Journal of Informetrics 14 (2020) 101014

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Topic-linked innovation paths in science and technology

Haiyun Xu

Chengdu Library of Chinese Academy of Sciences, No.16, South Section 2, Yihuan Road, 610041, Chengdu, China

ARTICLE INFO

Article history:
Received 20 June 2019
Received in revised form 16 January 2020
Accepted 16 January 2020

Keywords:
Evolution path
Scientific innovation
Science-technology linkage
Knowledge dissemination

ABSTRACT

In the modern world, science and technology jointly determine the evolutionary path of scientific innovation, with an increasingly close relationship between them. Therefore, it is important to study the identification method of the innovation path, based on the linkage of topics in science and technology. This study focuses on connected topics utilizing bibliometric analysis, thereby exploring the identification method for innovation paths based on the linkage of scientific and technological topics. The internal mechanism of knowledge dissemination and the relationship between science and technology are revealed and described in detail by measuring the linkage of knowledge units. For practical bibliometric analyses, research papers and patent literature were used to characterize scientific research and technological research to reveal the innovation path for the interaction of science and technology quantitatively, automatically, and visually. Experimental study shows that analysis of the topic-linked path of science and technology, along with the integration of multi-relationships, can effectively identify important science- and technology-related topics in a field in the evolution process, and help grasp the key points of basic research and applied research.

1. Introduction

The evolutionary path of the interaction of science and technology refers to the development and evolution of innovative topics and reflects the emergence, diffusion, and evolution of technological innovation. Science and technology, as the two foremost forces jointly determining the development of scientific innovation, merge into an innovation body by mutual promotion or restriction, rather than being in parallel (Chen, 2001). Therefore, the scientific innovation path contains both scientific discovery and technological innovation. The evolutionary scientific innovation process indicates a bidirectional diffusion of knowledge between science and technology, that is, a transfer of knowledge from science to technology and vice versa. Science and technology jointly determine the content and direction of scientific development. Accordingly, it is important to study the identification method of the innovation path based on science-technology topic linkage (Dong, Xu, Luo, Wang, & Fang, 2018; Kostoff & Schaller, 2001; Zhang, Guo, Wang, Zhu, & Porter, 2013, 2016).

The analysis of the evolutionary relationship between science and technology to support decisions cannot be separated from the essence of knowledge evolution. There are two main issues with the identification of scientific innovation paths: (1) previous identification methods primarily focused on one single innovation feature of either science or technology, ignoring the intrinsic linkage of the two areas. That is to say, researchers identified the evolutionary innovation path from the perspective of either scientific research or technological research. (2) The quantitative research on the linkage between science and technology is limited to counting the volume of literature, which cannot reveal the inherent content linkage of

E-mail address: xuhy@clas.ac.cn

https://doi.org/10.1016/j.joi.2020.101014
1751-1577/© 2020 Elsevier Ltd. All rights reserved.

science and technology. These two issues make it difficult to fully grasp the characteristics of scientific innovation and affect the accuracy of evolutionary innovation path identification.

There is abundant literature on the linkage between papers and patents, which shows that these linkages are indeed helpful for constructing high-quality roadmaps (Kostoff & Schaller, 2001) to detect technological opportunity (Albert, 2016). The aims of this study are (1) to discover linked topic evolution paths in science and technology, (2) reveal patterns in the development, transition, and evolution of science and technology, and (3) discover possible new knowledge growth points. This will help technology managers and policymakers to better understand the development of trends in science and technology and provide support for strategic decision-making, innovation resource allocation, and planning for industrial development. To achieve this goal, we propose the topic linkage path model of science and technology using bibliometric indicators. During the data analysis process, a multi-data fusion method was adopted that directly considers the content similarity between nodes.

The remainder of this paper is organized as follows. In Section 2, we introduce the theoretical and analytical framework, including a review of the current status of the research on the identification of innovation paths, the research on the relationship between science and technology, and the theoretical framework of the research. Section 3 discusses the empirical method to identify topic-linked evolution paths of science and technology, followed by an empirical study in genetically engineered vaccines (GEVs) in Section 4, which ends with a presentation of the results. In Section 5, the benefits and limitations of the method are highlighted, and the conclusions are presented.

2. Theoretical and analytical framework

2.1. Theoretical framework

2.1.1. Interactions of science and technology

Exploration of the relationship between science and technology is a process that ranges from theoretical analysis to empirical measurement, and it is furthermore a process that ranges from simple citation analysis to multi-feature analysis (Dong et al., 2018).

To reveal the internal relationships of knowledge in science and technology, it is necessary to delve into the knowledge system and to analyze it from the perspective of the dissemination of knowledge between knowledge units. In this study, we explore the more precise innovation paths in specific research domains through the mutual prompts of science and technology. We summarize the relationship between science and technology, as well as the relationships among their concepts as follows: the development of science stems from the accumulation and breakthroughs of basic research, and the development of technology comes from the improvement of applied research levels and developmental research levels (as shown in Fig. 1).

In this study, the phrase “basic scientific research” (“basic research” in Fig. 1) denotes scientific research with the purpose of increasing the understanding of “nature,” oriented toward discovering (basic) mechanisms of nature. We define “applied research” as research that applies scientific knowledge to solving technological issues (Tijssen, 2010; Tijssen & Winnink, 2016). There is no clear boundary between basic research, applied research, and science and technology because they all have a shared component, which is the application of basic research. The existence of this shared component means that basic research, applied research, and technology are linked. Science and technology are distinct but interrelated, and their interaction forms a complex system of knowledge exchange.

At the macro level, the interaction between science and technology is bidirectional, and this interaction determines the content and direction of technological development. At the micro level, the development of science and technology follows the rules of self-organization and evolves under chain logic (Arthur, 2009; Raan, 2000; Noyons & Raan, 1998). Science and technology are essentially different from each other because the development mechanism, the motivations and principles of development, and the evolution differ. Scientific research and technological innovation activities continue to improve in their own development trajectories in R&D activities. It is, therefore, necessary to study the development of scientific research and technological innovation through their interactions.

The interrelated and relatively independent relationship between science and technology makes it possible to reconstruct how science and technology jointly promote innovation development in their interaction. Hence, it is necessary to obtain a method to identify the scientific and technological innovation path based on the linkage of science and technology. However, there is still a lack of practical ways to find an innovation path in such a linked way. In the next section, we design the bibliometric process to find the innovation path in specific domains.

2.1.1.1. Differences between science and technology. The aim of science is to clarify the nature, characteristics, and laws of natural phenomena, in other words, the theoretical achievements. Moreover, technological achievements are identified as new technologies, new processes, new products, and new methods, which are products of technological innovation activities (Zhang, 1998). Traditional knowledge theory holds that technology is a type of applied science; therefore, science can develop without technology, whereas the opposite is not true. With a deeper understanding of the science-technology relationship, researchers began to explore the general processes and laws of the interaction between science and technology. Dong et al. (2018) systematically discussed and summarized the typical relationship patterns between basic research and applied research.

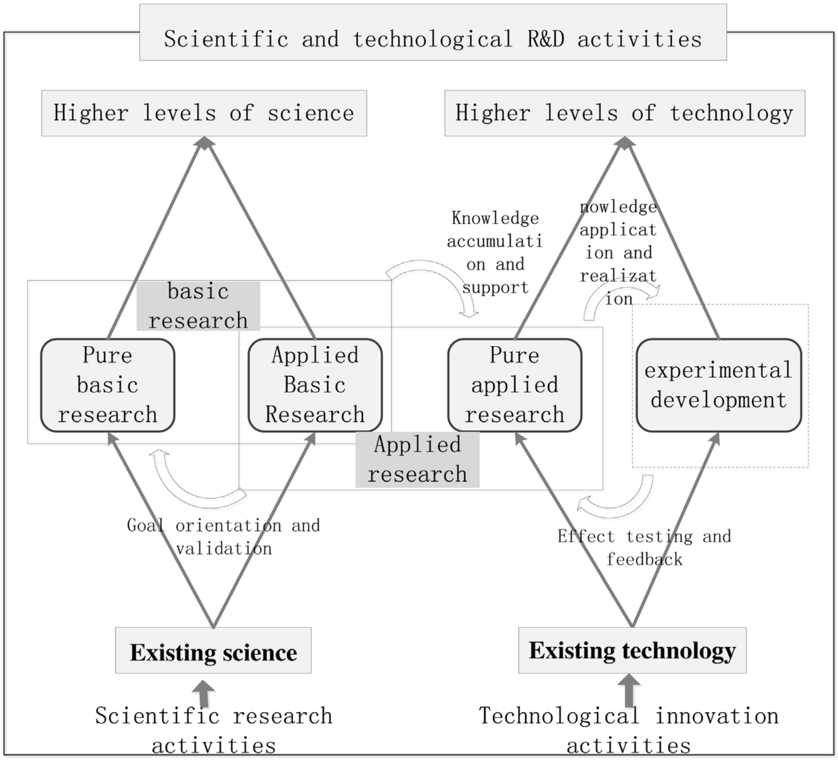
Scientific and technological R&D activities
Higher levels of science
Higher levels of technology
Knowledge accumulati on and support
nowledge applicat ion and realizat ion
basic research
Pure basic research
Applied Basic Research
Pure applied research
experimental development
Applied research
Goal orientation and validation
Effect testing and feedback
Existing science
Existing technology
Scientific research activities
Technological innovation activities

Fig. 1. Schematic diagram of the linkage between science and technologys.

Existing studies generally believe that the marginality and cross-cutting of science, as well as the interdisciplinary and integration of technology, lead to the horizontal and vertical integration of science and technology, which makes the relationship between science and technology a two-way and dynamic structural model (Li, 2011). Guan and He (2007) regard the exchange of information between science and technology as a double-helix model. Four conceptual views for the relationship between science and technology are defined by Gardner (1999): 1) the “demarcation view” when the two elements are considered to be independent, 2) the “ideal state view” where scientific developments precede technological advances, 3) the “materialist view” when technology develops before science, and 4) the “interaction model” when science and technology develop interactively.

