Journal of Informetrics 16 (2022) 101265

Contents lists available at ScienceDirect

Journal of Informetrics

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

Integrative model for discovering linked topics in science and technology

Haiyun Xu[a,*], Zenghui Yue[b], Hongshen Pang[c], Ehsan Elahi[d], Jing Li[e], Lu Wang[c]

[a] Business School, Shandong University of Technology, Zibo, Shandong 255049, China
[b] School of Medical Information Engineering, Jining Medical University, Rizhao, Shandong 276826, China
[c] Library, Shenzhen University, Shenzhen, Guangdong 518060, China
[d] School of Economics, Shandong University of Technology, Zibo, Shandong 255049, China
[e] Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041, China

ARTICLE INFO

Keywords:
Linked topics in science and technology
Scientific innovation
Science and technology linkage
Link prediction
Topic recognition

ABSTRACT

Linked topics in science and technology (LTSTs) can provide new avenues for technological innovation and are a key step in the transition from basic to applied research. This paper proposes a science and technology semantic linkage integration model for discovering LTSTs. Particularly, the integrative model fuses the term co-occurrence networks of basic and applied research, which expands the completeness of topic networks by enhancing the semantic characteristics of these networks. It is found that link prediction can further reinforce the semantic association of topic terms in networks between basic and applied topics. Simple fusion explicitly linked the topic terms, which can be used as automatic seed marking for subsequent link prediction to identify implicit linking of topic terms. Furthermore, an application to the gene-engineered vaccines field depicted that newly predicted implicit relations can effectively identify LTSTs. The results also show that implicit semantic recognition of LTSTs can be enhanced through simple fusion, while the recognition of LTST can be improved through link prediction. Therefore, the proposed model can assist experts to identify LTSTs that cannot be recognized through simple fusion.

1. Introduction

Scientific discovery calls upon the analysis of objective patterns in nature to produce reliable predictive models. Technological innovation involves the application of technology to devise new methods, tools, and processes of scientific discovery. Further to this dichotomic classification, Bush (1945) stated that the novel scientific discoveries could be extended beyond science and are critical for technological innovations and products. While after seven decades, Narayanamurti and Odumosu (2016) demonstrated that the “creation of new knowledge” (discovery/basic research) and the “accumulation and creation of knowledge that results in a new tool” (invention/applied research) are intricately connected, constituting a cycle. Research shows that the transformation of scientific discoveries into new technological innovations takes place in complex and indirect ways (Eulaerts et al., 2020). National innovation now changed into a systemic pattern with diverse sources apart from science discovery, therefore, cooperation between universities and industry is relevant (Chaminade & Lundvall, 2019). Currently, in China, investment in basic research has been increased in line with the policy supporting the development of basic research. China has formulated the “13th Five-Year” National Basic Research Special Plan (Ministry of Science and Technology, 2021) and during this period, China’s basic research funding nearly doubled (Ministry of Science and Technology et al., 2017). An increasing synergism in

* Corresponding author.
E-mail address: xuhaiyunnemo@gmail.com (H. Xu).

https://doi.org/10.1016/j.joi.2022.101265
Received 26 May 2021; Received in revised form 11 February 2022; Accepted 16 February 2022
Available online 2 March 2022

science and technology jointly determines the evolutionary direction of scientific innovation (Xu et al., 2020). The interaction of science and technology is now a defining characteristic to promote advancements in scientific and technological innovations through collaboration and sharing (Xu et al., 2017). In fact, science and technology jointly contribute to the continuous evolution of content and the direction of scientific developments. Therefore, it is imperative to study the methods to identify paths for innovation through linked topics in science and technology (LTSTs) (Dong et al., 2018; Kostoff & Schaller, 2001; Zhang et al., 2013).

LTSTs are at the interface of science and technology topic linkage and can be categorized into two types: Science–technology (S-T) and Technology–science (T-S) linkages. The S-T refers to technological breakthroughs that follow the scientific discovery. While the T-S refers to the scientific discoveries generated by technological innovation. A LTST usually refers to the convergence of cutting-edge emerging technological innovation with breakthroughs in the advancement of science and technology in a particular field (Stirling, 2007). A variety of published studies is available on the linkage between research articles and patents for getting technological opportunities (Albert, 2016).

Citation link analysis is the basic method to identify LTSTs; however, it has several limitations, including challenges related to incomplete information. Moreover, the link between scientific publications and patents based on citation is vulnerable due to differences in citation behavior (Li et al., 2014). Therefore, text mining can be used that provide more insightful information through content similarity linkage, particularly by validating topic linkage analysis when citations are missing in a patent. Particularly, the text mining analysis can represent certain implicit relation between patent text and research papers, which is not possible through citation analysis.

However, using text mining analysis, an insufficient mining of latent semantic relations between topic terms is the primary challenge for recognition of LTST. A solution of sufficient matching scientific terms with technological terms is imperative to identify LTSTs. The current study aims to design an analytical method that can better reveal the implicit semantic associations between scientific discoveries and technological innovations. To achieve this goal, we propose an integrated model for discovering LTSTs at a micro level using text mining analysis. The integrative model fuses the term co-occurrence networks of basic and applied research, in which the topic network is expanded, and the semantic association of the network is enhanced. Link prediction is used for further mining the latent semantic association between topic terms in science and technology topic networks. It helps to eliminate inconsistencies in the words and concepts shared between basic and applied scientific research.

The remaining article is structured as follows: Section 2 discusses related studies on topic recognition, the semantic association of science and technology, and the method of link prediction Section 3. explains the concept of LTST and a method for identifying the implicit semantic linkages of science and technology using link prediction Section 4. discusses GEVs as an example to verify the effectiveness of the proposed method. Finally, Section 5 concludes the main findings and provides the directions for future studies.

2. Literature review of related work

In this section, we discuss the relative concepts and theoretical basis of LTSTs and introduce the link prediction that will be used to propose an integrated model to detect LTSTs.

2.1. Scientific innovation and innovation topics

Currently, several scientific forecasting methods are targeting key areas at national levels that are effective. However, it is difficult to obtain satisfactory results of forecasting using quantitative analytical methods because of intensified segmentation, interdisciplinary and technological integration. From the perspective of scientific evolution, major innovations originated from micro-topics. Therefore, in addition to discipline itself, the domain topic foresight at the mesoscopic/micro-level is of great significance to scientific planning at the macro level. In research about topic classification on text corpus, “innovation topics” are usually described as “topic identification,” “topic finding,” “label identification,” “cluster labeling,” and “category labeling” (Stein & Zu Eissen, 2004). Following the Zhang, Huang, Porter et al. (2019), we preferred two definitions of “innovation topics” among several descriptive definitions. Firstly, “a topic is a collection of articles that are mathematically represented by their centroid, which is identified as the article sharing the highest similarity with all other articles on the topic”. Secondly, “a topic is geometrically represented as a circle, and its boundary is the largest Euclidean distance between its centroid and all articles (p.6)”. Historically, innovation topics have been identified using qualitative methods, such as summarized by domain experts. However, under the circumstance of the huge volumes of text data to be processed, expert consultation is nearly impossible, and quantitative methods become essential.

According to the objective and purpose of the researchScience and technology, as abstract concepts can be divided into “basic research,” “applied research,” “application-oriented basic research,” and “basic-oriented application research. The “Basic research” primary concerns research to increase the understanding to nature, while “applied research” applies scientific knowledge to address technological issues (Tijssen, 2010; Tijssen & Winnink, 2016). The concepts of “application-oriented basic research” and “basic-oriented application research” connect “basic research” and “applied research”, and blur the boundary between science and technology (Xu et al., 2020).

2.2. Theoretical basis

Since the 1940s, the relationship between science and technology, and basic research and applied research have been discussed widely. The National Science Foundation (NSF) is a pioneer in exploring the relationship between science and technology. In *Science, the Endless Frontier*, Bush (1945), presented a linear process innovation model and assumed progress in basic science is the "main source" of applied research, basic science achievement will automatically be transformed into economic activity. In fact, sustainable and strong competitiveness requires a strong interaction between basic research and applied research. To overcome the challenges posed by this linear model, in 2020, the U.S. Senate expanded the NSF and strengthened the U.S. science and technology research ecosystem with the promulgation of the Endless Frontier Act (Baltimore et al., 2021; Conn, 2021).

The Organization for Economic Co-operation and Development (OECD) and the European Union have also elaborated on the relationship between science and technology. Frascati Manual reported that the basic and applied research are not contradictory (OECD, 1970). Basic research will cater to the needs of the funding agent but probably does not consider practical applications. Similarly, European Commission (2005) found that the historical difference between basic research and applied research was no longer relevant as the emerging fields of science and technology often contain substantive content of both. The term "frontier research" was adopted to replace "basic research" in reflection of a new scenario, where frontier research creates new knowledge while also generating potentially useful knowledge.

Many scholars have conducted in-depth research on this issue. According to Pasteur's Quadrant Mode (Stokes, 1997), there is no clear boundary between basic and applied research. Guan et al. (2007) described the close connection of science and technology using the double-helix model as a metaphor. Moreover, Narayanamurti and Odumosu (2016) depicted that scientific research is comprised of both discovery and invention, which are equally important components of a dynamic cycle that inevitably draw upon one another.

Analysis of the relationship between science and technology continues without consensus. Some researchers or institutions maintain that science is the origin of theories and knowledge of all technology, and the source of technology, industrial and social development (Tao et al., 2019). But more researchers maintain that science and technology are interactive and are both critical; if either is missing, it would be hard to effectively support economic and social development. Xu et al. (2020) asserted that science and technology are distinct but interrelated, synergistic but not equivalent, and converge into one integrated driving force in the process of promoting innovation, rather than acting independently of each other. Science and technology (S&T), the two foremost forces that jointly determine the progress of innovations, are viewed as closely connected, interacting, and interdependent systems (Wang & Guan, 2011; Xu et al. 2020).

The interactive features of science and technology showed an intensifying trend. Therefore, many emerging research topics can be identified by discovering LTSTs (Winnink, 2017). At the macro level, the interaction between science and technology is bidirectional, particularly, the interaction determines the direction of scientific evolution (Xu et al., 2020). At the micro-level, science and technology are essentially different because the mechanism of dynamic evolution, motivations, and principles differ (Liu et al., 2019).

