and trends of scientific research topics is of crucial significance for detecting radical innovation topics. This paper uses scientific paper data from multiple time windows as the data source, comprehensively uses probabilistic topic models and word embedding methods for topic extraction and vectorization, first overcoming the blind spots in semantic expression and difficulties in screening and dimensionality reduction of keyword-centered topic identification methods, and completing the mapping from scientific text to mathematical vectors. Subsequently, this paper constructs dynamic topic networks under continuous time windows, comprehensively considers the structural characteristics of topics within time windows and evolution characteristics across time windows, and builds hierarchical indicator systems for measuring topic "novelty," "mutability," "impact," and "interdisciplinary nature" to identify radical innovation topics. From a methodological perspective, research on radical innovation topic identification under dynamic topic networks is an important supplement to existing methods based on text mining and network analysis perspectives.

From a results perspective, this paper uses scientific literature data from the blockchain field from 2011 to 2020 to identify two topics with the most significant radical innovation characteristics, namely neural networks and edge computing, and verifies the effectiveness of the method by combining existing research in the field and technology lists. However, this paper has certain limitations and room for further research. First, from the perspective of dynamic topic networks, no quantitative result verification method has been constructed; second, this paper currently only considers scientific literature data, with a single data source, and needs to further expand data dimensions in future research and map more characteristics of radical innovation onto multi-source, heterogeneous dynamic topic networks; finally, the method of this paper has only been empirically analyzed in the blockchain field, and needs to be applied to other technical fields in the future to further verify the systematicity and reliability of the method.

---

## References

[1] Guo Xiaochao, Lei Jing, Feng Yinhu, et al. A review of international radical innovation theory research based on knowledge map [J]. Scientific Management Research, 2020, 38(1): 20-26.

[2] Shao Yunfei, Zhan Kun, Wu Yanbo. Breakthrough technological innovation: Theoretical review and research prospects [J]. Technology Economics, 2017, 36(4): 30-37.

[3] Hain D S, Christensen J L. Capital market penalties to radical and incremental innovation [J]. European Journal of Innovation Management, 2020, 23(2): 291-313.

[4] Wan Ning. Analysis of the relationship between disruptive innovation, destructive innovation, and radical innovation [J]. Business, 2015(30): 122-123.

[5] Abernathy W J, Utterback J M. Patterns of innovation in technology [J]. Technology Review, 1978, 80(7): 41-47.

[6] McDermott C M, O' Connor G C. Managing radical innovation: An overview of emergent strategy issues [J]. Journal of Product Innovation Management, 2002, 19(6): 424-438.

[7] Leifer R. Radical innovation: How mature companies can outsmart upstarts [M]. Brighton: Harvard Business Press, 2000.

[8] Zhang Jinzhu, Zhang Xiaolin. A review of breakthrough innovation identification research based on patent science citations [J]. Library and Information Service, 2020, 64(5): 100-113.

[9] Schoenmakers W, Duysters G. The technological origins of radical inventions [J]. Research Policy, 2010, 39(8): 1051-1059.

[10] Dahlin K B, Behrens D M. When is an invention really radical? Defining and measuring technological radicalness [J]. Research Policy, 2005, 34(5): 717-737.

[11] Yoon J, Kim K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks [J]. Technological Forecasting and Social Change, 2016, 105: 179-191.

[12] Shibata N, Kajikawa Y, Takeda Y, et al. Detecting emerging research fronts based on topological measures in citation networks of scientific publications [J]. Technological Forecasting and Social Change, 2008, 28(11): 758-775.

[13] Liu Yahui, Xu Haiyun. A review of early identification of radical innovation and weak signal analysis [J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(1): 97-110.

[14] Zhang Y, Zhang G, Chen H, et al. Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research [J]. Technological Forecasting and Social Change, 2017, 119: 39-52.

[15] Li Hui, Xuan Hongsheng. A multi-attribute fusion method for technology innovation topic mining from patents—A case study of chip patents [J]. Library and Information Service, 2020, 64(11): 96-107.

[16] Chen H, Wang X, Pan S, et al. Identify topic relations in scientific literature using topic modeling [J]. IEEE Transactions on Engineering Management, 2021, 68(5): 1232-1244.

[17] Chen H, Zhang G, Zhu D, et al. Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014 [J]. Technological Forecasting and Social Change, 2017, 119: 39-52.

[18] Suominen A, Toivanen H. Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification [J]. Journal of the Association for Information Science and Technology, 2016, 67(10): 2464-2476.

[19] Jungs, Yoon W C. An alternative topic model based on common interest authors for topic evolution analysis [J]. Journal of Informetrics, 2020, 14(3): 101040.