2.1.2. Measurement the linkage between science and technology

2.1.2.1. Using citation relationships to identify relations between science and technology. Most existing bibliometric methods regard research papers as the representation of scientific research and patents as the representation of technological innovation (Meyer, 2002). Journal papers are considered to cover the knowledge production in scientific research, and patent documents mainly represent technological inventions and the fields in which these inventions can be applied (Kwon, Porter, & Youtie, 2016; Verbeek et al., 2002). Previous studies on linkages between the scientific publications and patents have mainly focused on the analysis of nonpatent literature references (NPRs).

Narin and Noma (1985) and Schmoch (1993) found that there were different kinds of interactions between science and technology, and the strength of the interaction varied in different fields after establishing the link between science and technology from the perspective of patent citations. Murray (2002) proposed a quantitative and qualitative combined methodology of patent–paper pairs to exemplify the intertwined and co-evolutionary scientific and technical ideas and communities. Besides publishing and citation analysis, founding, licensing, consulting, and advising were also applied, which provides more implications for understand the commercialization and technology transfer and the spillovers that arise. Glänzel and Meyer (2003) and Meyer and Debackere (2010) established a connection between subject areas and technical fields by patent citation analysis and found that applied science with high technology linkage was more influential than the pure basic science.

Owing to the deficiency of one-way analysis, some researchers use a two-way citation method to study the relationship between science and technology. McCalman (2001) pointed out that non-patent literature, especially highly cited scientific papers, was an important knowledge source for invention patents, and patents provided conditions and means for scientific development. Gao, Ding, Teng, and Pang (2012) tried to use the paper-patent hybrid co-citation analysis method to analyze the interaction between science and technology in terms of the co-citation situation and used this method to study the mutual fusion of science and technology in knowledge dissemination. Huang, Yang, and Chen (2015) found that the trend of correlation and integration of science and technology in the field of fuel cells was gradually increasing, although not very steeply, after analyzing the cross-citations of papers and patents in the field. The two-way citation linkages between· papers and patents, both scientific literature cited in patents and patents cited in scientific literature, are an important method to analyze the flow and spread of knowledge between science and technology.

There are several issues with using NPRs to study the associated linkages between science and technology: there are differences between countries in terms of how NPRs are cited (Michel & Bettels, 2001), about 30–40 % patents contain NPRs (Callaert, Van Looy, Verbeek, Debackere, & Thijs, 2006), and the contents of patents and cited papers are not always related (Meyer, 2000). The reasons for the lack of citation information are diverse, for example, due to citation incompleteness, citation relationships between topics in both "worlds" are not always known. Moreover, regardless of patent-to-paper citations or paper-to-patent citations, the links between the scientific publications and patents are noisy because the citation behavior varies from authors and inventors to reviewers and examiners (Li, Chambers, Ding, Zhang, & Meng, 2014).

In this situation, text analysis will supply more information, which could be reflected by content linkage. Specifically, it will improve the validity of topic linkage analysis for those situations where a citation is lacking in a patent. In this case, there is an implicit relationship between patent and papers that could not be connected by citation analysis but could be expressed by text analysis.

2.1.2.2. Exploration of the relationship between science and technology based on other features. Relationships between scientific publications and patents that are based on common features can effectively reflect the interaction between science and technology. Bassecoulard and Zitt (2004) used a chemical abstract database as the data source and linked science and technology by establishing the lexical correspondence between papers and patents. Zhang (2014) found that scientific knowledge and technical knowledge have a significant positive impact on each other after analyzing the interaction between science and technology and the role of academic inventors (article authors and patent inventors) in the interaction with researchers as the linked features. Stokes (1997) proposed a two-dimensional quadrant map of scientific research in 1997. The basic research that solves the application issue is visually summarized as the Pasteur quadrant, which was based on Bush's (1945) linear model of research and development.

These studies on the science-technology relation between research papers and patent document measurements mostly indicate a static relationship between science and technology, and therefore, cannot represent the temporally complex and dynamic nature of these relations.

Several concepts have similar connotations with innovation paths, such as technological paths, technological trajectories, technological maps, technological skeleton maps, and technology main paths (Tu & Hsu, 2016). Technology road mapping is a strategic management tool for situational analysis, strategic choice, path planning, process design, and policy and evaluation planning. It is increasingly used for long-term strategic planning of national scientific and technological development (Lee, Kang, & Park, 2007; Vishnevskiy, Grebenyuk, & Kindras, 2011). Technology roadmaps of knowledge discovery not only can grasp the direction of scientific and technological development and long-term development trend of a specific field but can also obtain more authoritative information with decision supporting value, such as candidate technologies and key technologies available in the field in different countries (Kostoff & Schaller, 2001; Letaba, Pretorius, & Pretorius, 2015). In this study, we use the term "innovation" for advances in scientific discovery and technological improvement. To strengthen the essence of knowledge evolution, which is the interaction of scientific discovery and technological innovation, it is better to use combinations and interactions of the features in both domains than to focus on one single feature from science or technology.

The relationship between scientific topics and technical topics in the path of scientific innovation is the result of diffusion of heterogeneous knowledge (Kang, 2016). Given the shortcomings of citation analysis, citation analysis cannot fully express the linkage of science and technology topics. This paper proposes the use of a topic linkage method based on text analysis to measure the theoretical basis of the relationship between scientific topics and technical topics. The latent Dirichlet allocation (LDA) topic model will be used to obtain innovative topics on the innovation path, and the multi-fusion method will be used to form a more accurate measure of topic linkage.

2.2. Analytical framework

The identification of a topic-linked evolution path of science and technology includes two important issues: 1) the identification of evolution paths in scientific discovery and technological innovation, and 2) the identification of topic linkages in the two innovation paths. This section discusses the theoretical background of this study, including the theoretical basis and the implementation method used.

2.2.1. Identifying evolution paths in science and technology

Citation analysis and text analysis are two different methods often used for identifying the evolution path of technological innovation. The former analyzes the evolution of the citation network based on the connectivity of the citation links and the network structure, whereas the latter identifies the evolutionary trajectories of innovation topics along the time axis, mainly based on the text-similarity of the topic content. In addition, the multivariate relationship fusion method provides a

Table 1
Methods of co-word analysis.

Basic co-word analysis	Coulter, Monarch, and Konda (1998) and Kim, Suh, and Sang (2008) clustered keywords based on the co-occurrence of the keywords in the thesis and patent documents and pointed out the evolution trend of the topic after analyzing the sequential changes of the keywords. Wu (2016) identified the domain topic evolution path by constructing weighted keyword co-occurrence networks.
Semantic-enhanced co-word analysis	Blei, Ng, and Jordan (2003, 2006) proposed Latent Dirichlet Allocation (LDA) model to discover the topic in the text, which could express the hierarchical semantic relationship among words based on the statistical probability level, and further proposed a dynamic topic model that could reveal the evolution of the topic. Wang, Li, Li, and Li (2012) constructed a thesaurus based on expert knowledge to explore the semantic relations among the topic terms and obtained the topic evolution by clustering the semantic similarity of the topic terms. Zhu, Zhang, and Wang (2017) expressed deep semantic features with the LDA model and obtained the topics in different time intervals for the domain topic evolution path.
Dynamic evolution of keyword analysis	Morris, Yen, Wu, and Asnake (2003) used literature-coupled clustering to identify research frontier topics and presented their evolution with the innovative timeline mapping method. Rosvall and Bergstrom (2008) proposed a community evolution visual analysis method based on an alluvial diagram in the field of geography, which could display the evolution process of topic structures. Cui et al., (2011) proposed a text visualization analysis method named TextFlow, which could analyze multiple topic evolution relations. TextFlow introduced the concept of topics merging and splitting in a massive text analysis, which could enable people to quickly grasp the development of massive information with intuitive streaming graphics. Wei et al. (2016) generated a variety of innovative evolutionary paths for time series evolution by calculating the coincidence of core nodes in a special topic.

new way to improve the accuracy of identifying an innovation evolution path by integrating several types of topic-related linkages.

2.2.1.1. Main path analysis in citation networks. The idea of "main path" was first proposed by Hummon and Dereian (1989), who extracted the key path of the citation network based on the connectivity of the citation network for tracing the most significant paths in a citation network; it is commonly used to trace the development trajectory of a research field. Pilkington and Meredith (2009) combined citation analysis and co-citation analysis with factor analysis. These authors concluded that using these combined methods, the relationships that are found are more accurate than those found by using only citation analysis. Liu and Lu (2012) proposed an integrated approach to main path analysis by providing the global main path, the backward local main path, multiple main paths, and key-route main paths. Each path is obtained from a perspective different from the original approach. The integrated approach provides several paths that are not captured by the original approach. Martinelli (2012) and Lu and Liu (2014) described the scientific structure and technological evolution path of specific fields by identifying the main path of the paper or patent citation network for multiple fields. Lai and Li (2015) carried out cluster analysis on patent citation networks and identified the technological evolution path by analyzing the sequential linkage of citation nodes. Zhu (2014) and Chen, Yang, Zhang, and Fan (2015) combined thematic content analysis with citation master path analysis and then extracted the main path of semantic enhancement by mining the linkage of the citations on the text content, thereby revealing the domain evolution path and trend more deeply. Martinelli & Nomaler (2014) proposed a genetic approach for identifying technologically important patents within a patent citation network using the patent's persistence index. Their empirical analysis shows that the method is successful in reducing the number of both nodes and considered links and can identify technological discontinuities. Park & Magee (2017) proposed a methodology of searching backward and forward paths from high-persistence patents, which are identified using a standard genetic knowledge persistence algorithm. The empirical results show that the method can dramatically reduce network complexity, without missing dominantly important patents, and can identify less complex main paths. Lathabai, George, Prabhakaran, and Changat (2018) proposed an integrated approach to path analysis using FV gradient weights for weight assignment of citation networks. The case study demonstrated that the FV gradient method can be used as a weight assignment scheme for the retrieval of paths that might not be highlighted by the use of SPX methods.