Scientific literature, conference contributions, and patents are the most important contributors to scientific progress. Patents are the main source of technological innovation and the way scientific discoveries translate into technological innovations (Ernst, 1999; Fabry et al., 2006; Schmookler, 1972; Niemann et al., 2017). The research articles contain results of research, and the patent publications represent technological inventions that can be turned into practical applications (Kwon et al., 2016; Verbeek et al., 2002). Sufficient and detailed information on technological activities can be observed through patent bibliometric analysis (Jaffe et al., 1998; Schmookler, 1950, 1953). Furthermore, the general trend of basic research can be acquired through paper bibliometric analysis (Tijssen, 2004). Research topics contained in research articles and patents exhibit individual explicit and implicit connections that are demonstrated through time processes, such as topic linkage and evolution (Liu et al., 2019).

A focus on LTSTs is imperative to promote economic and social development (Zhang, 2014). Therefore, it is necessary to identify LTSTs in specific research domains. Particularly, the methods that can improve the accuracy of LTST recognition should have to be focused. Unfortunately, continued theoretical debate has not directly improved practical means of LTST identification. Methods to improve the accuracy of LTST recognition, especially at the micro level, are being increasingly explored.

Science and technology are advancing rapidly, and large number of scientific and technological achievements have emerged. Expert consultation is time-consuming and labor-intensive. To a certain extent, quantitative and automatic analysis methods represented by text mining can help to identify LTSTs. However, current methods using text analysis remain unsatisfactory because the terminology used in scientific papers and patents applications differ remarkably. In particular, the patent text contains a large amount of jargon (Chen et al, 2020), which makes it difficult to identify the semantic association between research articles and the application of patents. Semantic patent analysis provides a lot of inspiration for research (Gerken & Moehrle 2012; Preschitschek et al., 2013; Brachtendorf et al., 2020).

The current article focuses on the science and technology semantic linkage integration model for the discovery of LTSTs. The integration model fuses the term co-occurrence networks of basic and applied research, which expands the completeness of topic networks through enhancing the semantic characteristics of these networks. The link prediction can further reinforce the semantic association of topic terms between basic and applied topics.

2.3. Methods

The relation between scientific discoveries and technological innovative topics is described in the scholarly literature and patent publications which is based on content correlation. Science and technology can be associated through the fusion of scientific and technological literature resources, citation linkage between patents and research articles, and similarity in topics between patents and research articles (Xu et al., 2020).

2.3.1. LTSTs by integration of scientific and technological literature

Linkages between scientific and technological topics can be established through bibliometric connections between scientific and technological literature. This method can provide a unified way of representation and query to enable the interconnection and interoperability of heterogeneous data resources (Duan & Li, 2007). For example, with the integration of the Online Public Access Catalogue (library online catalog), we can perform a comprehensive search of scientific and technological literature. However, this model cannot effectively distinguish between scientific and technological topics. Bassecoulard and Zitt (2004) established a correspondence between patent classification and scientific categories to connect science and technology between research articles and patents. Another study classified the integrated repository of journal articles and patent literature resources into two levels: domain and topic (Lai & Zeng, 2010). Therefore, at the meso- and macro-level, this classification system is time-efficient; however, this kind of analysis is not very effective at the micro-level. At the domain level, the category mapping by establishing semantic associations between terms in the Chinese Library Classification and concepts appearing in the International Patent Classification was performed. At the topic level, classification was performed by selecting thematic vocabulary to index journal articles and patent documents that evaluate similarities in the literature to establish a correspondence between the sources.

The LTST analysis of integrated scientific and technological literature resources usually relies on existing systems of classification and organization, such as the subject classification in Web of Science (WoS) and International Patent Classification (IPC) systems. In previous literature, the identification of science and technology linkages by matching IPC-ISI classification has been focused (Verbeek et al., 2002; Han & Magee, 2018). IPC is widely used to organize patents, while WoS classification is used to categorize research articles. However, matching IPC with WoS classification can obtain a rough scientific and technological linkage because of coarse granularity and slow updates. IPC adopts the classification principle that focuses on functionality and application rather than technical features. The granularity of WoS classification is in journals level rather than paper level. Therefore, WoS classification can effectively organize journals; however, it cannot get accurate subject classification.

The knowledge organization system is effective for describing and organizing literature resources, including their content and interrelationships. The classification and thesaurus are representative tools for concept and terminology organization. In particular, domain ontology can accurately match and integrate heterogeneous literature. However, because sophisticated and practical domain ontology is rare, the performance of this approach is unsatisfactory. Besides, the analysis of micro-level science and technology-related topics that relies on existing classification and organization systems is inadequate because cutting-edge emerging knowledge must be summarized appropriately in a short time.

2.3.2. LTSTs identified based on citation link analysis

In recent decades, the use of patent citation data in social science research has increased dramatically (Jaffe et al., 2019). Mutual citation between research articles and patents are essential for scientific dissemination (Glänzel & Meyer, 2003; Meyer & Debackere, 2010; Murray, 2002; Narin & Noma, 1985; Schmoch, 1993; Roach & Cohen, 2013; Hu & Rousseau, 2018; Du et al., 2019). In previous studies, the linkage between scientific research articles and patents has mainly focused on analyzing the non-patent literature references (NPRs), where the patent assignee cites scientific literature. Owing to the deficiency of one-way analysis, some researchers used a two-way citation method to study the relationship between science and technology (McCalman, 2001; Gao et al., 2012; Huang et al., 2015). The two-way citation linkages involved both research articles that cited patents and the patents cited in the research articles. Moreover, it is an important method to analyze the flow and spread of knowledge between science and technology. The citing of research articles in patents represents the ability of basic science to convert into applied science. The citing of patents in research articles represents ways in which applied science has the ability to support basic science (Xu et al. 2021).

Scientific and technological resources can be integrated through citation linkage between different knowledge systems (Leydesdorff, 2014; Morillo et al., 2001; Rinia et al., 2002; Steele & Stier, 2000). The integration is represented in the Web of Knowledge citation information resource platform established by ISI, Cite Seer developed by the NEC Research Institute of Princeton University in the United States, and SciFinder launched by the American Chemical Abstracts Service. These platforms use citations, co-citations, and other linkages for the integration of heterogeneous resources (Lai & Zeng, 2010). The Nature Index published by the Nature Publishing Group establishes the Lens Metric, which corresponds to the correlation between more than 100 million patent documents and more than one million academic papers in the Lens database. It allows a real-time insight into how science is shaping patent-based invention (LENS.ORG, 2019). However, LTST discovery based on citation relations has a limitation. The citation linkage between research articles and patents has a low coverage rate as it does not completely reflect the relation between the topics of the research articles and patents, and significant differences can exist in different fields (Xu et al., 2020).

Many issues arise when NPRs are applied to study linkages in science and technology. There is not a specific standard to represent NPRs and there are also differences between patent legislations to cite NPRs (Michel & Bettels, 2001). No more than 40% of patent publications include NPRs (Callaert et al., 2006), and there is no guarantee of the relation of content (Meyer, 2000). Furthermore, research articles and patents have considerable citation lags that may hinder future studies on cutting-edge topics. The motivation to publish research articles and patent applications is complex, and the citation relation between them will cause bias in science–technology topic linkage.

2.3.3. LTST identification by bridging researchers

Bridging researchers playing the dual role of article authors and patent inventors constitute an explicit link between science discovery and technology innovation (Cassiman, Glenisson, & Looy, 2007; Forti, Sobrero, & Franzoni, 2007; Boyack & Klavans, 2008; Noyons, Raan, Grupp, & Schmoch, 1994). These bridging researchers, referred to as "core scientists" (Furukawa & Goto, 2006) or "Pasteur scientists" (Stokes, 1997), play a pivotal role to overcome barriers between basic research, applied research, and technological development (Winnink, 2017). Moreover, these bridging researchers are important in moving scientific breakthroughs toward science-based technology innovations (Balconi, Breschi, & Lissoni, 2004; Lissoni, 2010; Meyer, 2006; Noyons, Van Raan, Grupp, & Schmoch, 1994; Tijssen, 2002; Zucker & Darby, 1996).

Forti et al. (2007) analyzed the structural patterns of the relationship and used a network analysis technique to characterize the network of authors of research articles and patents inventors. Wang and Guan (2011) measured the interactions in science and technology in nanotechnology field using author-inventor links. Zhang, Liu, & Wei, et al. (2019) investigated the bridging roles of authors of research articles and patents in knowledge diffusion networks through identifying bridging researchers and ranking their importance in the network. Luo et al. (2019) selected important researchers through the number of achievements, such as high betweenness centrality, invisible college, and the researchers who bridge science discovery and technical innovation, and thus identified the LTST. However, identifying bridging researchers is labor-intensive work, which heavily depends on the accuracy of name alignment between a large number of authors of research articles and patent inventors (Ranaei, Suominen, & Dedehayir, 2017; Luo et al., 2019).

2.3.4. LTST identification by the similarity of keywords

An in-depth analysis of the knowledge system considering the relation between knowledge units is imperative to discover the internal composition of scientific and technological knowledge and the disclosure of knowledge flow. Many studies have combined scientific and technological research topics to enable the identification of emerging and cutting-edge research. For instance, Ren et al. (2015) proposed a knowledge mining model based on the combination of research articles and patents, and identified emerging technologies from the joint research direction of research articles and patents. Xu et al. (2016) selected research topics based on keyword co-occurrence recognition in research articles and patents and mapped the similarity between the two to obtain the research frontier. Liu et al. (2019) constructed a multi-relational integrated keyword co-occurrence matrix and identified research topics in research articles and patents based on community detection algorithms, and used topic evolution visualization to draw a scientific topic evolution map. They analyzed the internal mechanism of science and technology interaction in micro granularity and identified science and technology interaction patterns from the perspectives of quantification, automation, and visualization.

Xu et al. (2019) proposed the CCorrLDA2 model to extract latent topics from the text of research articles and patents with LTST discovery converted into the optimal transportation problem. However, these studies did not probe the dynamic process of the linkage. Xu et al. (2020) used lexical linkage analysis to discover LTSTs, and further explored the identification approach for scientific evolution trajectory including LTSTs. The construction and analysis of network topology are the key points in the community detection and similarity calculation represented by co-occurrences analysis using natural language processing, mainly including co-word and co-sentence. Studies have found that the use of multi-layer network analysis (Tohalino, et al. 2018) and micro-topological structure analysis significantly improve the accuracy of topic analysis (Marinho, et al. 2016). Ba and Liang (2021) proposed a novel knowledge network coupling approach to gauge network linkage between science and technology by integrating knowledge and structural linkages.