[20] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. New York: ACM, 2013: 3111-3119.

[21] Zhang Y, Lu J, Liu F, et al. Does deep learning help topic extraction? A kernel k-means clustering method with word embedding [J]. Journal of Informetrics, 2018, 12(4): 1099-1117.

[22] Chen H, Jin Q, Wang X, et al. Profiling academic-industrial collaborations in bibliometric-enhanced topic networks: A case study for digitalization research [J]. Technological Forecasting and Social Change, 2022, 175: 121402.

[23] Dewar R D, Dutton J E. The adoption of radical and incremental innovations: An empirical analysis [J]. Management Science, 1986, 32(11): 1422-1433.

[24] Li H, Zhang Q, Zheng Z. Research on enterprise radical innovation based on machine learning in big data background [J]. Journal of Supercomputing, 2020, 76(5): 3283-3297.

[25] Cao Yiwen, Xu Haiyun, Wu Huawei, et al. Breakthrough prediction of emerging technology topics based on citation curve fitting—A case study of stem cell field [J]. Library and Information Service, 2020, 64(5): 100-113.

[26] Zhou K Z, Yim C K, Tse D K. The effects of strategic orientations on technology- and market-based breakthrough innovations [J]. Journal of Marketing, 2005, 69(2): 42-60.

[27] Xu Haiyun, Liu Yahui, Luo Rui. A review of early identification research on breakthrough scientific innovation [J]. Information Studies: Theory & Application, 2021, 44(4): 198-205.

[28] Arts S, Veugelers R. The technological origins and novelty of breakthrough inventions [C]//35th DRUID Celebration Conference. Spain: Barcelona, 2013: 1-30.

[29] Zhou Lei, Yang Wei, Zhang Yufeng. Research on breakthrough innovation identification framework based on patent mining [J]. Information Studies: Theory & Application, 2016, 39(9): 73-76, 46.

[30] Huang Lucheng, Jiang Linshan, Wu Feifei. Research on identification of disruptive technologies in embryonic stage [J]. Science & Technology Progress and Policy, 2019, 36(1): 10-17.

[31] Ponomarev I, Lawton B K, Williams D E, et al. Breakthrough paper indicator 2.0: Can geographical diversity and interdisciplinarity improve the accuracy of outstanding paper prediction? [J]. Scientometrics, 2014, 100(3): 755-765.

[32] Kleinberg J. Bursty and hierarchical structure in streams [J]. Data Mining and Knowledge Discovery, 2003, 7(4): 373-397.

[33] Zhang Jinzhu, Zhang Xiaolin. Breakthrough innovation identification based on cited scientific knowledge mutation [J]. New Technology of Library and Information Service, 2016(Z1): 42-50.

[34] Chen C, Chen Y, Horowitz M, et al. Towards an explanatory and computational theory of scientific discovery [J]. Journal of Informetrics, 2009, 3(3):

191-209.

[35] Ahuja G, Lampert C M. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions [J]. Strategic Management Journal, 2001, 22(6/7): 521-543.

[36] Zhang Jun. Analysis of disruptive innovation characteristics [J]. Market Modernization, 2007(27): 76.

[37] Zhang Dong. Measurement, identification and forecasting of breakthrough innovation for 2035 [J]. China Science and Technology Forum, 2020(8): 11-14.

[38] Zhuang Ziyin, Jia Hongjing, Xiao Chunhuan. Research progress on breakthrough innovation [J]. Economic Perspectives, 2020(9): 145-160.

[39] Della Malva A, Kelchtermans S, Leten B, et al. Basic science as a prescription for breakthrough inventions in the pharmaceutical industry [J]. Journal of Technology Transfer, 2015, 40(4): 670-695.

[40] Dess G G P S D. Porter' s generic strategies as determinants of strategic group membership and organizational performance [J]. Academy of Management Journal, 1984, 27(3): 467-488.

[41] Dosi G. Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change [J]. Research Policy, 1982, 11(3): 147-162.

[42] Fu Yuxiu, Zhang Hongshi. Breakthrough innovation: Concept definition and comparison [J]. Quantitative & Technical Economics Research, 2004(3): 73-83.

[43] Anderson P, Tushman M L. Technological discontinuities and dominant designs: A cyclical model of technological change [J]. Administrative Science Quarterly, 1990, 35(4): 604-633.

[44] Li Liangde, Chen Jin, Mo Xinwei. Research on breakthrough innovation management model [J]. Sino-Foreign Management, 2001(11): 38-41.

[45] Wang X, Wang Z, Huang Y, et al. Identifying R&D partners through subject-action-object semantic analysis in a problem & solution database [J]. Technology Analysis & Strategic Management, 2017, 29(10): 1167-1180.