2.2.1.2. Co-word analysis. Co-word analysis is the most commonly used method for innovative path identification based on topic analysis. Considering that the basic co-word analysis fails to characterize the semantic relationship among the topic terms fully, many scholars have already identified the topic evolution path with the help of semantic-enhanced topic term analysis, and some scholars have also studied the dynamic process and different modes of the path. Table 1 describes the three most important variants of co-word analysis.

2.2.1.3. Multivariate-relationship. Which relatedness measure yields the most accurate clustering of publications? One perspective is that there is no absolute notion of accuracy (Gläser, Glänzel, & Scharnhorst, 2017). Following this perspective, each relatedness measure yields clustering solutions that are accurate in their own right, and it is not meaningful to ask which clustering solutions are overall the most accurate ones. Different citation-based and text-based relatedness measures emphasize different aspects of the way in which publications are related to each other, and the corresponding clustering solutions provide a legitimate viewpoint on the organization of scientific literature (Waltman, Boyack, Colavizza, & Van Eck, 2017).

Multivariate data fusion refers to the comprehensive analysis of different types of relational data through specific models, revealing the research object by integrating all information together, which could make up the deficiency of single relationship-based analysis. The multi-relationship fusion provides a more effective analysis basis for innovation evolution research, through combining multiple relationships among different entities based on different attributes of scientific papers into a new relationship (Xu et al., 2017).

Jensen, Liu, Yu, and Milojevic (2016) correlated many attributes such as documents, topic words, authors, and citations, using the meta path method, presented the relatedness and similarity of different bibliometric entities, and initially applied it to the exploration of topic evolution. They also applied this method to topic evolution exploration. Liu, Wang, and Bai (2016) proposed a multi-dimensional subject evolution analysis model, which constructed topic evolution visualization maps of the subject strength, structure, and content. On this basis, evolution analysis of multi-disciplinary topics was carried out, which further met the deep semantic information needs in technological innovation. Kang (2017) constructed a hybrid network analysis model combining coupling analysis and co-citation analysis from the perspective of heterogeneous hybrid networks. Moreover, this model broke through the limitations of single homogenous networks in knowledge discovery with strong operability.

2.2.2. *Identifying science-technology related topics*

According to the characteristics of knowledge carriers, knowledge dissemination has different characteristics in terms of content, hierarchy, and structure. They can be divided into two dissemination types: homogenous dissemination and heterogeneous dissemination. In terms of bibliometric methods, homogenous dissemination is the interaction between the same literature types, mainly referring to the knowledge flow between patent documents or non-patent literature, and heterogeneous dissemination refers to the knowledge dissemination between patent documents and non-patent literature. Heterogeneous dissemination easily increases the uncertainty of the future development of technology and the probability of innovation damage, the nonlinear change of technology track, and the opportunities for cross-domain application of new technology. It also stimulates production of new knowledge, new thinking, and new technologies, which makes it easier to simulate breakthrough innovations.

Most existing research has conducted non-patent literature citation analysis, which can better reveal the various relationships within heterogeneous knowledge dissemination. Van Looy, Magerman, and Debackere (2007) used the amount of non-patent literature as a scientific strength indicator, confirming that scientific intensity can be used as a dynamic indicator for assessing the status of technological development. McCalman (2001) pointed out that heterogeneous dissemination is a two-way and interactive spiraling process. Non-patent literature, especially high-quality scientific papers, is an important knowledge base and source of invention patents, whereas patents provide conditions and means for scientific development. Therefore, studying the internal mechanism of knowledge dissemination by analyzing and summarizing the characteristics of heterogeneous knowledge can better reflect the deep knowledge flow implied in a specific technical area. Kang (2016) constructed data linkage rules from three aspects—data linkage times, correlation strengths, and correlation coefficients, which overcome the shortcomings of linkage times or correlation strengths that only reflect the similarity and avoid the random factor caused by the connection strength with a combination of the application of correlation coefficients and linkage times.

2.2.2.1. *Identifying linked topics in science and technology.* Themes in science and technology topics both contain innovation knowledge from the authors or inventors. Thus, one will miss many valuable scientific papers that have not yet been cited by patents. There already has been some pioneering work on discovering the linkage between science and technology based on topic similarity. Xu, Zhu, Qiao, Shi, and Gui (2012) proposed a simple procedure for constructing topic linkages between papers and patents by analyzing these two types of information resources simultaneously. A novel statistical entity-topic model, CCorrLDA2 model, was proposed by Xu, Luo, and Li, 2019, Xu, Zhai et al., 2019) to discover the hidden topics from scientific publications and patents. However, these prior works did not go into the dynamic process of the linkage.

The topics of similar fields of science and technology transcribe similar ideas, yet the texts are distinct—a paper describes experimental results, whereas a patent defines utility and makes claims of inventiveness. In the science-technical correlation calculation, the key to realizing interactive pattern identification is to match the corresponding scientific terms related to technical terms semantically and implicitly in the domain vocabulary. Moreover, the integration of scientific-technical semantic linkages is a good choice to avoid the preceding deficiency. However, at present, there is still no scientific vocabulary and technical vocabulary that can be contrasted and connected in most fields. Therefore, an effective connection between scientific terms and technical terms in specific fields will be formed. Scientific-technical semantic linkage integration can be achieved through a feasible algorithm design.

Here, the primary issue is that single linkage analysis in text analysis, such as co-word analysis, has difficulty in identifying a meaningful topic because of insufficient semantics. Therefore, we chose to use multivariate relationship fusion to identify the topic by text analysis on research articles and patent literature, thus solving the inaccuracy of text topic discovery. If the construction of the scientific-technical topic network can be realized, and the connection of scientific topic words and technical topic words can be realized at the topic level, the essential relationship between science and technology, the

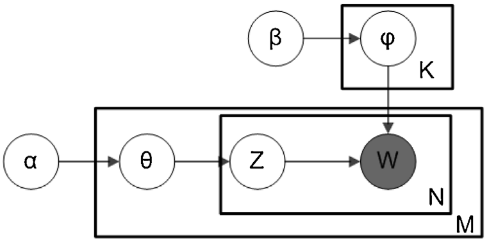

Fig. 2. Rational of LDA schematic.

diffusion and evolution path of scientific and technological innovation, will be more accurate. For the identification of topics, we rely on "LDA topic modeling" and "multi-relation fusion."

2.2.2.2. LDA topic modeling. LDA is an unsupervised machine learning method proposed by Blei, Ng, and Jordan (2008). LDA is generally regarded as a topic model for processing documents, primarily by representing documents as a topic vector to achieve a reduction of high-dimension features. For a vocabulary size V, each word is represented by a V-dimensional vector, and for a document composed of N words, it is represented as $w = (w_1, w_2, \ldots w_N)$, for a corpus composed of M documents. For $D=\{w_1, w_2, \ldots, w_M\}$, the process of generating a document is as follows. First, generate a document with an N length window from a global Poisson distribution (parameter β). Generate θ of the current document in the Dirichlet distribution (parameter α). Then, perform the following two steps for each word of the current document with an N length window. First, from the multinomial distribution with θ as the parameter, generate a topic (Topic) Zn. Second, generate a word (Word) Wn from a multinomial distribution with β and Zn as parameters together (as shown in Fig. 2).

The LDA model is an important algorithm in the field of natural language processing. The application scope includes text topic recognition, text classification, and text semantic similarity calculation. LDA has been successfully used in many types of related topical analysis.

Griffiths and Steyvers (2004) used LDA to extract topics from PNAS abstracts from 1991 to 2001. By analyzing the topic dynamics, they illustrated the relationships between different scientific disciplines and assessed their trends. Weng, Lim, and Jiang (2010) applied LDA to the goal of topic distillation, which automatically identified topics that twitterers were interested in based on the tweets they published. Wei and Croft (2006) used LDA to improve ad-hoc information retrieval, and experiments demonstrated that an LDA-based document model consistently outperformed the cluster-based approach.

To meet different needs, LDA-based model extensions have emerged. The combination of LDA and other textual information, such as the author information, forms an author-topic model that improves the accuracy of topic recognition (Rosen-Zvi, Griffiths, & Steyvers, 2004). The advantage of an unsupervised learning algorithm is that it is not necessary to manually label the training set during training; it is only necessary to determine the number of topics in the document set. In addition, after using the LDA model, all topics in the document set can be described with a set number of words. Mcauliffe and Blei (2008) introduced supervised LDA, which is more advanced than unsupervised LDA. The LDA topic model treats each document as a word frequency vector with the bag-of-word model. Although it is easy to transform the text into digital information, it does not consider the order of words.

2.2.2.3. Multi-relation fusion. With the explosive growth of the amount of scientific literature and the continuous enrichment of the types of documents, the relation types that can be analyzed by scientific measurement are constantly growing. In addition to citation analysis and text topic analysis, multiple coupling analysis among different entities has also been successfully applied in some research, and multi-source heterogeneity has become an important and common form of data.

Multi-data fusion refers to the integration of multiple related data through data fusion algorithms to obtain more accurate semantic linkages between entities or knowledge units through richer information extraction and screening (Xu et al., 2017). Therefore, the weakening of the single linkage relationship leads to an analysis bias due to the insufficient expression of the entity linkage feature to ensure that subsequent topic recognition becomes more objective and more realistic. However, the current topic acquisition methods for this document rely mostly on single linkage analysis, so it is difficult to obtain the topics of scientific or technological developments accurately. Therefore, finding multi-relationships between the entities of a document is one of the key technologies for accurate topic identification in massive scientific or technological works of literature.