Lexical analysis is an alternative solution to avoid citation bias from citation lag and different citation behavior. However, the issue of inconsistent expression between scientific and technological knowledge systems needs to be addressed. Owing to the inconsistency of grammar and semantics between patents and research articles, unifying the concept of a word in advance is necessary. Otherwise, only morphological similarities may be recognized. In general, semantic similarity is difficult to identify and important information is easy to miss; therefore, it reduces the reliability of this method. Correlations exist between the content of research articles and patents through keyword similarities. The topic-term-based method can reflect the strength of the linkage in science and technology; however, it is an unreliable method because of inconsistency between the terms used in patents and research articles.

2.4. Link prediction

This section discusses the link prediction method and the link prediction algorithm based on local information similarity, which will help to further eliminate inconsistencies in the words and concepts shared between basic and applied scientific research.

2.4.1. Link prediction method

Link prediction is used to predict the possibility of a linkage between two nodes in a network that are not connected through using known network nodes, network structure, and other information (Getoor & Diehl, 2005). Link prediction methods can be static or dynamic. The static prediction identifies links that already exist but are undetected or not perceived (i.e., unknown links). The dynamic prediction identifies links that do not currently exist but are likely to exist in the future. Prediction is closely related to network evolution (Lv & Zhou, 2013). Dataset division methods differ between the two types of prediction. The static link prediction mostly uses random sampling, whereas dynamic link prediction considers the time-sequence state (Zhang & Ma, 2015). Zhang et al. (2021) used a refined link prediction approach based on weighted resource allocation to reveal emergence. Network systems are described in one of four ways: unweighted and undirected, weighted and undirected, unweighted and directed, and weighted and directed.

2.4.2. Link prediction algorithm based on local information similarity

Existing link prediction algorithm indicators based on similarity are mainly classified into three types: similarity indicators based on local information, those based on path, and those based on a random walk (Clauset et al., 2008; Guimerà & Sales-Pardo, 2009). Similarity indicator calculation based on local information only considers the network's local topology information. For example, the Common Neighbor (CN) indicator is a basic type of local information indicator based on the principle that nodes with many neighbors are common and more likely to be connected. The influence of the degree of nodes at both ends is based on CNs. The following similarity indicators can be produced using different methods. The Salton indicator, Jaccard indicator, Sorenson indicator, Hub Promoted Index (HPI), Hub Depression Index (HDI), and Leicht–Holme–Newman (LHN-I) index. Another similarity index that only considers the degree of nodes is the preferential attachment (PA) index. The Adamic–Adar indicator (AA) and Resource Allocation indicator (RA) considered the degree of information of the CNs of the two nodes (Lv & Zhou, 2013). Local naive Bayesian models, particularly Local Naive Bayes Common Neighbors (LNBCN), Local Naive Bayes Adamic–Adar (LNBAA), and Local Naive Bayes Resource Allocation (LNBRA) reveal the different roles of various CNs (Liu et al., 2011). The link prediction similarity index based on local information is listed in Table 1.

S_{xy} represents the similarity value of nodes (x and y). This value represents the possibility of connecting the edges between these nodes in the future. The greater the value, the higher the probability of generating connecting edges. The neighbor of node x in the network is defined as $\Gamma(x)$, and $k_x = |\Gamma(x)|$ for the degree of x. Furthermore, z represents the CN of nodes (x and y), s is the prior conditional probability of the network, and R_z is the contribution of the CN node to generate a connected edge.

3. Methodology

This section discusses the research framework using the integrative model to discover LTSTs and introduces the implementation details.

Table 1
Link prediction similarity index based on local information.

Similarity Indicators	Equations
CN	$S_{xy} = \lvert\Gamma(x)\cap\Gamma(y)\rvert$
Salton	$S_{xy} = \frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\sqrt{k_x k_y}}$
Jaccard	$S_{xy} = \frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\lvert\Gamma(x)\cup\Gamma(y)\rvert}$
Sorenson	$S_{xy} = \frac{2\times\lvert\Gamma(x)\cap\Gamma(y)\rvert}{k_x + k_y}$
HPI	$S_{xy} = \frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\min\{k_x, k_y\}}$
HDI	$S_{xy} = \frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\max\{k_x, k_y\}}$
LHN- I	$S_{xy} = \frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{k_x k_y}$
PA	$S_{xy} = k_x k_y$
AA	$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{\log k_z}$
RA	$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{k_z}$
LNBCN	$S_{xy} = \lvert\Gamma(x)\cap\Gamma(y)\rvert\times\log s + \sum_{z\in\Gamma(x)\cap\Gamma(y)} \log R_z$
LNBAA	$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{\log k_z}\times(\log s + \log R_z)$
LNBRA	$S_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{1}{k_z}\times(\log s + \log R_z)$

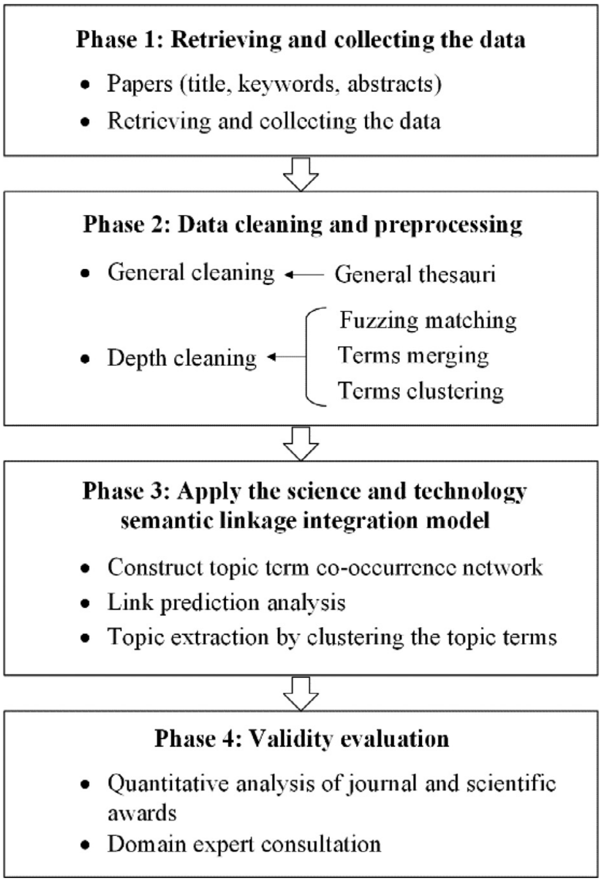

Fig. 1. Research framework using an integrative model.

3.1. *Research framework using the integrative model to discover LTSTs*

Fig. 1 illustrates a research framework using the integrative model to discover LTSTs. The framework consists of four phases. Phase 1 is related to retrieving and collecting the data. Detailed information is provided in Section 4.1. Phase 2 is related to data cleaning and preprocessing, which includes general and in-depth cleaning (a detailed explanation is provided in Section 4.2). Phase 3 is related to the science and technology semantic linkage integrative model; Section 4.2 can be referred to for the details. Phase 4 is about validation and evaluation. The validation process will combine the quantitative analysis of authoritative journals and scientific awards with expert consultation to assess the recognition results of the proposed model, as explained in Section 4.4.

3.2. *Developing science and technology semantic linkage using an integrative model*

Mechanisms for the evolution of science and technology are different and their evolutionary trajectories are also different. However, scientific discoveries and technological innovations evolve dependently and interactively. Both their relative independence and the linkage of papers and patents should be maintained while measuring their scientific achievements. In the process of linked-degree calculation between science and technology, matching implicitly related to the scientific and technological terms is critical to detect LTSTs. Therefore, this study firstly emphasis on scientific and technological networks independently, and proposes the science and technology semantic linkage integration model (Fig. 2), which performs knowledge network integration operations twice. Simple fusion first flags explicit linked topic terms, which can be used as an automatic seed marking for the next link prediction to identify the implicit linkage of topic terms.

The implementation steps of the integrative model are given below:

Step 1: Obtaining explicit linkage using the simple fusion

We construct a scientific topic term co-occurrence network "S" and a technological topic term co-occurrence network "T" (the upper layer of Fig. 2). For a large scientific corpus, data cleaning of topic terms cannot be completed without manual involvement, which is very time-consuming. The vantage point is powerful for extracting and indexing data (Vantage Point, 2020). Thus, we employed Vantage Point and the semi-automated term clumping process proposed by Zhang et al. (2014) to clean data which is

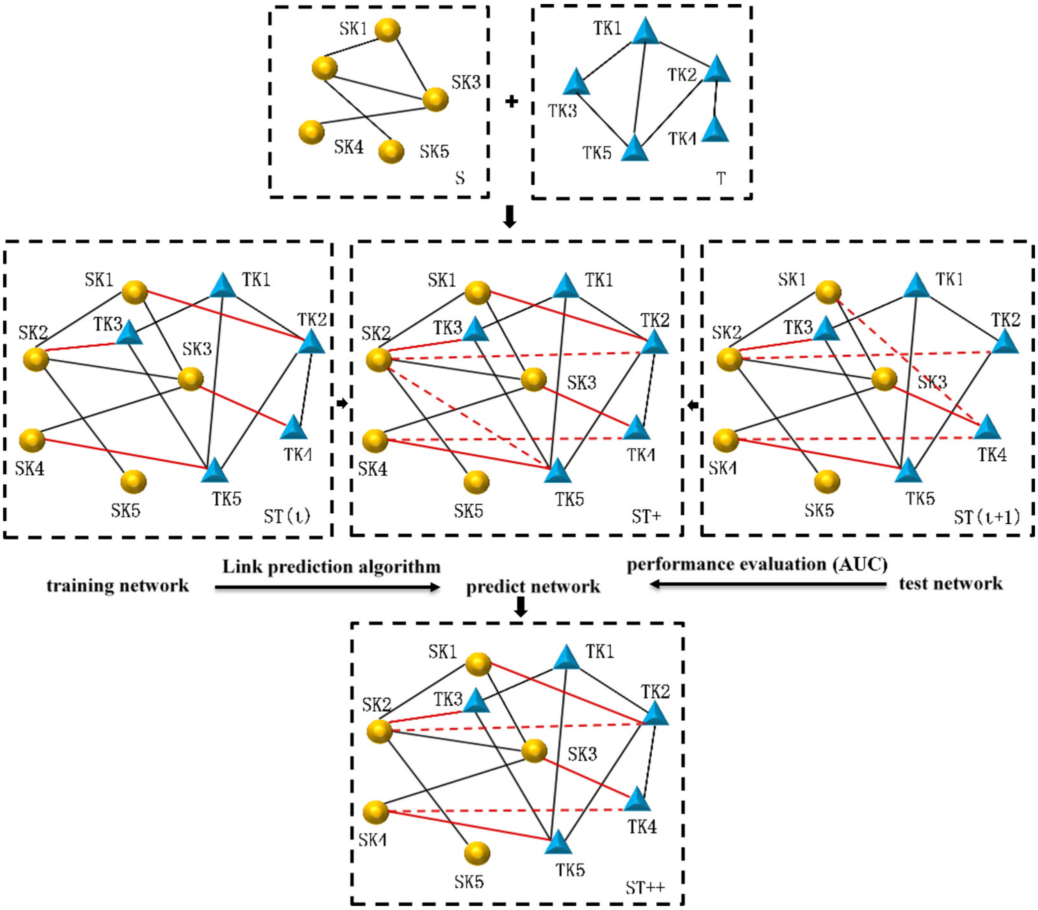

Fig. 2. Science and technology semantic linkage integrative model (Dot node represents for scientific topic terms. Triangle node represents technological topic terms. Solid black lines represent the direct connection of the scientific topic terms or technological topic terms. Solid red lines represent the direct connection between scientific and technological topic terms. Similarly, the red dotted line represents links from linkage prediction.

combined with instruction from domain experts. Afterward, the scientific topic term co-occurrence network "S" and the technological topic term co-occurrence network "T" could be built using the co-occurrence matrix function.

