Information Processing and Management 59 (2022) 102862

Contents lists available at ScienceDirect

Information Processing and Management

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

A methodology for identifying breakthrough topics using structural entropy

Haiyun Xu[a,*], Rui Luo[b], Jos Winnink[c], Chao Wang[d], Ehsan Elahi[e]

[a] Business School, Shandong University of Technology, Zibo, 255000, China
[b] Information Center, Jiangsu Academy of Agricultural Sciences, Nanjing, 210014
[c] Centre for Science and Technology Studies (CWTS), Leiden University, Leiden 2300 AX, the Netherlands
[d] Information Research Institute of Shandong Academy of Science, Jinan, 250014, China
[e] School of Economics, Shandong University of Technology, Zibo, 255049, Shandong, China

ARTICLE INFO

Keywords:
structural entropy
scientific breakthrough
link prediction
knowledge networks

ABSTRACT

This research uses link prediction and structural-entropy methods to predict scientific breakthrough topics. Temporal changes in the structural entropy of a knowledge network can be used to identify potential breakthrough topics. This has been done by tracking and monitoring a network's critical transition points, also known as tipping points. The moment at which a significant change in the structural entropy of a knowledge network occurs may denote the points in time when breakthrough topics emerge. The method was validated by domain experts and was demonstrated to be a feasible tool for identifying scientific breakthroughs early. This method can play a role in identifying scientific breakthroughs and could aid in realizing forward-looking predictions to provide support for policy formulation and direct scientific research.

Introduction

Scientific breakthroughs allow research to be channeled in directions that were previously inaccessible. Thus, such developments are highly innovative and represent the forefront of scientific research. The early identification of breakthrough topics is important for policy formulation and strategic management by governments, businesses, and other organizations. Identifying major breakthroughs early gives policy-makers ample time to react. Compared to incremental innovation, breakthrough innovation is more difficult to identify because it takes time to recognize a development and gather the information needed for analysis. In some cases, luck and intuition play a role in discovering this information, especially when serendipity is involved (van Andel, 1994; Merton et al., 2004; Fukawa, 2006; Seymour, 2009).

Studies have primarily explored the identification of scientific breakthroughs using qualitative methods that depend on expert judgment. However, regarding the background of interdisciplinary integration, relying solely on experts is excessively time-consuming and often leads to contradictory results. An abundance of powerful data-processing resources, analytical tools, and algorithms have emerged that provide effective support for, or alternatives to, expert opinion. Currently, most methods used to identify breakthrough topics rely heavily on specific attributes (e.g., recency, novelty, integrative, knowledge configurability, and word frequency). However, these methods fail to consider processes of gestation, propagation, development, and mutation.)

The early identification of scientific breakthroughs relies on a deep understanding of the laws of technological innovation.

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

https://doi.org/10.1016/j.ipm.2021.102862
Received 28 November 2021; Accepted 30 December 2021
Available online 10 January 2022

Scientific innovation is a nonlinear developmental process, and different knowledge topics interrelate to form complex networks. In these networks, newly emerging topics affect the structure of the original network (Chen, 2012, 2015; Dahlin & Behrens, 2005; Wan, 2017; Xu et al. 2019; Xu et al. 2021; Zhang et al. 2021). Thus, the greater the degree of innovation, the greater is the impact of the topic on the knowledge network structure, thus, scientific breakthroughs always affect the structure of knowledge networks (Luo, 2020). Therefore, evaluating their impact on knowledge-network structures can identify potential breakthrough topics. Structural entropy, considers the number of communities and their sizes, to encapsulate a richer representation of the network's structure into a single value (Almog et al., 2019), conversely, it regards the network holistically and dynamically as a complex system by considering the evolutionary processes (Guan, 2014; Huo, 2019; Xu et al. 2019; Xu et al. 2021), therefore, it would help to identify scientific breakthroughs during their early stages. The increase in the network structure's entropy can be an indicator of the network's evolution, in addition, the growth rate is high during early network evolution but as it grows, the network structure stabilizes, and the growth rate decreases (Luo et al., 2013).

In this study, we attempt to combine link prediction and structural entropy methods to predict scientific breakthrough topics. Compared with extant prediction indicators, the structural entropy index regards the knowledge network as a complex system from a holistic perspective. This method is used to recognize emerging topics that significantly impact the knowledge-network structure and regards them as potential breakthrough topics. Finally, the prediction results of our proposed method are evaluated by domain experts to assess their validity and reliability.

Our objective is to provide a method that allows the detection of scientific breakthroughs in their early stages. The method is based on tracking and monitoring critical points in the evolution of complex knowledge networks to identify potential breakthrough topics. The moment at which a significant change in a network's structural entropy occurs i.e., the tipping point, can be used to denote the point at which breakthrough topics emerge. Therefore, unlike the extant prediction indicators, the structural entropy index holistically regards the knowledge network as a complex system. Our proposed method considers the mechanism of the emergence of scientific breakthroughs from a process perspective, and we believe this is beneficial for the early identification of scientific breakthrough topics. In this respect, our method stands out from most current research as they are inclined toward extant hot-topic monitoring and are not forward-looking.