Xu, Luo et al. (2019), Xu, Zhai et al. (2019) chose authors, references, and citation literature as the intermediate MEs of topic term strengthened relations and additional relations. Seven types of topic linkages are formed by the topic terms and these MEs. The fusion relationship can make up for the lack of information in a single linkage relationship by obtaining more accurate topic linkages. An empirical analysis proved that multi-relation fusion can effectively improve the effect of topic clustering.

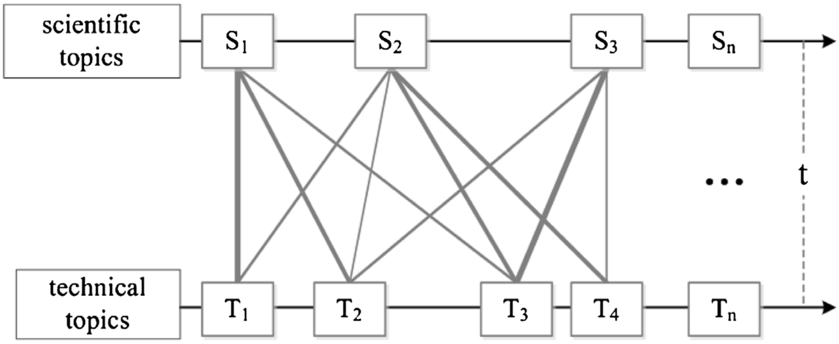

Fig. 3. Schematic diagram of the science-technology interaction pattern.

3. Methodology

3.1. Topic linkage path model of science and technology

For the practical bibliometric application, research papers and patent literature were used to characterize scientific research and technology research, respectively, to reveal innovation paths under the interaction of science and technology in a quantitative, automated, and visualized manner.

Moreover, this model is established to analyze the relationship between topics in science and technology, while maintaining their relative independence status, therefore, we did not just blend the papers and patent literature together to perform a single topic identification. Fig. 3 is a schematic diagram of science-technology interaction pattern identification. $S_1, S_2, S_3 \ldots S_n$ are scientific topics, and $T_1, T_2, T_3 \ldots T_n$ are technical topics. All the scientific topics and technical topics are distributed on the horizontal timeline (T), and the line thickness indicates the different intensities of topic linkage between scientific topics and technical topics. We can calculate the relatedness of science topics and technology topics by text analysis, and if the relatedness is up to a certain threshold, we can consider that there is meaningful relatedness between the science and technology topics.

3.2. Construction of topic linkage between science and technology

Finding meaningful topics is a critical step in the analysis of science-technology linkages based on text information. Using single data relationships, it is rarely possible to identify meaningful topics. Conversely, the multi-data fusion method plays a significant role in the identification of meaningful topics using text analysis.

The calculation methods of topic linkage degree between any scientific topic (topic S) and any technological topic (topic T) are shown in Fig. 4. In this study, there are three types of topic linkages: co-word linkage, co-author linkage, and co-citation linkage. Co-word linkage means that two topics are semantically related because they contain the same topic terms. Two

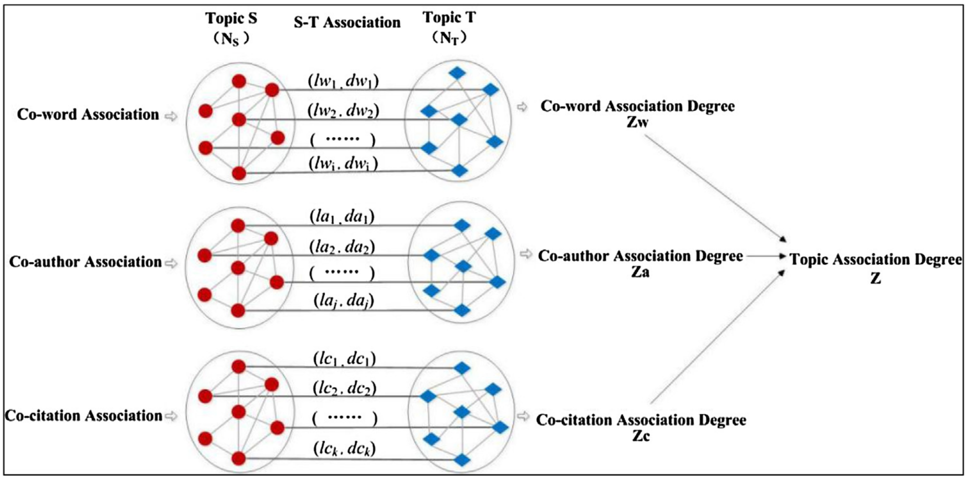

Fig. 4. Schematic diagram of topic linkage degree calculation.

topics that share the same author are supposed to be semantically related by co-author linkage. Co-citation linkage indicates that two topics are semantically related because reference or citation relationships exist between them (Xu et al., 2017).

In Fig. 4, topic S contains N_s topic terms (red dots), and topic T contains N_T topic terms (blue diamonds), where the linked paths between them are denoted by l. l_w, l_a, and $l_c$specifically represent the paths of co-word linkage, co-author linkage, and co-citation linkage, respectively. d_w, d_a, and d_c represent the corresponding linked weights. According to the three types of linked relationships, we calculate the co-word linkage degree Zw, co-author linkage degree Za, and co-citation linkage degree Zc, respectively, and then calculate them synthetically. Finally, the topic linkage degree Z between topics S and T can be obtained. The concrete calculating methods are as follows.

3.2.1. Calculation of co-word linkage degree

As shown in Fig. 4, topics S and T contain some of the same topic terms, so there are multiple linked paths between them. Eq. (1) for calculating the co-word linkage degree Zw is as follows:

$$\mathrm{Zw} = \frac{p}{N_s N_T} \sum_{i=1}^{p} d_{wi} \tag{1}$$

where d_{wi} indicates the co-occurrence linkage degree of topic terms. $N_s N_T$ indicates the number of all linked topic terms between topics S and T. The function of $\frac{p}{N_s N_T}$ is to change the absolute value of the topic linkage degree into a relative value, avoiding the scenario that the absolute value obtained by summation is too high because some topics contain too many topic terms. The linked weight d_{wi} is determined by the frequency of the topic terms connected by path l_{wi}. If there exists a co-word linkage connected by path l_{wi} between X (number of topic S terms) and Y (number of topic T terms), the value of d_{wi} is X + Y.

3.2.2. Calculation of co-author linkage degree

As shown in Fig. 6, there are several co-author linkage paths between topics S and T because some of the topic terms share co-authors. Eq. (2) for calculating the co-author linkage degree Za is as follows:

$$\mathrm{Za} = \frac{p}{N_s N_T} \sum_{j=1}^{p} d_{aj} \tag{2}$$

where d_{aj} indicates the absolute value of the co-author linkage degree. $N_s N_T$ indicates the number of all possible topic term linkage paths between topics S and T. The function of $\frac{p}{N_s N_T}$ is to change the absolute value of the topic linkage degree into a relative value, avoiding the scenario that the absolute value obtained by summation is too high because some topics contain too many topic terms. The path weight d_{aj} is determined by the frequency of the topic terms connected by path l_{aj}. If there exists a co-author linkage connected by path l_{aj} between X (number of topic S terms) and Y (number of topic T terms), the value of d_{aj} is X + Y.

3.2.3. Calculation of co-citation linkage degree

As shown in Fig. 6, there are several co-citation linkage paths between topics S and T because of the mutual reference of topic terms (including the relationship of citing and cited). Eq. (3) for calculating the co-citation linkage degree Zc is as follows:

$$\mathrm{Zc} = \frac{p}{2N_s N_T} \sum_{k=1}^{p} d_{ck} \tag{3}$$

where d_{ck} indicates the absolute value of the co-citation linkage degree. $N_s N_T$ indicates the number of all possible topic term linkage paths between topics S and T. The function of $\frac{p}{N_s N_T}$ is to change the absolute value of the topic linkage degree into a relative value, avoiding the scenario that the absolute value obtained by summation is too high because some topics contain too many topic terms. The path weight d_{ck} is determined by the frequency of the topic terms connected by path l_{ck}. If there exists a co-author linkage connected by path l_{ck} between the X topic S terms and Y topic T terms, the value of d_{aj} is (X + Y)/2. (The frequency of topic terms is calculated repeatedly when calculating references, so it needs to be halved.)

Finally, the scientific-technological topic linkage is obtained, and Eq. (4) for calculating the integrative topic linkage degree is as follows:

$$Z = \alpha \frac{p}{N_s N_T} \sum_{i=1}^{i=p} d_{wi} + \beta \frac{p}{N_s N_T} \sum_{j=1}^{j=p} d_{aj} + \gamma \frac{p}{2N_s N_T} \sum_{k=1}^{k=p} d_{ck} \tag{4}$$

where α, β, and γ are the weight coefficients of the co-word linkage degree, co-author linkage degree, and co-citation linkage degree, respectively, which are determined by the specific meaning of their linkage degrees and the effect on the topic linkage degree.