The co-occurrence topic terms in the S network and T network can be used as intermediaries for the simple integration of the S and T networks. Thus, the simple fused science and technology topic term co-occurrence network "ST", including $ST_{(t)}$ and $ST_{(t+1)}$, is obtained (left and right sides of the middle layer of Fig. 2). In the ST network, the solid black line indicates the direct connection of scientific topic terms or a direct link of technological topic terms. In Fig. 2, the solid red line indicates the direct connection between scientific and technological topic terms.

Step 2: An implicit linkage prediction of ST network to create ST+ network

Based on the similarity of local information, the link prediction algorithm is used to predict implicit semantic linkage in the ST network. We use the link prediction algorithm based on the similarity of local information to predict implicit semantic linkage in the ST network. Both the original simple fusion linkage and prediction relation are obtained. Afterward, the topic term co-occurrence linked network ST+ is obtained. It is shown in the central network of the middle layer of Fig. 2.

Step 3: Obtaining the science–technology topic terms-enhanced association ST++ network

The scientific and technological topic terms with indirect or implicit linkage are indicated by the red dotted lines in the ST+ network (Fig. 2). Subsequently, pairs of topic terms that have an indirect linkage degree above a certain threshold are extracted (e.g., SK2 and TK2; and SK4 and TK4 topic terms pairs). The newly discovered pairs are integrated with the ST network. Finally, the science and technology topic terms enhanced association network ST++ is obtained for the effective mapping of scientific and technological topic terms. It has been shown in the bottom layer of Fig. 2.

Step 4: Clustering topic terms to extract topics of ST++ network

The force-directed layout in the ECharts tool can be used to visualize the layout of the ST++ network (Apache ECharts, 2020). Force-directed layout simulates a spring/charge model, which adds repulsion and attraction between two nodes of each edge. Afterward, combined with the instruction of a domain expert, community division is carred out. Finally, semantic-enhanced LTSTs are obtained. Based on identifying major communities, we further identify key LTST chains, which are constituted by connected LTST terms, then finer scale LTSTs will be obtained.

3.3. Implicit semantic LTST recognition using link prediction

Analysis of link prediction

After obtaining the undirected co-occurrence network for different periods, we used an undirected network link prediction algorithm to determine the implicit relation between scientific and technological topic networks. The time-series topic term co-occurrence network in the targeted domain was used as the analysis object. A total of 13 similarity indicators using local information were employed to perform dynamic link prediction. Subsequently, the prediction performance indices were compared. The optimal index was selected to predict the implicit association of the network.

In the domain topic term co-occurrence network G (V, E), V represents the topic term in the network (set of nodes), and E represents the topic term co-occurrence relation (set of edges). Over time, the association relation of the topic term co-occurrence network will change dynamically (with nodes increase, decrease, appear, disappear or stay the same). We used the topic term co-occurrence network G_t at time t as the training network. Afterward, we used a link prediction algorithm to assign a similarity value S_{xy} to each pair of topic terms (x, y) in G_t. This value represents the possibility of producing a co-occurrence association between the topic term pairs in the future. The greater the value, the higher will be the probability that the topic terms will have a co-occurrence relationship. To evaluate the dynamic link prediction performance of each algorithm, we used the co-occurrence network G_{t+1} at time $t + 1$ as the test network. Furthermore, we evaluated the accuracy of the link prediction algorithm using information about the test network and prediction network that combined with the area under the receiver operating characteristic curve (AUC) evaluation index (Ren et al., 2015).

Link prediction can only predict the co-occurrence relation between existing topic terms. It failed to predict the possibility of co-occurrence association between existing topic terms and new terms. Therefore, we preprocessed the training and test networks to ensure that the topic terms of the two sets have identical nodes (topic terms), and the only difference is the edge of the network before dynamic link prediction. The used test networks were processed by topic term co-occurrence networks.

AUC definition

In the test network (G_{t+1}), the probability of similarity exists between newly added edges (topic term co-occurrence edges exist in G_{t+1} but not in G_t) is higher than between any non-existent edge. If the similarity value of the newly added edges in the test network is greater than the similarity value of non-existent edges, 1 point is added. If the two edges have the same value, 0.5 point is added. Similarly, suppose that the number of newly added edges in the test network is m, and the number of non-existing edges is n, then the new edges and non-existent edges are independently compared $m \times n$ times. If there are n' times, the similarity value of the newly added edges in the test network is greater than the similarity value of non-existing edges. The two similarity values are equal to n" times. Therefore, AUC is calculated using the given equation 1.

$$\text{AUC} = \frac{n' + 0.5n''}{m \times n} \tag{1}$$

The dynamic link prediction accuracy of the algorithm will be higher with the larger value of AUC (B Zhang & Ma, 2015). The topology of the topic term co-occurrence network changes dynamically over time. Depending on the network, the links prediction performance will also be changed accordingly. To ensure the reliability of the performance evaluation of dynamic link prediction, the domain time-series topic term co-occurrence network is divided into several sub-networks (for initial time t_0 and termination time t_q). To perform the first link prediction and to calculate AUC_1, we used the topic term co-occurrence network at time t_0 as the training network and the topic term co-occurrence network at time t_1 as the test network. Subsequently, to perform the second link prediction and to calculate AUC_2, we used the topic term co-occurrence network at time t_1 as the training network and the topic term co-occurrence network at time t_2 as the test network. Finally, to calculate the value of AUC_q, we used the topic term co-occurrence network at time t_{q-1} as the training network and the topic term co-occurrence network at t_q as the test network for the q^{th} link prediction. Finally, we made a comprehensive judgment of the dynamic link prediction performance based on the AUC values.

4. Empirical analysis

4.1. Research data

Currently, the GEV field is widely dominated by countries, enterprises, and academia. In our prior investigations, the number of papers and patent data in the field of GEV was moderate, and the research results have certain continuity (Xu et al., 2020). Therefore, We selected the research articles and patents in the GEV domain as the empirical field. This section describes the specific retrieval strategies.

Paper data

We retrieve paper data from the Web of Science database, the advanced search query is as follow:

TI =((Genetic* adj engineer* or DNA adj engineer*) and (vaccine* or antigen*)) or 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*).

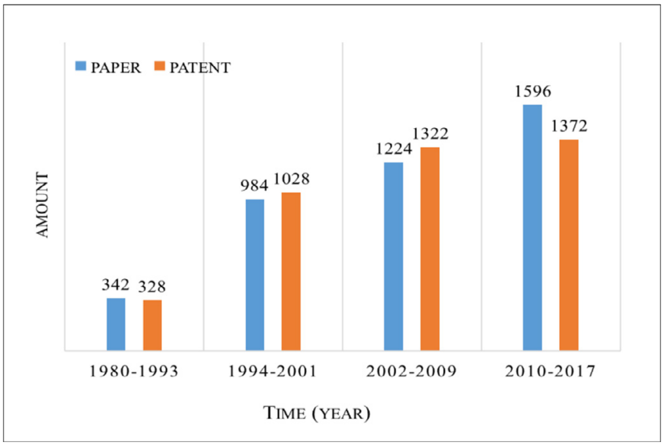

Fig. 3. Number of distributions of research articles and patents in the four-time periods.

Note: TS denotes title, abstract, author keywords and keywords plus.

In above search query, the document type is limited to research articles or proceedings paper. The time is limited to the publication year until 2017. The retrieval date is January 6, 2018. The export format is a full record plain text.

Patent data

We also use an advanced search method to retrieve patent data from the Derwent Innovation Index database. The search query is given as:

TI = ((Genetic* adj engineer* or DNA adj engineer*) and (vaccine* or antigen*)) or 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).

Note: TI, TS, and PN represent title, title and abstract, and patent number, respectively.

The time limit of the patent application is until 2017. The retrieval date is January 6, 2018, and the export format is in full record and plain text.

Firstly, we retrieved and downloaded the data. The duplicate record was removed. Finally, 4,146 patents and 4,050 research articles were obtained. Because of the long time required to collect research data, essential topics of a specific stage may be overlooked under the entire period, resulting in the inability to identify the topics in the growing phase. Therefore, we divided the time axis into four time periods. It takes at least 8 years for vaccines to be developed and marketed; therefore, the safety and effectiveness of vaccines can be better guaranteed (Jin, 2016). In addition, dividing the time stages into equal-length intervals is conducive to comparing the analysis of the results of each stage. Therefore, the data before 1993 were classified into a period from 1980 to 1993, which is defined as the first period (period 1). After 1994, every 8 years was regarded as a period. Finally, we obtained the number distribution of patents and research articles in four periods, as shown in Fig. 3.

4.2. *Construction of topic-term co-occurrence network*

Firstly, text data were imported into Clarivate's Derwent Data Analyzer™, and the multiword list in the field of "combined keywords + phrase" was selected as the field-of-topic term. The list in the "combined keywords + phrase" field was extracted and cleaned from titles and abstracts using a series of term-clumping processes, which is good at cleaning words and terms and synthesizing technical synonyms (Zhang et al., 2014). The term-clumping processes extract and clean topic terms through text mining, including data cleaning based on general thesauri, depth cleaning based on fuzzy matching, terms merging, and terms clustering based on principal component analysis. Thereafter, we removed the stop words, normalized the word forms on singular and plural issues, and performed manual semantic cleansing to remove irrelevant words and merge synonyms.