[46] Hu Zhengyin, Fang Shu. Review of research progress on patent text technology mining [J]. New Technology of Library and Information Service, 2014(6): 62-70.

[47] Blei D M. Probabilistic topic models [J]. Communications of the ACM, 2012, 55(4): 77-84.

[48] De Battisti F, Ferrara A, Salini S. A decade of research in statistics: A topic model approach [J]. Scientometrics, 2015, 103(2): 413-433.

[49] Watts R J, Porter A L. Innovation forecasting [J]. Technological Forecasting and Social Change, 1997, 56(1): 25-47.

[50] Zhao Rongying, Guo Fengjiao, Zhao Yuehua. Mainstream research fields and hot frontiers in scientometrics [J]. Library and Information Service, 2015, 59(2): 66-74.

[51] Tortoriello M, McEvily B, Krackhardt D. Being a catalyst of innovation: The role of knowledge diversity and network closure [J]. Organization Science, 2015, 26(2): 423-438.

[52] Zhang Jinzhu, Zhang Xiaolin. Using citation scientific knowledge mutation to identify breakthrough innovation [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(3): 259-266.

[53] Lu Chao, Hou Haiyan, Ding Y, et al. Review of foreign research on emerging topic discovery [J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(1): 97-110.

[54] Griffiths T L, Steyvers M. Finding scientific topics [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101: 5228-5235.

[55] Wei X, Croft W B. LDA-based document models for ad hoc retrieval [C]//The 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: Association for Computing Machinery, 2006: 178-185.

[56] Levy O, Goldberg Y, Dagan I. Improving distributional similarity with lessons learned from word embeddings [J]. Transactions of the Association for Computational Linguistics, 2015, 3: 211-225.

[57] Zhang Y, Zhang G Q, Zhu D H, et al. Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics [J]. Journal of the Association for Information Science and Technology, 2017, 68(8): 1925-1939.

[58] Shen Jun, Wang Xukun, Gao Jiping, et al. Topic analysis from the perspective of technology coordinates—A case study of 3G mobile communication technology [J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(6): 603-611.

[59] Jiang Xin, Wang Dezhuang, Ma Haiqun. Research on the application of social network analysis methods in library and information science [M]. Beijing: Intellectual Property Publishing House, 2019.

[60] Kong X J, Shi Y J, Yu S, et al. Academic social networks: Modeling, analysis, mining and applications [J]. Journal of Network and Computer Applications, 2019, 132: 86-103.

[61] Song Ge. Research on innovation potential indicators from the perspective of network structure [J]. Library and Information Service, 2014, 58(3): 64-71.

[62] Lu Yongxiang. The significance of disciplinary intersection and interdisciplinary science [J]. Bulletin of Chinese Academy of Sciences, 2005(1): 58-60.

[63] Shang Qi, Chen Hongmei. Patent intelligence demonstration of blockchain technology innovation trend [J]. Journal of Intelligence, 2019, 38(4): 23-28, 59.

[64] Yan Kailun. Research on transaction propagation algorithms and decentralized machine learning frameworks for blockchain [D]. Guilin: Guangxi Normal University, 2021.

[65] He Shuai, Huang Xiangnian, Liu Qianbo, et al. Research on improvement of DPoS blockchain consensus mechanism [J]. Application Research of Computers, 2021, 38(12): 3551-3557.

[66] Zhu Shukun. Design and implementation of electric vehicle energy trading scheme based on blockchain and convolutional neural network [D]. Wuhan: Central China Normal University, 2020.

[67] NSTC. Critical and emerging technologies, CETs [R]. Washington, DC: National Science and Technology Council, 2022.

[68] Wu Jigang, Liu Tonglai, Li Jingyi, et al. Research progress of blockchain technology in mobile edge computing [J]. Computer Engineering, 2020, 46(8): 1-13.

[69] He Shen, Lu Lu, Li Zheng, et al. Blockchain + edge computing technology white paper [R]. Hangzhou: China Mobile 5G Joint Innovation Center, 2020.

[70] Wang Chenxu, Cheng Jiacheng, Sang Xinxin, et al. Blockchain data privacy protection: Research status and prospects [J]. Journal of Computer Research and Development, 2021, 58(10): 2099-2119.

## Author Contributions

Chen Hongshu: Methodology conception, paper writing;
Song Yahui: Data acquisition, experimental work, paper writing;
Jin Qianqian: Methodology validation, paper writing;
Wang Xuefeng: Methodology validation, paper writing.

*Note: Figure translations are in progress. See original paper for figures.*