3.3. Topic distribution time

After obtaining the topics based on multi-relation fusion of science and technology, the topic linkage between science and technology is calculated along the time axis. The topics on the timeline are not an even distribution, and the frequency

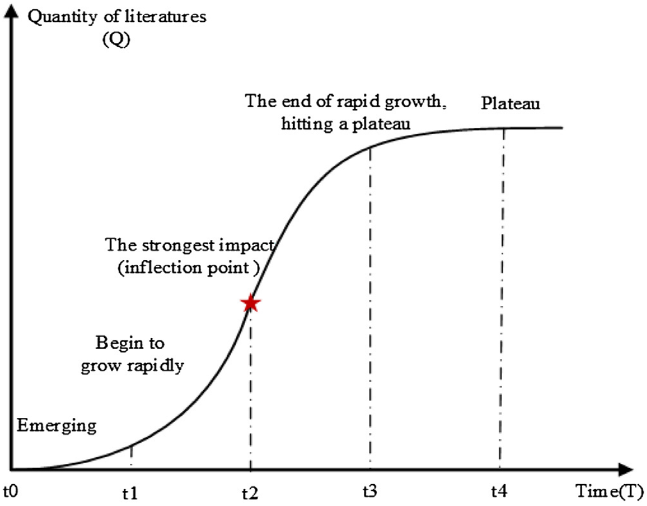

Fig. 5. Schematic diagram of the topics' time distribution.

of the topic distribution usually changes with time and shows multiple modes of trends, such as decrease, rise, first rise then decrease, and first decrease then rise. In this situation, the stability and persistence of linked topics in science and technology are important; otherwise, these topics will not have enough impact. In contrast to co-citation clusters, direct citation has stability with a high degree of dynamic variation (Small et al., 2014). In the topic prominence analysis performed by Klavans and Boyack (2017), the combination of stability and dynamics associated with direct citation clusters is appropriate for the purposes, whereas new topics need not be realigned with the influx of publications from the most recent year. This study intends to show the interrelationships between topics, and the same topic could appear in different time periods. However, to clearly show the chronological correlations between important topics, only the most influential time point of the important topics on the innovation path will be shown. In this study, the inflection point in the life cycle curve was chosen as the most influential time point.

The average time of cumulative values of scientific literature is typically used as a time point in scientific and technical involution path maps. However, the publication of scientific papers or the trend of patent applications does not accumulate in a linear way. Following the law of index development, it presents a skewed distribution, that is, a large amount of literature concentrated in recent years. Therefore, choosing the average time of publications on the timeline will cause a lagging trend for topic distribution time. Furthermore, the earliest publication year or application year may only indicate sporadic research, which did not have any significant influence.

Usually, soon after a major development has been made or when a breakthrough has been achieved, a research theme will attract the attention of scientists who conduct research on the same or a closely related topic. Over time, the subject will receive increasing attention, and the number of scientific publications will grow and eventually be leveled off.

This process can be represented by a growth function curve (Andersen, 1999). As shown in Fig. 5, the inflection point of the curve marks the moment the growth rate starts to decrease, As shown with an asterisk ("*") in Fig. 5. The topic is in a period of rapid growth near the inflection point, and the published scientific documents have already accumulated to a certain quantity but have not yet reached the maximum quantity. When the number of publications is close to the maximum, it means that there are many followers. In fact, the topic has become a mature research hotspot and has passed its most influential moment. In this study, the count of publications for the topic represents the attention and impact that the topic attracted. Therefore, theoretically, the inflection point captures the topic distribution time at an early and influential time point.

4. Empirical analysis

4.1. Data sources

In this study, the GEV field was selected as our experimental field. GEV is a research area that is expected to be the source of new innovative vaccines. Papers and patent documents are collected in the GEV field for a period as analysis data sets. The Web of Science database (WoS) is selected to search scientific papers, and the Derwent Innovations Index database (Derwent) is used to search patent documents. Both WoS and Derwent are databases provided by Clarivate Analysis. Data collection was performed on January 6, 2018, and the publication year is up to 2017. The resulting dataset consists of 4146 records for scholarly publications and 4050 records for patent publications. The next two database-specific queries were used to collect the data.

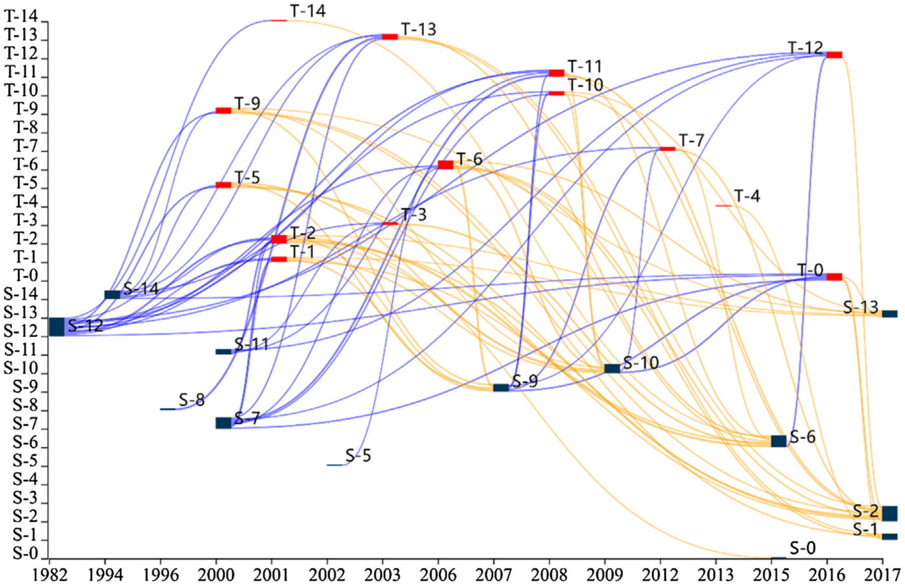

Fig. 6. Topic-linked path of science and technology (threshold = 0.5).

4.1.1. *Selecting scientific publications*

To select scientific publications in the WoS database, we used the following query:

TS=((Genetic* adj engineer* or DNA adj engineer*) and (vaccine* or antigen*)) or TI=((nucleic* adj acid* or RNA) and (vaccine* or antigen*)) or TI=((plasmid* adj DNA) and (vaccine* or antigen*)) not TI=(test or immunoassay adj detect* or detect*); the document type is limited to Article OR Proceedings Paper.

4.1.2. *Selecting patent publications*

To select patent publications in the Derwent database, we used the following query:

TS=((Genetic* adj engineer* or DNA adj engineer*) and (vaccine* or antigen*)) or TI=((nucleic* adj acid* or RNA) and (vaccine* or antigen*)) or TI=((plasmid* adj DNA) and (vaccine* or antigen*)) not TI=(test or immunoassay adj detect* or detect*) not PN=(WO2015052345-A1 or WO2015051850-A1 or CN104165998-A or WO2014177042-A1).

4.2. *Identifying topics*

4.2.1. *Topic identification based on multi-relation fusion*

The topic visualization effect along the timeline would result in poor visualization, with more than 30 topics with lines interlinking the topics. To overcome this visualization issue, we used the following four-step approach.

Step 1 Fields such as title, abstract, author, and cited journal were extracted from the articles and patent datasets. Likewise, keyword phrases from the patents and the abstracts of the articles were extracted using the KEA (http://community.nzdl.org/kea/) algorithm and were added to the subject field of patents and papers.

Step 2 With the LDA topic model, 15 scientific topics, Si, i $\in$ [0, 14], and 15 technical topics, Tj, j $\in$ [0, 14], were acquired. Then, according to the distribution probability of document-topic, the documents were allocated to different topics.

Step 3 The co-word similarity, co-author similarity, and co-citation similarity between the topics, Si and Tj, were calculated. Consider the co-author similarity as an example. First, all of the authors in Si and Tj were identified and passed into arrays i and j, respectively. Then, the intersection and union between arrays i and j was calculated. In the same way, the co-word similarity and co-citation similarity between the scientific topic Si and the technical topic Tj were calculated. Finally, all scientific topics Si, technical topics Tj, the co-author similarity matrix, the co-citation similarity matrix, and the co-word similarity matrix were traversed, and the scientific and technical topics were finally calculated.

Step 4 According to the topic fusion correlation degree in Eq. (4), we calculated the fused scientific topic and technical topic-linked degree of multi-relationship. The values for the parameters alpha, beta, and gamma used in Eq. (4) were calculated for the GEV case as follows:

Because the co-word linkage is a direct linkage to topic terms, while co-author linkage and co-citation linkage are indirect association, it is supposed that there is a decreasing order in topic linkage calculation for co-word linkage, co-author linkage, and co-citation linkage. In the GEV field, according the data feature, the weight coefficients of co-word linkage degree, co-author linkage degree, and co-citation linkage degree are set as $\alpha = 0.5$, $\beta = 0.3$, and $\gamma = 0.2$.

Table 2
Topics' distribution years.

Science topics	Publication year	Article counts	Technical topics	Application Year	Patent counts
0	2015	148	0	2016	295
1	2017	303	1	2001	526
2	2017	206	2	2001	302
3	2015	148	3	2003	246
4	2013	181	4	2013	306
5	2002	218	5	2000	178
6	2015	278	6	2006	356
7	2000	381	7	2012	274
8	1996	135	8	2009	183
9	2007	665	9	2000	402
10	2009	183	10	2008	552
11	2000	810	11	2008	229
12	1982	444	12	2016	198
13	2017	275	13	2003	89
14	1994	141	14	2001	188

4.2.2. *Labeling topics*

In this study, we used the LDA model to identify scientific topics and technical topics (see Appendix). We further used the LexPageRank (Erkan & Rade, 2004) method to obtain an automatic summary for each topic. Finally, under the guidance of domain experts, the content of 15 scientific topics and 15 technical topics with clear connotations was obtained. The complete topic connotations and topic feature terms of 15 scientific topics and 15 technical topics, as well as their connotations and key keywords, are attached in the appendix.

4.3. *Topic-linked path of science and technology*

Table 2 shows the topics' distribution years. The number of papers published and patent applications in the GEV field did not keep pace. In 2000, the number of scientific papers was the largest, with 800 publications. In 2008, the number of patents was the highest, with 552 applications.