In the later stage, the number of documents was increased, and the density of the subject word network became larger. Due to the large number of papers and patent topic words, to better display the analysis results, we set the thresholds for the word frequency of research articles and patent topic words as research articles topic words ≥ 15 and patent topic words ≥ 20, respectively. The number of topic terms in the fourth stage is listed in Table 2. For the S and T network in the fourth stage, the S network has 587 words, and the T network has 498 words. We present some specifications of the networks in Table A1 and Table A2 in the appendix. The partial topic term co-occurrence structure in research articles is given in Table A1. The partial topic term co-occurrence structure of patents is given in Table A2. The co-occurrence topic terms in S and T network can be used as intermediaries for the simple integration of S and T networks. Finally, a total of 977 topic terms were obtained in four-time periods and four topic terms co-occurrence networks were built.

Table 2
Number of topic terms in research articles and patents.

Stages	Number of topic terms of the research articles	Number of topic terms of the patents
Phase 1	330	236
Phase 2	528	452
Phase 3	567	494
Stage 4	587	498

Table 3
Average of AUC.

Indexes	Average of AUC	Indexes	Average of AUC
CN	0.733	AA	0.737
Salton	0.717	RA	0.739
Jaccard	0.520	PA	0.711
Sorenson	0.580	LNBCN	0.738
HPI	0.716	LNBAA	0.738
HDI	0.690		

Table 4
Topology indicators before and after link prediction.

Network	Average degree	Average path length	Density	Diameter
Before link prediction	50.401	2.224	0.049	5
After link prediction	757.212	1.265	0.737	3

4.3. Link prediction

Table 3 depicts the calculation of the results of the average area under the curve (AUC). The link prediction is performed on the ST network in the fourth stage. In the first step, all co-occurrence topic terms datasets are divided into training and test sets at a ratio of 9:1 for the relatively large-scale dataset. In the second step, a list of classic algorithms given in Table 1 was used to perform multiple training experiments on the training set and the results were compared with the test set to calculate the AUC. The AUC indicator was used to evaluate the effect of the algorithm. The link prediction result of the latter stage was obtained through the previous stage, which was compared with the actual value to calculate the AUC indicator for evaluating the effect of the algorithm to obtain the test results of each stage. Finally, the average of the three stages was taken as the final AUC value (Table 3). In the third step, we carried out 100 independent repeat experiments and calculated the average AUC value. The link prediction algorithm having the best performance was selected as the better index. In the fourth step, we applied the pre-screening algorithm on the ST network. The AUC was used to evaluate the accuracy of link prediction.

In the fourth step, during the application of the science and technology semantic linkage integrative model, the link prediction was performed on the ST network with 977 term words. When link prediction was used, all frequency topic term nodes and link prediction for different topic term pairs were predicted. The topology indicators including average degree, average path length, density, diameter before and after link prediction are given in Table 4. All these four topology indicators before and after link prediction indicated the improvement of the network connectivity. We present the specifications of the networks in Tables A3, A4, and A5 in the appendix.

Table A3 represents the current term co-occurrence matrix of the S topic terms and T topic terms for the fourth period Table A4. shows that the S topic terms and T topic terms co-occurrence frequency currently does not appear, but the similarity value using the WRA algorithm may co-occur in the future. WRA algorithm was used; moreover, the similarity value used for two terms may have a co-occurrence relationship in the future after link prediction, regardless of whether the S topic terms and T topic terms have already appeared together (Table A5).

4.4. Results

To determine the advantage of our proposed integrated model for the detection of LTSTs, we carried out a comparison experiment. Firstly, we explain the co-word networks performed in our proposed model before and after link prediction. Fig. 4 illustrates the topics discovered using the co-word networks before link prediction. Similarly, co-word networks after the link prediction are given in Fig. 5. To obtain finer-scale LTSTs on the base of major communities, we identified the key LTST chains, that are constituted by connected LTST terms. By considering to display more nodes while keeping a clear view of the two-co-word networks, after repeating multiple tests, we found 200 nodes with the strongest co-occurrence association that could provide a good illustration. We used different node shapes to represent three types of topic terms such as triangle nodes for scientific topic terms, square nodes for technological topic terms, and dot nodes for LTSTs. The topic term linkages identified through link prediction are represented by dotted lines.

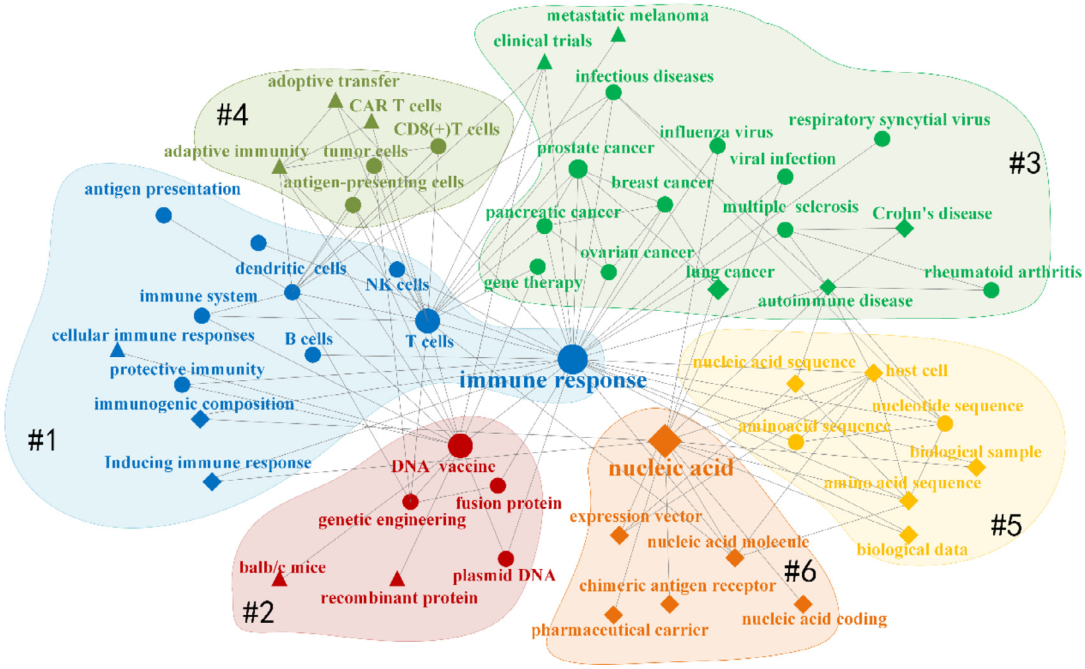

Fig. 4. Diagram of LTSTs before link prediction (▲ triangle node represents scientific topic terms, ◆ square node represents technological topic terms, and ● dot node represents LTSTs terms).

Combined with expert opinions, we have described the content and evolution trends of the topics in detail to analyze the function of link prediction for recognition of the LTST.

4.4.1. LTSTs discovered before link prediction

Based on the force-directed layout of the ECharts tool, six community divisions were found. Specifically, # 1 for immune response, # 2 for DNA vaccine preparation technology, # 3 for therapeutic areas of vaccines, # 4 for adoptive cellular immunotherapy, # 5 for molecular biology research in vaccine development, and # 6 for chimeric antibody receptor preparation (Fig. 4). Among the six community divisions, Fig. 4 illustrates the research emphasis on basic and applied of the 6 topics are notably different; however, there is no purely basic or purely applied research topic among the six topics. #2 and #4 are more focused on basic research (the topic terms contain more triangles). The #5 and #6 are more focused on applied research (the topic term contains more squares). Similarly, #1 and #3 are more inclined to LTSTs (the topic terms contain more circles, but also mixes of triangles and squares).

4.4.2. LTSTs chains before link prediction

CAR-T cell

There are two representative LTSTs chains of topic terms, given as:

CAR-T cell—T cell —NK cell—immune response—lung cancer;

CAR-T cell—adaptive immunity—dendritic cells—T cell—NK cell—immune response—lung cancer

Chimeric Antigen Receptor T-Cell Immunotherapy (CAR-T) plays a significant role in the treatment of acute leukemia and non-Hodgkin's lymphoma. Owing to the complex genetic heterogeneity of solid tumors, immuno-suppressive micro-environment, and other factors, CAR-T faces many challenges in the treatment of solid tumors. Natural killer cells (NK) are an essential part of the body's innate immune system. Compared with T cells, NK cells are more effective in killing tumor cells and are less immunogenic. Chimeric antigen receptor NK cells (CAR-NK) technology is an adaptive immune cell therapy that enhances the targeting and killing function of NK cells. Therefore, compared with CAR-T, the CAR-NK shows better prospects of applying to cellular immunotherapy for refractory malignant tumors. Scientists have also carried out many clinical trials for solid tumors, such as lung cancer (Mehta & Rezvani, 2018; Patel et al., 2019). Moreover, dendritic cells, as antigen-presenting cells, play an important role in cellular immunity mediated by T cells and NK cells (Fernandez et al., 1999). Studies have used anti-tumor synergy between immune cells to treat cancer with super-dendritic cells and CAR-NK. This has overcome the immuno-suppressive effect of the solid tumor micro-environment (Belounis et al., 2020).