The remainder of this paper is organized as follows. First, this study details the characteristics of scientific breakthroughs and reviews extant prediction methods. Second, we explain the principles and processes of constructing the prediction model based on structural entropy and link prediction. Then, the field of genetically engineered vaccines (GEV) is used as a testbed, and the prediction results are compared with expert evaluations. Finally, we sum up the advantages and disadvantages and elaborate on future research directions.

The highlights of this study are

Identifying a scientific breakthrough early and helping to establish forward-looking predictions.
Depicting the non-linear characteristics of complex knowledge networks through structural changes.
Regarding the knowledge network as a complex system from a holistic perspective.
Observing the incubation mechanism of emergent scientific breakthroughs from a dynamic evolutionary perspective.

Theoretical framework

This section introduces the theoretical framework of our research. This framework is based on the existing scientific literature on scientific breakthroughs and on the methods for identifying scientific breakthroughs gleaned from the literature. We also classify scientific breakthroughs based on the knowledge network to which they belong. In addition, a new method is introduced that should overcome the shortcomings of existing methods.

The concept of a scientific breakthrough

No single definition of the concept of a scientific breakthrough is agreed on by all scholars. Scientific breakthroughs are often related to scientific revolutions and changes in scientific paradigms (Kuhn, 1962; Fortunato et al., 2018; Min, Bu, & Sun, 2021). Breakthrough developments are considered to overcome obstacles for further scientific and technical progress. Scientific breakthroughs may produce new theories or may improve existing ones (Merton, 1973; Wray, 2011). The cognitive structure of a scientific field can change and research can expand into new scientific fields thanks to new insights resulting from a scientific breakthrough. The concept of revolutions in science is incompatible with the current dominant theoretical framework about how scientific research develops (Andersen et al., 2006; Luo et al. 2021). Conceptually, this study uses a breakthrough as a signal of revolutionary scientific development. The focus is on fundamental scientific research on which new technological developments are based that can serve to produce new (consumer) products. Koshland (2007) introduced the Charge-Chance-Challenge theory in which he divides scientific discoveries into three types: charge, chance, and challenge, the latter being radically discoveries that cause scientific revolutions and subvert the existing theoretical framework in a field of science. Wuestman et al. (2020) concluded that instead of distinguishing between charge, challenge, and chance types, breakthroughs could better be understood as being question-driven or research object-driven, i.e., introducing a new question/research object or a known question/research object and having a contribution that contradicts or supports state-of-the-art literature. The notion of a scientific breakthrough is usually based on new scientific principles or technical means that emerge that lead to the future development of science and technology (Jiang et al., 2017; Li et al., 2016), is

Table 1
Scientific breakthrough identification and prediction methods.

Methods	Method description	The mechanism and advantages	Limitation	Representative study
Expert judgment	Expert consultation and evaluation	Use the knowledge and experience of experts to identify and predict the development-trend of science and technology. Experts have a relatively complete knowledge system and experience.	Strong subjectivity; under the background of interdisciplinary integration, the efficiency is low and the accuracy is not strong.	Manyika et al. 2013; Thomson Reuters (2015); MIT Technology Review (DeepTech, 2020); 2019 Science Development Report (Chinese-Academy-of-Sciences, 2020)
Citation Network Analysis	Based on the citation relationship between the documents, a citation network is constructed to analyze the co-citation network or the coupling network.	The citation relationship reflects the inheritance and development of domain knowledge, can reflect the development of the domain, and help predict the direction of the domain.	There is a hysteresis. Only when the literature reaches a certain number of citations can it get attention, therefore, it is easy to ignore potential research frontiers.	Small, 1977; Dahlin & Behrens, 2005; Small et al. 2014; Schneider & Costas, 2017; Wang et al. 2017; Staudt et al., 2018; Winnink, 2017; Min et al. 2019. .
Topic mutation monitoring	Pay attention to the changes of topic terms and outlier data, analyze the changes of sudden bursting, and find breakthrough frontiers.	Pay attention to the mutations and sudden phenomena of knowledge units (topic words). The research object is topic words, with fine-grained analysis, more dynamic and time-sensitive.	It is easy to ignore the semantic connection between documents and topic terms, and it cannot reveal the overall knowledge structure of the subject field well.	Kleinberg, 2003Yoon & Kim, 2011; Luo et al., 2019; Jia et al., 2021.
Sleeping beauty literature analysis	In-depth analysis of sleeping beauty papers and patents, searching for potential breakthrough research that may be hidden in such literature	Pay attention to the wake-up mechanism of sleeping beauty literature. Improve the phenomenon of delayed recognition in the field of science and technology and shorten the recognition time lag for breakthrough innovations.	The sleeping beauty literature is only a small part of the scientific literature and can only be used as a supplement to identify scientific breakthroughs. Most of the breakthroughs attract attention in a relatively short time.	Palomeras, 2003; van Raan, 2017; Du 2017; van Raan & Winnink, 2018; Vanhoucke & Batselier, 2019, .
Machine learning algorithm model	Using multi-source data, with the help of machine learning algorithm models, predict the development trend of topics.	It can efficiently process large amounts of complex data and optimize the parameter model of the algorithm.	It takes many experiments to find a suitable algorithm model, and it needs the support of domain expertise.	Wolcott et al., 2016; Mao et al. 2019; Xu & Wang 2019; Eulaerts et al., 2019; Joanny et al., 2019; Yang et al., 2020; Wang et al., 2021; Luo, 2020; Luo et al. 2021; Liang et al., 2021.