This paper visualizes the innovation evolution path of the GEV field. Fig. 6 illustrates the overall development trend of candidate linked topics through the evolution paths, which was drawn by D3.js (Bostock, Ogievetsky, & Heer, 2011). D3 (Data-Driven Documents) is a favored data visualization libraries of data scientists. D3 focuses on visualization and interactive presentation and provides a rich set of visualization types. D3 remaps the original data provided by a layout method into a new data format to fit a specific type of chart. D3 has multiple visualization methods. In addition to line diagrams, column diagrams, and pie charts, common visual displays include complex diagram styles such as tree diagrams, force-oriented diagrams, chord diagrams, and word cloud diagrams (Bostock et al., 2011). D3 is not only used to construct big data visualization frameworks (Bao & Chen, 2014) but is also the basis of visualization software in many professional fields, such as visualization of protein sequences (Mukhyala & Masselot, 2014), network visualization (Múrias dos Santos, Cabezas, & Tavares, 2015), and computational biology visualization (Shank, Weaver, & phylotree, 2018). However, the application of D3 requires certain data programming skills and is not suitable for analysts of statistical analysis software.

In the evolution network, S-i (i ∈ [0,14]) represents the science topic, and T-j (j ∈ [0,14]) represents the technology topic. Fig. 6 shows the GEV science and technology innovation path with a threshold of 0.5, which provides a relatively clear evolution path after linkage fusion, in order to demonstrate the similarity between the 15 important scientific topics and the emergence of 15 important technical topics. In Fig. 6, 15 scientific topics and 15 technical topics are evenly arranged on the vertical axis, where the position does not have a quantitative meaning. A continuation path represents the path of scientific and technological innovation. To highlight the interaction between science and technology, we used colored lines for linked topic pairs that are not less than the threshold (0.5). Blue lines indicate where the scientific topic emerges before the technical topic. When the technical topic emerges before the scientific topic, the topics are connected by yellow lines.

Fig. 6 clearly shows that before 2000, the innovative path mostly evolved from scientific topics to technical topics. After 2008, the spread and diffusion of knowledge from technology topics to scientific topics took precedence. This also illustrates the obvious interaction between basic research and applied research in the GEV field. In the early stage (before 2000), there were many major developments (breakthroughs) in the basic research field that began to spread and extended into applied research. In the later period (after 2008), many breakthroughs in applied technology occurred. At the same time, a better theoretical understanding of feedback from applied to basic research began to occur, which is illustrated by the large number of yellow lines.

Combined with Fig. 6 and under the instruction of domain experts, we will give a detailed description of the most important science and technology linked topics and their development points in recent years.

The most linked science and technology topic pair are T-2 (2001) and S-6 (2015), which shows that the applied research of technology promotes the advancement of basic science research. With the development of genetic engineering and molecular biology since 2000, the GEV field has provided new ways to develop safer and more effective vaccines and has become the main method of vaccine production. In the field of medical treatment, in addition to human and animal anti-infectives, cancer therapy has also become a key area for vaccine development. Some oncology research departments have also tried to help stimulate the body's immune response to control it through tumor vaccines because of the ineffectiveness of conventional treatments such as surgery, chemotherapy, and radiotherapy. Vaccines such as tumor DNA vaccines, dendritic cell tumor vaccines, and the like, have been developed. Although the effectiveness of tumor vaccines has achieved good results in animal studies, the results of treatment in clinical studies have not been satisfactory. The general issue is insufficient antigenic protein expression and insufficient transfection efficiency. It has become urgent for tumor immunotherapy to find and screen effective tumor antigens, to activate the cellular and humoral immune responses of the body, to break the body's tolerance to tumors, to more effectively perform immune surveillance, to explore more effective methods for improving the immune microenvironment, and to reduce the inhibition of tumor immune response and tumor vaccine entry into the body, which would further promote the development of related basic research and new technologies.

In the absence of external interference, the number of T cells that recognize tumor cells in the human body is very small. In recent years, scientists have invented adoptive cell immunotherapy that uses genetic engineering to make ordinary T cells capable of recognizing tumor cells. All the above mechanisms will initiate immunity to tumor cells, including activation of immune responses by costimulatory activation of CD28 molecules on the surface of T cells, regulation of T-cell immunosuppression by CD4+, and recognition of the specificity chimeric antigen receptor (CAR) of the tumor antigen by genetic manipulation of T cells. Meanwhile, the receptor plus a model delivery region causes T-cell activation in the intracellular domain; therefore, cell therapy shows good capacity for targeting and killing tumor cells in some blood and tumors. Meanwhile, some significant effects have been produced in patients; for example, some clinical issues such as cytokine release syndrome and its unique side effects have yet to be studied. Research on tumor immunotherapy has also become a hotspot in cancer therapy in recent years.

From the content perspective, S-8 (1996) and T-13 (2003) are the basic research on the mechanism to promote the application. S-8 (1993) mainly involves the study of the mechanism of antigen-presenting cells. After the foreign antigen is taken up and treated by antigen-presenting cells (dendritic cells, etc.), it must be combined with the peptide binding region of major histocompatibility complex (MHC) to form an antigen peptide-MHC molecular complex transporter. It is expressed on the surface of antigen-presenting cells to be recognized by T cells. Scientific research on the mechanism of antigen presentation and the various types of molecules involved contributes to the design of novel vaccines. Improvements in vaccine delivery systems have become a new trend in vaccine applications (T-13). Dendritic cells are the strongest antigen-presenting cells in the body, and they can induce naive T cells into cytotoxic T lymphocytes. Inducing a specific immune response, dendritic cells as a therapeutic vaccine have become a research hotspot.

S-7 (2000) and T-11 (2008) are studies on the preparation method to promote GEV application technology. The application of gene recombination technology to prepare human-mouse chimeric antibody technology is not only widely used in the in vitro diagnosis of diseases but also finds increasing applications in the treatment of diseases. T-cell therapy (CAR-T), which uses genetically engineered technology to express tumor-specific chimeric antigen receptors, is progressing rapidly, binding tumor antigens from T cells that originally expressed chimeric antigen receptors in an antigen-dependent, non-MHC-restricted manner. Initiating and killing the tumor response and to increase the cytotoxicity and proliferative activity of T cells, a costimulatory molecule such as CD28 is added to the chimeric receptor, and the chimeric antigen receptor is modified along with the progress of the cell immunotherapy tumor. The role of T cells in clinical applications is receiving increasing attention.

In general, the GEV field presents a trend that scientific research and technological applications mutually prompt each other. The basic technological development of genetic engineering has enabled GEV to be developed in a variety of disease fields, especially in anti-infectives. In addition, the field of tumors is also a research hotspot of GEV. However, due to the complex tumor types, unclear mechanisms of action, and gene mutation of tumors, there are still many scientific issues in the clinical application of GEV. It is urgent to promote the progress of basic research. Meanwhile, the progress of basic research also provides new ideas for the design of new cancer vaccines.

5. Discussion and conclusions

5.1. Discussion

All the progress in this study will help intelligence analysts and technology managers to understand the development trends of science and technology in a more efficient manner. It provides support for strategic decision-making, innovation resource allocation, and planning for industrial development.

However, this study has certain limitations. The multi-relationship fusion only considered three types of relationships; thus, there is still information missing. Additionally, despite using topic linkage analysis with multi-relationship fusion to strengthen the semantic relevance of the topic terms, the study has not considered the difference between the terms that are synonymous in basic and applied scientific research. This is an important factor in the identification of the interaction of science and technology. The method based on word correlation can reflect the correlation degree between science and technology to some extent; however, the result is not completely reliable because of inconsistencies in the words and concepts of the paper and patent documents.

5.1.1. Similarities and differences of the evolution mechanisms in science and technology

Science and technology are interrelated to a certain degree, and at the same time, the dynamics and mechanisms of their respective developments are fundamentally different. In the next two sections, we focus on the similarities and differences.

5.1.1.1. Similarities of the evolution mechanisms in science and technology. Scientific advancement and technological evolution follow the same logic; they both follow the self-organizing dynamics mechanism with the ordering and logic of knowledge structure. Noyons and Raan (1998) and Raan (2000) stated that the growth of scientific knowledge is largely a self-organizing process within a scientific cognitive system, with a dynamic growth and aging process. Jing, Ma, and Zhang (2009) proposed that science is an unstable, chaotically logical system, where stable and orderly construction can be achieved through logical reorganization. Popper (2005) pointed out that science comes from problems, and scientific theory is a tentative response to problems. He also agreed that scientific development is a chain process from problem to problem; therefore, he introduced the idea of "evolution" into the process of scientific cognition.

Knowledge growth within technology also has its own self-organizing mechanism. Moreover, the evolution of science is not a linear gradually evolving system; sometimes, there is scientific revolution (Kuhn, 1962). Arthur (2009) believes that technological innovations have common abstract attributes and structures and follow three basic principles: combination, circulation, and attachment to natural phenomena. On the one hand, technology is self-organizing and can be gathered through some simple rules. Conversely, technology is self-created because all technologies are derived from existing technologies; that is, any new methods to meet human needs, or to achieve a certain goal that cannot be completed without the existing technologies.

Both the theoretical systems of science and technology are purposeful and follow the same logic, are based on existing research, and present a chain rule. In the process of scientific development, on the one hand, each subject subsystem has its own unique internal dynamic mechanism. This promotes the subject knowledge evolution according to internal logical rules, leading to the emergence of a discipline frontier (Leydesdorff, 2013). On the other hand, there are interaction mechanisms among discipline subsystems. There will be crossovers or integrations of the disciplinary knowledge within various science subsystems.

5.1.1.2. Differences in the evolution mechanisms of science and technology. This consistency does not signify that science and technology are identical. In reality, the development mechanisms of science and technology are quite different because scientific development relies more on reasoning, argumentation, and logical support, while technological development depends on the combination and superposition of prior existing technologies. In addition, scientific research has its own thinking logic, including analysis, synthesis, induction, deduction, analogy, abstraction, and generalization. However, technology is a type of knowledge set about certain practices and activities, and it is not just the application of knowledge from other fields. Sometimes, technological knowledge emerges before scientific knowledge in some fields.