Autoimmune disease

The LTSTs chains of topic terms are given as:

balb/c mice — DNA vaccine — dendritic cells — immune response — autoimmune disease

The causes of autoimmune diseases are complex, and genes and epigenetics are the primary pathogens. Inflammation, B cells, T cells, and related cytokines play a vital role in the occurrence and development of diseases. Targeted DNA vaccines for autoimmune

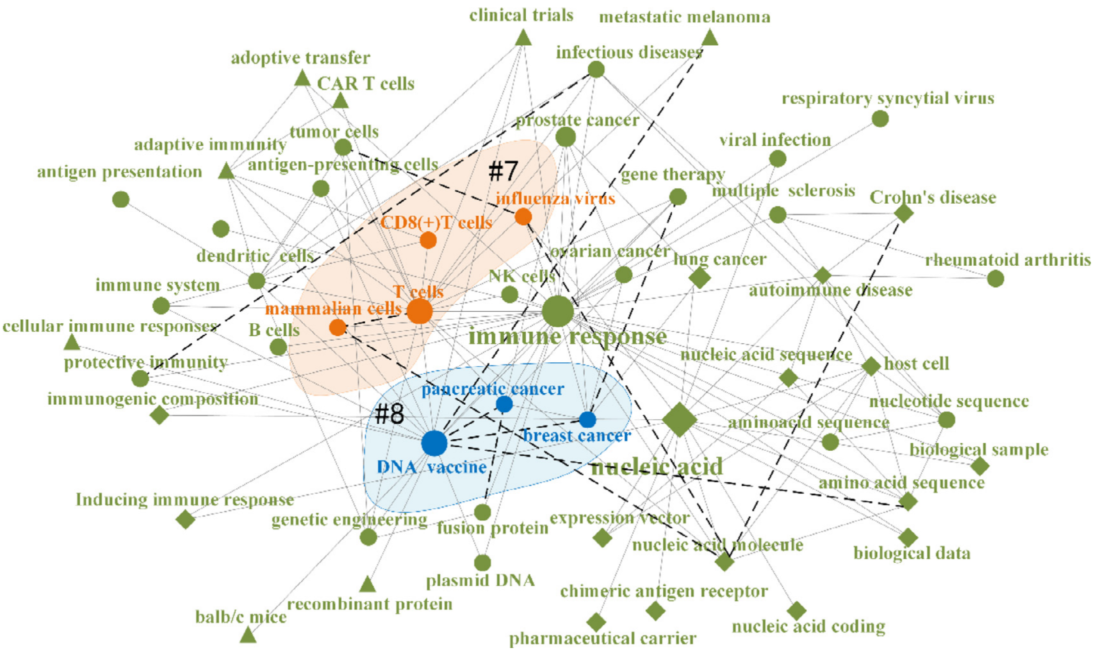

Fig. 5. LTSTs discovered after link prediction (▲ triangle node represents for scientific topic terms, ◆ square node represents technological topic terms, and ● dot node represents LTSTs terms. A solid line represents the direct connection between scientific and technological topic terms. While dotted line represents links from linkage prediction).

diseases have become the focus of current research (Hobernik & Bros, 2018). For example, the recombinant interferon DNA vaccine on a mouse was identified during one of the experiments. The specificity of the vaccine can block the production of IFN-α, thereby delaying the differentiation of mononuclear phagocytes into activated dendritic cells. It also weakens the activation of reactive T and B lymphocytes downstream, and thus slowing disease progression (Pan et al., 2004).

Adaptive immunity

The topic terms are linked as follows: *adaptive immunity — dendritic cells — CD8(+) T cells — T cell — immune response — influenza virus or viral infection*

Dendritic cells are the main initiators of the adaptive immune system. They are responsible for the uptake, processing, and classification of antigens that provide antigen information to T cells and initiate antigen-specific T-cell immune responses. The dendritic cells have potent abilities to activate primary T cells. Mature dendritic cells can express high levels of antigen-presenting molecules (MHC-I and MHC-II), thus stimulating a cellular immune response with CD8 + T cells as the main body. Virus clearance depends on robust, polyclonal, specific CD8 + cytotoxic T cells to fight viral antigens. In the early stages of a virus infection, the active expansion of specific CD8 + T cells is necessary for the body to resist viruses and avoid severe diseases (Hashimoto et al., 2019). Hence, understanding the action mechanisms of CD8 + T cells helps develop vaccine components that can withstand new virus strains, especially for the types of pathogenic membrane proteins that are easily mutated, such as universal influenza vaccines (He et al., 2016).

4.4.3. LTSTs discovered after performing link prediction

After link prediction, two new research topics were discovered. It has shown in Fig. 5 particularly, # 7 Anti-influenza research of CD8(+) T cells and # 8 Anti-tumor DNA vaccine.

Group # 7: Anti-influenza research of CD8(+) T cells

The cellular immunity of mammals mainly depends on the lymphatic system. It is, therefore, a connection between “T cell” and “mammalian cells” found. T cells mainly mediate cellular immunity (B cells participate in humoral immunity), and CD8(+) T cells are kind of the transmembrane glycoprotein of the T cell receptor (TCR) of the co-receptor often exhibits strong antiviral effects. Therefore, it is often used as a test indicator in the antiviral process such as HIV and hepatitis B (HBV). In the research and development of influenza vaccines, because the subtypes of influenza virus strains that circulate every year are different, research and development are done every year. However, the current influenza virus is only targeted at a few subtypes, and there is no broad-spectrum influenza vaccine. CD8(+) T cells can kill the virus, but also have the memory of the virus strain. The current research on CD8(+) T cells and the research on anti-influenza vaccines based on CD8(+) T cells have also become the focus.

Group # 8: Anti-tumor DNA vaccine

Through link prediction, a direct connection is established between DNA vaccine and breast and pancreatic cancer, etc., making it easier to identify the topic.

4.4.4. LTSTs chains after link prediction

Based on identifying major communities, we further identify key LTSTs chains, which are constituted by connected LTSTs terms, and the microcosmic LTSTs are obtained. After link prediction, the topic becomes more straightforward and simple, for instance before link prediction, the link of the topic terms is given as:

adaptive immunity — dendritic cells — CD8(+) T cells — T cell — immune response — influenza virus/viral infection

After link prediction, the link of the topic terms is given as:

adaptive immunity — CD8 (+) T cells — dendritic cells — T cell —— influenza virus

Therefore, through link prediction, the connection between the T cell and the influenza virus is established directly. In contrast, without link prediction, the "immune response" must be used as a connected topic term during the link prediction. Similarly, we observed that the topic association becomes more direct after link prediction because, after the link prediction, the DNA vaccine is directly linked to the corresponding cancer diseases. For instance, before the link prediction, the topic terms are linked as follows:

DNA vaccine — immune response — pancreatic cancer/prostate cancer/breast cancer/ovarian cancer.

After link prediction, the topic terms are linked as follows:

DNA vaccine —- pancreatic cancer/ breast cancer

4.4.5 Newly discovered LTSTs chains using link prediction

Tumor vaccines based on viral vectors

After link prediction, the link of topic terms is given as:

genetic engineering — tumor cells —- viral infection

Virus-infected macrophages and dendritic cells are typical antigen-presenting cells, which can deliver foreign antigens to host T cells and B cells, suggesting that it may be a promising vaccine carrier. In recent years, recombinant virus vectors have been widely used in the development of vaccines. To obtain the recombinant virus, they use genetic engineering technology to insert exogenous protective antigen genes into the viral genome, which is a type of carrier vaccine that expresses the corresponding target protein after immunizing the body and inducing an immune response (Rollier et al., 2011).

Metastatic melanoma

After link prediction, the topic terms are linked as: *metastatic melanoma —- DNA vaccine*

Melanoma is a type of invasive skin cancer. Major research is ongoing to find melanoma-associated antigens and implement melanoma immunotherapy. The DNA vaccine is an important method for tumor-specific immunotherapy (Lopes et al., 2019). For example, Scancell Immuno Body's new melanoma drug SCIB1 (DNA vaccine) can target dendritic cells, activate cell-killing CD8 T cells, and cause a broader immune effect to fight tumors. In 2018, the vaccine has entered phase-II clinical trials in the UK (Patel et al., 2018).

5. Conclusion and Discussion

5.1. Main conclusion

The main aim of the current study is "to identify LTSTs at the micro level using text mining analysis". Particularly, we propose an integrated model for the detection of LTSTs and apply the method as a case study in the field of gene-engineered vaccines (GEV). The model is based on the semantic coupling of scientific terms and technological concepts. The integrative model fuses the term co-occurrence networks of basic and applied research, in which the topic network is expanded and the semantic association of the network is enhanced. Furthermore, the link prediction is used to mine the latent semantic association between topic terms in science and technology topic networks. The GEV field is used to test the effectiveness of our proposed method. Emerging LTSTs can be identified by comparing the topic identification of the non-fusion network with the topic identification of the fusion network. Domain experts confirm that the integrative model with link prediction is better than the simple fusion model at identifying LTSTs. The results suggest that the proposed method can be applied to all fields, but the analysis effect may be quite different in different fields. Previous studies have shown that in the field of biology, science and technology are more closely related, so the recognition effect may be more significant. In future, empirical analysis will be conducted in other fields. However, although two-fold fusions can significantly improve the accuracy of LTST recognition, full recognition of LTSTs by text analysis is still challenging, owing to the natural language attributes of the text, it cannot reach 100% matching.

5.2. Major contributions

Improved effect of LTST discovery owning to integrative model

The main contribution of this article is the proposed practical LTST detection model based on text mining analysis. It includes two semantic-related enhancement processes: explicit semantic enhancement and implicit semantic enhancement. The implicit semantic recognition of LTSTs can be enhanced through simple fusion. In addition, the link prediction is applied to the semantic linkage of topic terms in science and technology, which helps erase inconsistencies in the words and concepts shared between basic and applied scientific research by alignment of the synonym phenomenon of scientific and technical terms. Therefore, our proposed method can identify LTSTs concisely and directly. It discovers topics that cannot be recognized by simple fusion. Moreover, LTST recognition can be further improved through link prediction.

LTSTs discovering support science and information policy

The methodology proposed in this article provides crucial intelligence support for strategic decisions in science and technology. Particularly, the allocation of innovative resources, and the layout of industrial development. This improvement is helpful for policy-makers and stakeholders to efficiently familiarize themselves with the scientific landscape, especially topics with high transformation opportunities from the laboratory to practical applications for the support of decision making. Furthermore, for science and technology innovation itself, our findings can help to reveal the flow and diffusion of scientific and technological innovation factors. It will promote an efficient flow of knowledge between basic and applied research and accelerate the knowledge transformation of scientific, innovative achievements.

5.3. Limitations

Although the study has useful implications, it has several limitations that need to be focused on in future studies.

Use of simple fusion and link prediction only once with no further iterations

Theoretically, though repeated iterations of simple fusion and link prediction, the recognition effect of implicit LTSTs could be further enhanced. Simple fusion and link prediction cannot replace each other. The simple fusion requires a more standardized vocabulary of scientific and technological terminology. The results of link prediction will be further improved with the higher validity of simple fusion. New matched semantic topic terms from link prediction can be added to the linked vocabulary in science and technology that is used in simple fusion. Iterating this process in this manner should further improve the accuracy with which LTSTs are identified. In a future study, we will use a combination of simple fusion and link prediction in an iterative way.

Simplified topic network building and community division

Firstly, a direct co-occurrence relation and an indirect semantic correlation exist between scientific and technological topic terms. However, this study only used the local link prediction method. Therefore, it is inferior in identifying meaningful indirectly linked topic terms in the science–technology network. Secondly, different topic terms belong to multiple topics. It means that the topic communities are overlapped rather than independent. Theoretically, the link prediction will promote the existence of more overlapping communities by increasing the association between different communities. Therefore, the independent community division method in this study affects the analysis of topic association. In future studies, the overlapping community identification method should be used to improve the analysis of topic association.

Ignored publication time delay of patent and research articles

The current study did not consider the delay of patent publication, which requires 18 months from application to publication. The publication time is used in this study; simultaneously, the delay in patent publication differs between countries or regions. There is a time lag for both patent and article publication; however, the delay time between the compared paper and the patent data pair is not consistent.

5.4. Further research

For future study, it is recommended to further mine the latent semantic association between topic terms in science and technology through more semantic dictionaries and grammar rules, and use weak signal analysis to further reduce the time lag issue.

Semantic dictionaries and grammar rules should add to the integrative model

In the topic recognition methods, it is quite difficult for semantic dictionaries to guarantee keyword coverage. Particularly, the co-word analysis cannot completely reflect the relation between keywords. Therefore, it is required to make full use of the existing semantic dictionary or domain ontology. The clustering based on the semantic dictionary should be combined with similarity analysis of co-occurrence knowledge units. Furthermore, comprehensive topic term acquisition should be combined with a broader extraction scope to include grammar and semantic levels. The calculation of topic term correlation requires the use of the existing semantic dictionary and domain ontology and combining the semantic dictionary similarity with the knowledge unit co-occurrence analysis. The fusion of multi-relationships related to the topic should also be considered.

Focus on the weak signal analysis to reduce the time lag

To overcome the time lag caused by citation analysis, this study uses text analysis and link prediction analysis, so the time lag issue can be reduced to a certain extent; however, this study focuses on analysis of high-frequency words, which may ignore the analysis of weak signals that can better predict future topics and topic associations. We will try to use weak signal analysis in future. Theoretically, it should be possible to further reduce the issue of time lag.

Author contribution

Haiyun Xu: proposed the research idea, designed the research, drafted and revised the manuscript.
Zenghui Yue: drafted the manuscript and performed partial data analysis.
Hongshen Pang**:** performed partial data collection and partial data analysis.
Ehsan Elahi: performed partial data analysis, drafted and revised the manuscript.
Jing Li:performed partial data analysis, drafted and revised the manuscript.
Lu Wang: performed partial data collection and analysis.

Acknowledgments

This article is an outcome of the projects "Study on the Recognition Method of Innovative Evolving Trajectory Based on Topic Correlation Analysis of Science and Technology (No. 71704170)" and "Study on Regularity and Dynamics of Knowledge Diffusion Among Scientific Disciplines (No. 71704063)" funded by the National Natural Science Foundation of China, and Taishan Scholar Foundation of Shandong Province of China (202103069).

Appendix

Table A1
Topic terms co-occurrence of data structure in patents (partial).

Topic term	Topic term	Weight
Immune response	Nucleic acid	71
CAR T-cells	T-cells	65
T-cells	Adaptive immunity	65
Nucleic acid	Nucleic acid sequence	52
DNA vaccine	Plasmid DNA	51

Table A2
Topic terms co-occurrence of data structure in research articles (partial).

Topic term	Topic term	Weight
Adaptive immunity	T-cells	65
CAR T-cells	T cells	65
Adaptive immunity	CAR T cells	50
Adoptive transfer	T cells	48
DNA vaccine	plasmid DNA	47

Table A3
The current topic term co-occurrence matrix (partial).

	Antigenic variation	Expression cassette	Vaccine therapy	Heavy chain	T-cells	Autoimmune disease
Antigenic variation	0	0	0	0	0	0
Expression cassette	0	0	0	0	2	1
Vaccine therapy	0	0	0	0	0	0
Heavy chain	0	0	0	0	0	0
T-cells	0	2	0	0	0	10
Autoimmune disease	0	1	0	0	10	0
MHC Class-I	0	0	0	0	0	0
Reperfusion injury	0	0	0	0	0	1

Table A4
The future topic term co-occurrence matrix (partial).

	Antigenic variation	Expression cassette	Vaccine therapy	Heavy chain	T-cells	Autoimmune disease
Antigenic variation	0.212	0	0	0	0.022	0
Expression cassette	0	0.672	0.003	0	0	0
Vaccine therapy	0	0.003	0.247	0	0.147	0.010
Heavy chain	0	0	0	0.032	0	0.021
T-cells	0.022	0	0.147	0	9.250	0
Autoimmune disease	0	0	0.010	0.021	0	4.419
MHC Class-I	0.134	0.004	0.048	0	0.522	0.051
Reperfusion injury	0	0.037	0	0	0.050	0

Table A5
The topic term co-occurrence matrix after link prediction (partial).

	Antigenic variation	Expression cassette	Vaccine therapy	Heavy chain	T-cells	Autoimmune disease
Antigenic variation	0.212	0	0	0	0.022	0
Expression cassette	0	0.671	0.003	0	*0.360*	*0.441*
Vaccine therapy	0	0.003	0.247	0	0.147	0.010
Heavy chain	0	0	0	0.032	0	0.021
T-cells	0.022	*0.360*	0.147	0	9.250	*1.836*
Autoimmune disease	0	*0.441*	0.010	0.021	*1.836*	4.419
MHC Class-I	0.134	0.004	0.048	0	0.522	0.051
Reperfusion injury	0	0.037	0	0	0.050	*0.170*

References

Albert, T. (2016). *Measuring technology maturity: Operationalizing information from patents, Scientific Publications, and the Web*. Springer.
Apache ECharts (incubating) TM (2020). Features. https://echarts.apache.org/en/feature.html.
Ba, Z., & Liang, Z. (2021). A novel approach to measuring science-technology linkage: From the perspective of knowledge network coupling. *Journal of Informetrics, 15*(3), Article 101167.
Balconi, M., Breschi, S., & Lissoni, F. (2004). Networks of inventors and the role of academia: an exploration of Italian patent data. *Research Policy, 33*(1), 127–145.
Baltimore, D., Conn, R., Press, W. H., Rosenbaum, T., Spergel, D. N., Tilghman, S. M., & Varmus, H. (2021). *Should the Endless Frontier of Federal Science be Expanded?* arXiv preprint arXiv:2103.09614.
Bassecoulard, E., & Zitt, M. (2004). *Patents and Publications*. Springer.
Belounis, A., Ayoub, M., Cordeiro, P., Lemieux, W., Teira, P., Haddad, E., & Duval, M. (2020). Patients' NK cell stimulation with activated plasmacytoid dendritic cells increases dinutuximab-induced neuroblastoma killing. *Cancer Immunology, Immunotherapy, 69*(9), 1767–1779.
Boyack, K. W., & Klavans, R. (2008). Measuring science-technology interaction using rare inventor-author names. *Journal of Informetrics, 2*(3), 173–182.
Brachtendorf, L., Gaessler, F., & Harhoff, D. (2020). *Truly standard-essential patents? A semantics-based analysis* https://ssrn.com/abstract=3603956.
Bush, V. (1945). *Science, the endless frontier*. Ayer Company Publishers.
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.
Cassiman, B., Glenisson, P., & Looy, B. V. (2007). Measuring industry-science links through inventor-author relations: A profiling methodology. *Scientometrics, 70*(2), 379–391.
Chaminade, C., & Lundvall, B.-Å. (2019). *Science, technology, and innovation policy: old patterns and new challenges*. Oxford University Press.
Chen, L., Xu, S., Zhu, L., Zhang, J., & Guancan, Y. (2020). A deep Learning based method for extracting semantic information from patent documents. *Scientometrics, 125*(1), 289–312.
Clauset, A., Moore, C., & Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature, 453*(7191), 98–101.
Conn, R. W. (2021). Expanding the endless frontier. *Science, 372*(6537) 9-9.
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.
Du, J., Li, P. X., Guo, Q. Y., & Tang, X. L. (2019). Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis. *Journal of Informetrics, 13*(1), 132–148.
Duan, J., & Li, X. (2007). Practising for resource integration of science and technology literature. *Information Sciences, 25*(8), 1230–1234.
Ernst, H. (1999). *Evaluation of dynamic technological developments by means of patent data. The dynamics of innovation* (pp. 103–132). Springer.
Eulaerts, O., Joanny, G., Giraldi, J., Fragkiskos, S., & Perani, S. (2020). *Weak signals in science and technologies - 2019 report*. EUR 29900 EN). Publications Office of the European Union.
European Communities. (2005). Frontier Research: The European challenge: high-level expert group report. publications office.
Fabry, B., Ernst, H., Langholz, J., & Köster, M. (2006). Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry. *World Patent Information, 28*(3), 215–225.
Fernandez, N. C., Lozier, A., Flament, C., Ricciardi-Castagnoli, P., Bellet, D., Suter, M., & Zitvogel, L. (1999). Dendritic cells directly trigger NK cell functions: cross-talk relevant in innate anti-tumor immune responses in vivo. *Nature medicine, 5*(4), 405–411.
Forti, E., Sobrero, M., & Franzoni, C. (2007). The effect of patenting on the networks and connections of academic scientists. In *Proceedings of the Acad. Manage. Conf., Philadelphia, PA*.
Furukawa, R., & Goto, A. (2006). The role of corporate scientists in innovation. *Research Policy, 35*(1), 24–36.
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.
Gerken, J. M., & Moehrle, M. G. (2012). A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. *Scientometrics, 91*(3), 645–670.
Getoor, L., & Diehl, C. P. (2005). Link mining: a survey. *Acm Sigkdd Explorations Newsletter, 7*(2), 3–12.
Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of 'reverse' citation relations. *Scientometrics, 58*(2), 415–428.
Guan, J., & He, Y. (2007). Patent-bibliometric analysis on the Chinese science—technology linkages. *Scientometrics, 72*(3), 403–425.
Guimerà, R., & Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences, 106*(52), 22073–22078.
Han, F., & Magee, C. L. (2018). Testing the science/technology relationship by analysis of patent citations of scientific papers after decomposition of both science and technology. *Scientometrics, 116*(2), 767–796.
Hashimoto, M., Im, S. J., Araki, K., & Ahmed, R. (2019). Cytokine-mediated regulation of CD8 T-cell responses during acute and chronic viral infection. *Cold Spring Harbor perspectives in biology, 11*(1), Article a028464.
He, F., Leyrer, S., & Kwang, J. (2016). Strategies towards universal pandemic influenza vaccines. *Expert Review of Vaccines, 15*(2), 215–225.
Hobernik, D., & Bros, M. (2018). DNA vaccines—how far from clinical use? *International journal of molecular sciences, 19*(11), 3605.
Hu, X. J., & Rousseau, R. (2018). A new approach to explore the knowledge transition path in the evolution of science & technology: From the biology of restriction enzymes to their application in biotechnology. *Journal of Informetrics, 12*(3), 842–857.
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.