At the macro level, there are two innovative elements in the process of scientific and technological innovation. One spreads from science to technology, and the other from technology to science, which jointly determine the content and direction of technological development. At the micro level, the development of science and technology follows the rules of self-organization and evolves under chain logic. Meanwhile, science and technology are essentially different from each other because of the development mechanisms, motivations, and principles of development and evolution. Scientific research and technological innovation activities continue to improve in their own development trajectories in R&D activities. Therefore, it is necessary to study the development trend of science and technology innovation through the interactive path.

The interrelated and relatively independent relationship between science and technology make it possible to determine how science and technology jointly promote innovation in their interaction. It is necessary to find a method of identifying the scientific and technological innovation path based on the linkage of science and technology. However, there is still a lack of practical ways to find the innovation path under such a relationship.

When analyzing the evolution path of innovative topics based on text analysis, we do not merely mix scientific papers with patent information to form a single evolutionary path of science and technology but rather maintain the relative independence of the two paths and further an analysis of their relationship. Through correlation of the relatively independent and self-contained system, we can maximize the discovery of the interaction of science and technology and then discover new knowledge growth points.

5.2. Conclusions

This study focused on linked science and technology topics by exploring the identification method of innovation paths based on the linkage of science and technology topics. We systematically analyzed the relationship between basic research and applied research to clarify the relationship between science and technology, specifically the general process and principle of the interaction between science and technology. To identify the innovation evolution path from an interactive perspective, we revealed the interaction of science and technology by analyzing the linkage between topics at a micro level. The internal mechanism of knowledge dissemination and the relationship between science and technology were revealed and described in detail by measuring knowledge unit linkage. An experimental study in the area of GEV shows that our method is indeed successful in revealing interactions between science and technology at a micro level. This study makes three major contributions:

First, compared to the analysis on the relevance of scientific and technical topics with the use of non-patent literature, in our study, text analysis based on topic linkage could supply more information and make the calculation of scientific relevance more accurate by making up for the shortages of patent citation analysis. In contrast to a single path of scientific and technological innovation, our method provides advantages in demonstrating the co-evolution characteristics of technological innovation, thus eliminating the current one-sidedness of using only one innovative element in the innovation path discovery. The topic linkage of science and technology, by content linkage, could further eliminate the limitation of depending on numerical statistics.

Second, topic identification and their similarity calculations using texts are the key points in the study of linkage of innovation topics between science and technology. Using a single relationship, it is difficult to obtain meaningful topics; therefore, we adopted a text topic calculation method with integration of multi-relationships. This method can largely eliminate insufficient semantics in text topic analysis and effectively identify the science and technology-related topics, in addition to the contents of important topics in the evolution process. An experimental study showed that multi-source heterogeneous data fusion is an efficient way to measure the linkage of innovation topics between science and technology.

Third, the inflection point in the life cycle curve is selected as the most influential time point on the innovation path. In contrast, the average time for publication will cause a lagging trend for the topic distribution time, and the earliest publication year, or application year, may indicate only sporadic research and not any certain influence. The inflection point can display the topic distribution time at an early and particularly influential time point.

Our method can be used as a tool to gain a better understanding of the linkage between scientific discoveries and technological developments in different innovation paths and to discover evolution trends in these links. In this context, the method can also be used as a support tool for strategic decision-making, the allocation of innovation resources, and industrial development planning. It is possible to discover new unknown technology pathways that are supported by scientific discoveries.

5.3. Future work

In the future, we plan to improve and extend the method by adding full-text information to analyze the main points of innovation paths of science-supported technological developments. To cope with the differences in the granularity of topic terms, we will include more effective algorithms for heterogeneous network analysis in the method. Currently, there are numerous powerful algorithms for heterogeneous network analysis in the field of machine learning. This allows a greater semantic relevance for the identification of links between science and technology. In addition to the already used parameters (topic terms, authors, and references), we will use parameters such as the organizations to which authors are affiliated, the scientific journals in which the scientific discoveries are published, and the scientific disciplines to which the discovery belongs. In addition, link prediction analysis will be applied to further enhance the semantic linkage of topic terms between science and technology. It is thought that link prediction can provide alignment to the synonym phenomenon of scientific and technical terms; therefore, it can erase inconsistencies in the words and concepts shared between basic and applied scientific research.

Additionally, it is significant to recognize the causal linkage between topics in science and technology along the time axis, which will greatly improve the prediction accuracy of the innovation paths. For the scientific and technical topics on topic-linked innovation paths, most research only considers the timeline and content similarity while ignoring the causal linkage between them. In the future, we will implement temporary network analysis to find such cause-effect links between scientific discoveries and technological developments.

The emphasis will be on revealing patterns in the development, transition, and evolution of science and technology and discovering potential new knowledge growth points. Particular attention will be paid to the time span from the innovation path from scientific discovery to technological application. We will also use this method to conduct research into evolution paths from science to technology, which then go back to science.

Author contribution

Haiyun Xu: Conceived and designed the analysis, Collected the data, Contributed data or analysis tool, Performed the analysis, Wrote the paper.

Jos Winnink: Conceived and designed the analysis, Performed the analysis, Wrote the paper.

Zenghui Yue: Collected the data, Contributed data or analysis tool, Performed the analysis.

Ziqiang Liu: Collected the data, Contributed data or analysis tool, Performed the analysis.

Guoting Yuan: Collected the data, Contributed data or analysis tool, Performed the analysi.

Acknowledgments

This article is an outcome of the project "Study on the Recognition Method of Innovative Evolving Trajectory based on Topic Correlation Analysis of Science and Technology" (No. 71704170) supported by National Natural Science Foundation of China, and the project "Study on Regularity and Dynamics of Knowledge Diffusion among Scientific Disciplines" (No.

71704063) supported by National Natural Science Foundation of China and the Youth Innovation Promotion Association, CAS (Grant No. 2016159).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.joi.2020.101014.

References

Albert, T. (2016). *Measuring technology maturity: Operationalizing information from patents, scientific publications, and the web*. Springer.

Andersen, B. (1999). The hunt for S-shaped growth paths in technological innovation: a patent study. *Journal of Evolutionary Economics*, *9*(4), 487–526. http://dx.doi.org/10.1007/s001910050093

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.

Bao, F., & Chen, J. (2014). Visual framework for big data in d3. Js. In *2014 IEEE Workshop on Electronics, Computer and Applications* (pp. 47–50).

Bassecoulard, E., & Zitt, M. (2004). *Patents and publications*. Netherlands: Springer.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Paper Presented at the International Conference*,

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2008). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent dirichlet allocation*. JMLR.org.

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309.

Bush, V. (1945). *Science, the endless frontier: A report to the President*. US Govt. print. off.

Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., & Thijs, B. (2006). Traces of prior art: An analysis of non-patent references found in patent documents. *Scientometrics*, *69*(1), 3–20.

Chen, C. (2001). *A new introduction to natural dialectics*. Shenyang: Northeastern University press (in Chinese).

Chen, L., Yang, G., Zhang, J., & Fan, Y. (2015). Research on multiple main paths method oriented to analysis of technology evolution. *Journal of Library & Information Services in Distance Learning*, *59*(10), 124–130, 115. (in Chinese).

Coulter, N., Monarch, I., & Konda, S. (1998). *Software engineering as seen through its research literature: A study in co-word analysis*. John Wiley & Sons, Inc.

Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z., et al. (2011). TextFlow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2412–2421. http://dx.doi.org/10.1109/TVCG.2011.239

Dong, K., Xu, H., Luo, R., Wang, C., & Fang, S. (2018). A review of the research on the relationship between science and technology. *Journal of The China Society for Scientific and Technical Information*, *37*(6), 642–652 (in Chinese).

Erkan, G., & Rade, D. R. (2004). LexPageRank: Prestige in multi-document text summarization. *EMNLP*, *4*.

Gao, J. P., Ding, K., Teng, L., & Pang, J. (2012). Hybrid documents co-citation analysis: Making sense of the interaction between science and technology in technology diffusion. *Scientometrics*, *93*(2), 459–471.

Gardner, P. L. (1999). The representation of science-technology relationships in Canadian physics textbooks. *International Journal of Science Education*, *21*(3), 329–347.

Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of’ reverse’ citation relations. *Scientometrics*, *58*(2), 415–428.

Gläser, J., Glänzel, W., & Scharnhorst, A. (2017). Same data—Different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics*, *111*(2), 981–998.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5228–5235.

Guan, J., & He, Y. (2007). Patent-bibliometric analysis on the Chinese science — Technology linkages. *Scientometrics*, *72*(3), 403–425.

Huang, M. H., Yang, H. W., & Chen, D. Z. (2015). Increasing science and technology linkage in fuel cells: A cross citation analysis of papers and patents. *Journal of Informetrics*, *9*(2), 237–249.

Hummon, N. P., & Dereian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, *11*(1), 39–63.

Jensen, S., Liu, X., Yu, Y., & Milojevic, S. (2016). Generation of topic evolution trees from heterogeneous bibliographic networks. *Journal of Informetrics*, *10*(2), 606–621.

Jing, J., Ma, F., & Zhang, X. (2009). *Information science*. Beijing: science press (in Chinese).

Kang, Y. (2016). Heterogeneous knowledge flow network measurement based on perspective of intergration innovation. *Journal of the China Society for scientific and technical information*, *35*(9), 963–970 (in Chinese).

Kang, Y. (2017). Analysis of technology opportunity based on c̈oupling and co-citationḧybrid network. *Journal of the China Society for scientific and technical information*, *36*(2), 170–179 (in Chinese).

Kim, Y. G., Suh, J. H., & Sang, C. P. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Applications*, *34*(3), 1804–1812.

Klavans, R., & Boyack, K. (2017). Research portfolio analysis and topic prominence. *Journal of Informetrics*, *11*, 1158–1174. http://dx.doi.org/10.1016/j.joi.2017.10.002

Kostoff, R. N., & Schaller, R. R. (2001). Science and technology roadmaps. *Engineering Management IEEE Transactions on*, *48*(2), 132–143.