Jaffe, A. B., & De Rassenfosse, G. (2019). Patent citation data in social science research: Overview and best practices. In B. Depoorter, & P. Menell. (Eds.), *Research handbook on the economics of intellectual property law*. Edward Elgar Publishing.
Jaffe, A. B., Fogarty, M. S., & Banks, B. A. (1998). Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation. *The Journal of Industrial Economics, 46*(2), 183–205.
Jin, Z. (2016). *How to ensure the safety of domestic vaccines—It takes 8 or even 20 years from research and development to market, depending on how to ensure the safety of domestic vaccines* https://epaper.gmw.cn/gmrb/html/2016-04/26/nw.D110000gmrb_20160426_2-06.htm.
Kostoff, R. N., & Schaller, R. R. (2001). Science and technology roadmaps. *Engineering Management IEEE Transactions on, 48*(2), 132–143.
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, Y., & Zeng, J. (2010). Research on integration framework between periodical thesis and patent literature. *Library and Information Service, 54*(4), 109–112.
LENS.ORG (2019). *About the Lens* Retrieved from https://www.lens.org/lens/search?q=stem%20cells&preview=true .
Leydesdorff, L. (2014). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science & Technology, 58*(9), 1303–1319.
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.
Lissoni, F. (2010). Academic inventors as brokers. *Research Policy, 39*(7), 843–857.
Liu, Z., Xu, H., Luo, R., Dong, K., & Zhu, L. (2019). Research on scientific and technological interaction patterns based on topic relevance analysis. *Journal of the China Society for Scientific and Technical Information, 38*(10), 997–1011.
Liu, Z., Zhang, Q.-M., Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A local naïve Bayes model. *EPL (Europhysics Letters), 96*(4), 48007.
Lopes, A., Vandermeulen, G., & Préat, V. (2019). Cancer DNA vaccines: Current preclinical and clinical developments and future perspectives. *Journal of Experimental & Clinical Cancer Research, 38*(1), 146.
Luo, R., Xu, H., & Liu, Z. (2019). Science and technology linked topic recognition through identification of bridging researchers. In *Proceedings of the 11th National Symposium on Scientometrics and Evaluation of Science and Education, Chongqing, China*.
Lv, L., & Zhou, T. (2013). *Link Prediction*. Higher Education Press.
Marinho, V. Q., Hirst, G., & Amancio, D. R. (2016). Authorship attribution via network motifs identification. In *Proceedings of the 2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE.
McCalman, P. (2001). Reaping what you sow: An empirical analysis of international patent harmonization. *Journal of International Economics, 55*(1), 161–186.
Mehta, R. S., & Rezvani, K. (2018). Chimeric antigen receptor expressing natural killer cells for the immunotherapy of cancer. *Frontiers in Immunology, 9*, 283.
Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy, 29*(3), 409–434.
Meyer, M. (2006). Are patenting scientists the better scholars?: An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. *Research Policy, 35*(10), 1646–1662.
Meyer, M., & Debackere, K. (2010). *Can applied science be 'good science'? Exploring the relationship between patent citations and citation impact in nanoscience*. Springer.
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.
Ministry of Science and Technology, Ministry of Education, Chinese Academy of Sciences, & National Natural Science Foundation of China (2017). *Notice on the issuance of the "Thirteenth Five-Year Plan" National Basic Research Special Plan* http://www.most.gov.cn/xxgk/xinxifenlei/fdzdgknr/qtwj/qtwj2017/201706/t20170608_133409.html.
Ministry of Science and Technology (2021). *During the "Thirteenth Five-Year Plan" period, my country's basic research funding has nearly doubled: A new leap in scientific and technological innovation capabilities* http://www.gov.cn/xinwen/2021-02/27/content_5589117.htm.
Morillo, F., Bordons, M., & Gómez, I. (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics, 51*(1), 203–222.
Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Research Policy, 31*(8-9), 1389–1403.
Narayanamurti, V., & Odumosu, T. (2016). *Cycles of invention and discovery: Rethinking the endless frontier*. Harvard University Press.
Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics, 7*(3-6), 369–381.
Niemann, H., Moehrle, M. G., & Frischkorn, J. (2017). Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application. *Technological Forecasting and Social Change, 115*, 210–220.
Noyons, E., Van Raan, A., Grupp, H., & Schmoch, U. (1994). Exploring the science and technology interface: Inventor-author relations in laser medicine research. *Research Policy, 23*(4), 443–457.
OECD (1970). *Proposed standard practice for surveys of research and experimental development. DAS/SPR/70.40*. Paris: Directorate for Scientific Affairs.
Pan, J., Zhang, M., Wang, J., Wang, Q., Xia, D., Sun, W., & Cao, X. (2004). Interferon-γ is an autocrine mediator for dendritic cell maturation. *Immunology letters, 94*(1-2), 141–151.
Patel, P. M., Ottensmeier, C. H., Mulatero, C., Lorigan, P., Plummer, R., Pandha, H., & Adams, S. E. (2018). Targeting gp100 and TRP-2 with a DNA vaccine: Incorporating T cell epitopes with a human IgG1 antibody induces potent T cell responses that are associated with favourable clinical outcome in a phase I/II trial. *Oncoimmunology, 7*(6), Article e1433516.
Patel, S., Burga, R. A., Powell, A. B., Chorvinsky, E. A., Hoq, N., McCormack, S. E., & Cruz, C. R. Y. (2019). Beyond CAR T cells: other cell-based immunotherapeutic strategies against cancer. *Frontiers in Oncology, 9*, 196.
Preschitschek, N., Niemann, H., Leker, J., & Moehrle, M. G. (2013). Anticipating industry convergence: Semantic analyses vs IPC co-classification analyses of patents. *Foresight, 15*(6), 446–464.
Ranaei, S., Suominen, A., & Dedehayir, O. (2017). A topic model analysis of science and technology linkages: A case study in pharmaceutical industry. *2017 IEEE technology & engineering management conference*.
Ren, Z., Zhan, s., & Fan, T (2015). Research of analysis and mining method based on integration of papers and patents. *Technology Intelligence Engineering, 1*(4), 88–94.
Rinia, E. J., Leeuwen, T. N. V., Bruins, E. E. W., Vuren, H. G. V., & Raan, A. F. J. V. (2002). Measuring knowledge transfer between fields of science. *Scientometrics, 54*(3), 347–362.
Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science, 59*(2), 504–525.
Rollier, C. S., Reyes-Sandoval, A., Cottingham, M. G., Ewer, K., & Hill, A. V. (2011). Viral vectors as vaccine platforms: deployment in sight. *Current opinion in immunology, 23*(3), 377–382.
Schmoch, U. (1993). Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics, 26*(1), 193–211.
Schmookler, J. (1950). The interpretation of patent statistics. *J. Pat. Off. Soc'y, 32*, 123.
Schmookler, J. (1953). The utility of patent statistics. *J. Pat. Off. Soc'y, 35*, 407.
Schmookler, J. (1972). *Patents, invention, and economic change*. Harvard University Press.
Steele, T. W., & Stier, J. C. (2000). The impact of interdisciplinary research in the environmental sciences: a forestry case study. *Journal of the American Society for Information Science, 51*(5), 476–484.
Stein, B., & Zu Eissen, S. M. (2004). Topic identification: Framework and application. *Proceedings of the International Conference on Knowledge Management, 399*, 522–531.
Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface, 4*(15), 707.
Stokes, D. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.
Tao, C., Zhang, Z., & Chen, Y. (2019). Few opinions on building the basic science research power of china. *World Sci-Tech R&D, 41*(1), 1–15.
Tijssen, R. J. (2002). Science dependence of technologies: Evidence from inventions and their inventors. *Research Policy, 31*(4), 509–526.
Tijssen, R. J. (2004). Is the commercialisation of scientific research affecting the production of public knowledge?: Global trends in the output of corporate research articles. *Research Policy, 33*(5), 709–733.

Tijssen, R. J. (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.
Tijssen, R. J., & Winnink, J. (2016). Twenty-first century macro-trends in the institutional fabric of science: Bibliometric monitoring and analysis. *Scientometrics, 109*(3), 2181–2194.
Tohalino, J. V., & Amancio, D. R. (2018). Extractive multi-document summarization using multilayer networks. *Physica A: Statistical Mechanics and its Applications, 503*, 526–539.
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.
Wang, G., & Guan, J. (2011). Measuring science-technology interactions using patent citations and author-inventor links: An exploration analysis from Chinese nanotechnology. *Journal of Nanoparticle Research, 13*(12), 6245–6262.
Winnink, J. (2017). *Early-stage detection of breakthrough-class scientific research: using micro-level citation dynamics*. Leiden University [PhD Thesis].
Xu, H., Winnink, J., Yue, Z., Liu, Z., & Yuan, G. (2020). Topic-linked innovation paths in science and technology. *Journal of Informetrics, 14*(2), Article 101014.
Xu, H., Winnink, J., Yue, Z., Zhang, H., & Pang, H. (2021). Multidimensional Scientometric indicators for the detection of emerging research topics. *Technological Forecasting and Social Change, 163*, Article 120490.
Xu, H., Yue, Z., Wang, C., Dong, K., Pang, H.-S., & Han, Z. (2017). Multi-source data fusion study in scientometrics. *Scientometrics, 111*(2), 773–792.
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.
Xu, X., Zheng, Y., & Liu, Z. (2016). Study on the Method of Identifying Research Fronts Based on Scientific Papers and Patents. *Library and Information Service, 60*(24), 97–106.
Zhang, B., & Ma, F. (2015). A Review on Link Prediction of Scientific Knowledge Network. *Journal of Library Science in China, 41*(3), 99–113.
Zhang, G., Liu, L., & Wei, F. (2019). Key nodes mining in the inventor-author knowledge diffusion network. *Scientometrics, 118*(3), 728–735.
Zhang, J. (2014). *Research on the linkage between science and technology*. Nanjing Agricultural University [PhD Thesis].
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 & Strategic Management, 25*(6), 707–724.
Zhang, Y., Huang, Y., Porter, A. L., Zhang, G., & Lu, J. (2019). Discovering and forecasting interactions in big data research: A learning-enhanced bibliometric study. *Technological Forecasting and Social Change, 146*, 795–807.
Zhang, Y., Porter, A. L., Hu, Z., Guo, Y., & Newman, N. C. (2014). Term clumping" for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting and Social Change, 85*, 26–39.
Zhang, Y., Wu, M., Miao, W., Huang, L., & Lu, J. (2021). Bi-layer network analytics: A methodology for characterizing emerging general-purpose technologies. *Journal of Informetrics, 15*(4), Article 101202.
Zucker, L. G., & Darby, M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences, 93*(23), 12709–12716.