Kuhn, T. S. (1962). *The Structure of scientific revolutions*. Chicago: The University of Chicago Press.

Kwon, S., Porter, A., & Youtie, J. (2016). Navigating the innovation trajectories of technology by combining specialization score analyses for publications and patents: Graphene and nano-enabled drug delivery. *Scientometrics*, *106*(3), 1057–1071.

Lai, R. J., & Li, M. F. (2015). Technology evolution of lower extremity exoskeleton from the patent perspective. *Key Engineering Materials*, *625*, 536–541.

Lathabai, H. H., George, S., Prabhakaran, T., & Changat, M. (2018). An integrated approach to path analysis for weighted citation networks. *Scientometrics*, *117*(3), 1871–1904.

Lee, S., Kang, S., Park, Y. S., et al. (2007). Technology roadmapping for R&D planning: The case of the Korean parts and materials industry. *Technovation*, *27*(8), 433–445.

Letaba, P., Pretorius, M. W., & Pretorius, L. (2015). Analysis of the intellectual structure and evolution of technology roadmapping literature. *Portland International Conference on Management of Engineering and Technology*, 2248–2254.

Leydesdorff, L. (2013). *Statistics for the Dynamic Analysis of Scientometric Data: The evolution of the sciences in terms of trajectories and regimes*. Springer-Verlag New York, Inc.

Li, R. (2011). *A study on optimizing the model for exploring linkage between science and technology based on patent citation analysis*. Beijing: Graduate School of Chinese Academy of Sciences., in Chinese.

Li, R., Chambers, T., Ding, Y., Zhang, G., & Meng, L. (2014). Patent citation analysis: Calculating science linkage based on citing motivation. *Journal of the Association for Information Science and Technology*, *65*(5), 1007–1017.

Liu, J. S., & Lu, L. Y. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the Association for Information Science and Technology*, *63*(3), 528–542.

Liu, Z., Wang, X., & Bai, R. (2016). Research on visualization analysis method of discipline topics evolution from the perspective of multi-dimensions: A case study of the big data in the field of library and information science in china. *Journal of Library Science in China*, *42*(226), 67–83 (in Chinese).

Lu, L. Y. Y., & Liu, J. S. (2014). A survey of intellectual property rights literature from 1971 to 2012: The main path analysis. *Paper Presented at the Portland International Conference on Management of Engineering & Technology*.

Martinelli, A. (2012). An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry. *Research Policy*, *41*(2), 414–429.

Martinelli, A., & Nomaler, Ö. (2014). Measuring knowledge persistence: A genetic approach to patent citation networks. *Journal of Evolutionary Economics*, *24*(3), 623–652.

Mcauliffe, J. D., & Blei, D. M. (2008). Supervised topic models. In *Advances in neural information processing systems*. pp. 121–128.

McCalman, P. (2001). Reaping what you sow: An empirical analysis of international patent harmonization. *Journal of International Economics*, *55*(1), 161–186.

Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy*, *29*(3), 409–434.

Meyer, M. (2002). Tracing knowledge flows in innovation systems, Scientometrics, 54(2): 193-212. *Scientometrics*, *54*(2), 193–212.

Meyer, M., & Debackere, K. (2010). *Can applied science be 'good science'? Exploring the relationship between patent citations and citation impact in nanoscience*. New York, Inc: Springer-Verlag.

Michel, J., & Bettels, B. (2001). Patent citation analysis: A closer look at the basic input data from patent search reports. *Scientometrics*, *51*(1), 185–201.

Morris, S. A., Yen, G., Wu, Z., & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the Association for Information Science and Technology*, *54*(5), 413–422.

Mukhyala, K., & Masselot, A. (2014). Visualization of protein sequence features using JavaScript and SVG with pViz. *Bioinformatics*, *30*(23), 3408–3409.

Múrias dos Santos, A., Cabezas, M. P., Tavares, A. I., et al. (2015). tcsBU: A tool to extend TCS network layout and visualization. *Bioinformatics*, *32*(4), 627–628.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Research Policy*, *31*(8–9), 1389–1403.

Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics*, *7*(3-6), 369–381.

Noyons, E. C. M., & Raan, A. F. J. V. (1998). Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research. *Journal of the Association for Information Science and Technology*, *49*(1), 68–81.

Park, H., & Magee, C. L. (2017). Tracing technological development trajectories: A genetic knowledge persistence-based main path approach. *PloS one*, *12*(1).

Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis. *Journal of Operations Management*, *27*(3), 185–202.

Popper, K. R. (2005). The logic of scientific discovery. *Yinshan Academic Journal*, *12*(11), 53–54.

Raan, A. F. J. V. (2000). On growth, ageing, and fractal differentiation of science. *Scientometrics*, *47*(2), 347–362.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., et al. (2004). The author-topic model for authors and documents. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, 487–494.

Rosvall, M., & Bergstrom, C. (2008). Maps of random walks on complex networks reveal community structure. *Proceeding of the National Academy of Sciences of the United States of America*, *105*(4), 1118–1123.

Schmoch, U. (1993). Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics*, *26*(1), 193–211.

Shank, S. D., Weaver, S., & phylotree, P. S. L. K. (2018). js-a JavaScript library for application development and interactive data visualization in phylogenetics. *BMC Bioinformatics*, *19*(1), 276.

Small, H., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research policy*, *43*(8), 1450–1467.

Stokes, D. E. (1997). *Pasteur's quadrant*. Brookings Institution Press.

Tijssen, R. J. W. (2010). Discarding the 'basic science/applied science' dichotomy: A knowledge utilization triangle classification system of research journals. *Journal of the American Society for Information Science and Technology*, *61*(9), 1842–1852. And.

Tijssen, R. J. W., & Winnink, J. (2016). Twenty-first century macro-trends in the institutional fabric of science: Bibliometric monitoring and analysis. *Scientometrics*, *109*(3), 2181–2194.

Tu, Y. N., & Hsu, S. L. (2016). Constructing conceptual trajectory maps to trace the development of research fields. *Journal of the Association for Information Science and Technology*, *67*(8), 2016–2031.

Van Looy, B., Magerman, T., & Debackere, K. (2007). Developing technology in the vicinity of science: An examination of the relationship between science intensity (of patents) and technological productivity within the field of biotechnology. *Scientometrics*, *70*(2), 441–458.

Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics*, *54*(3), 399–420.

Vishnevskiy, K., Grebenyuk, A., Kindras, A., et al. (2011). Integration of roadmapping and scenario planning for implementing science, technology and innovation strategic priorities - the case of Russia. *International Journal of Foresight and Innovation Policy*, *10*(2/3/4), 126.

Waltman, L., Boyack, K., Colavizza, G., & Van Eck, N. (2017). A principled methodology for comparing relatedness measures for clustering publications. *Paper Presented at the Proceedings of the 16th International Conference on Scientometrics and Informetrics (ISSI 2017)*.

Wang, Z. Y., Li, G., Li, C. Y., & Li, A. (2012). Research on the semantic-based co-word analysis. *Scientometrics*, *90*(3), 855–875.

Wei, X., & Croft, W. B. (2006). LDA-based document models for ad-hoc retrieval. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 178–185.

Wei, L., Xu, H., Hu, Z., Dong, K., Wang, C., & Pang, H. (2016). Multiple-pattern analysis and prediction of topic evolution path based on topic correlation: A case study of information science research. *Journal of Library & Information Services in Distance Learning*, *60*(3), 71–81 (in Chinese).

Weng, J., Lim, E. P., Jiang, J., et al. (2010). Twitterrank: Finding topic-sensitive influential twitterers. *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, 261–270.

Wu, C. C. (2016). Constructing a weighted keyword-based patent network approach to identify technological trends and evolution in a field of green energy: A case of biofuels. *Quality & Quantity*, *50*(1), 213–235.

Xu, H.-Y., Yue, Z.-H., Wang, C., Dong, K., Pang, H.-S., & Han, Z. (2017). Multi-source data fusion study in scientometrics. *Scientometrics*, *111*(2), 773–792. http://dx.doi.org/10.1007/s11192-017-2290-5

Xu, S., Zhu, L., Qiao, X., Shi, Q., & Gui, J. (2012). Topic linkages between papers and patents. *Proceedings of the 4th International Conference on Advanced Science and Technology*, 176–183.

Xu, W., Luo, D., & Li. (2019). Topic identification based on multi-semantic relation fusion. *Journal of Library Science in China*, *45*(1), 82–94 (in Chinese).

Xu, S., Zhai, D., Wang, F., An, X., Pang, H., & Sun, Y. (2019). A novel method for topic linkages between scientific publications and patents. *Journal of the Association for Information Science and Technology*, *70*(9), 1026–1042.

Zhang, Y. (1998). Theory of three main sociological differences between science and technology. *Social Science in Nanjing*, (8) (in Chinese).

Zhang, J. (2014). *Research on the linkage between science and technology*. Nanjing: Nanjing Agricultural University (in Chinese).

Zhang, Y., Guo, Y., Wang, X., Zhu, D., & Porter, A. L. (2013). A hybrid visualisation model for technology roadmapping: Bibliometrics, qualitative methodology and empirical study. *Technology Analysis and Strategic Management*, *25*(6), 707–724.

Zhang, Y., Zhang, G., Chen, H., Porter, A. L., Zhu, D., & Lu, J. (2016). Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological Forecasting and Social Change*, *105*, 179–191.

Zhu, Q. (2014). *Research on semantic enrichment of citation analysis method and application experiments*. University of Chinese Academy of Sciences Beijing (in Chinese).

Zhu, M., Zhang, X., & Wang, H. (2017). A LDA based model for topic evolution: Evidence from information science journals. *Paper Presented at the International Conference on Modeling, Simulation and Optimization Technologies and Applications*.