nonlinear and may cause discontinuities (Kuhn, 1962; Liang et al., 2017). Scientific breakthroughs are difficult, if not impossible, to predict (Fu & Zhang, 2004; Winnink, 2017), but the impact of scientific breakthroughs can even be observed shortly after they occur (Winnink, 2017; Winnink et al., 2019; Min, Bu & Sun, 2021).

Dynamics of science and technology

The way knowledge is produced, organized, and disseminated in science, technology, social science, and humanities are fundamental issues in the dynamics of science (Coccia, 2020a; Gibbons et al., 1994). Multiple studies have proved that the dynamics of science is dependent on manifold factors, including historical, institutional, political, and research contexts, such as mental ability, the existing status of culture, institutions, and research funding (Börner et al. 2012, Coccia et al. 2015, Coccia, 2018, 2020b,a,). Scientometric mapping methods (Boyack et al. 2005) showed the hierarchy of the Sciences using bibliometric evidence (Fanelli, & Glänzel, 2013). Roshani et al. (2021) analyzed the relationship between research funding and citation-based performance.

The use of mathematical models to study the structure and dynamics of science and technology systematically has a long tradition in scientometrics (Börner et al. 2011). Tria et al (2014) proposed a mathematical model to predict statistical laws for the rate at which novelties happen. Coccia and Finardi (2012) applied exponential models of growth to measure the rate of scientific and technological advances of some emerging nanotechnological research fields in biomedicine to detect path-breaking technological trajectories.

Extant identification methods

Many scholars have studied how to identify and predict breakthrough innovations from different perspectives. The main commonly used quantitative methods for detecting groundbreaking research frontiers are citation analysis, topic mutation analysis, sleeping beauty literature analysis, technical evolution methods, and analyses based on machine learning models. Table 1 shows the characteristics and the advantages and disadvantages of the different methods.

Extant methods for identifying breakthroughs in science and technology

As shown in Table 1, considerable effort has been expended in identifying and predicting scientific breakthroughs. These studies provided insights into the characteristics of scientific breakthroughs, that could also become the basis for breakthrough innovations and would reveal the characteristics of such innovations. However, most research is conducted with a retrospective view and, consequently, has a limited capability of predicting future developments.

First, existing prediction methods generally rely on the judgment of experts. Expert judgment and other methods based on difficult functional elements require field experts to decompose and interpret actual developments and issues. Citation analysis does not comprehensively investigate the innovativeness of scientific papers. Works by prominent authors tend to receive a disproportionate number of citations in contrast to the limited attention given to publications by relatively unknown researchers covering the same topics (Savov et al., 2020). Because text-processing technology has not yet adequately attained the ability to judge, it has not completely surpassed the benefits of human interpretation.

Second, there is still a lack of research on the process mechanics of incubating emergent scientific breakthroughs. Most existing identification methods focus on the relationships among specific characteristics and driving factors that may be related to scientific breakthroughs, such as novelty, knowledge reorganization, delayed recognition, and knowledge intersection. The characteristics of signals that are used for early breakthrough publications, e.g. frequency of keyword changes and the intersection of multiple knowledge items, are commonly leveraged (Yang et al., 2020; Wang et al., 2021). These methods identify existing patterns and make predictions using only the most common highly associated breakthrough-innovation phenomena. Thus, these studies fail to distinguish the direct causes or indirect phenomena leading to scientific breakthroughs. A new and more advanced method should also include analyses of the dynamic mechanisms. Similarly, most extant recognition methods draw judgments based on the current states of their topics. Unfortunately, this has led to several studies that ignored the evolutionary processes and dynamic changes of breakthrough frontier topics.

Third, most methods identify hot topics during the rapid development or maturity phases of a research field rather than during its early stages. This leads to fewer forward-looking results and fails to meet the needs of decision-makers and scientists in forecasting (Liang et al., 2021). Many scientific breakthroughs display few signs to measure during their early stages, and the earlier the stage, the less information is available. Meanwhile, the entire innovation process is accompanied by many uncertainties, and the information carried by the early signals is especially fuzzy. In short, current studies fail to capture those early signals from the emerging stages of frontier-topic growth, and this leads to critical delays in recognition.

The influence of breakthroughs on the structure of knowledge networks

From the perspective of network dynamics, the entry of this new information influences the structure of the cognitive network into which it is induced. Also, the stability of the network may be influenced. According to Chen (2012), the impact of new knowledge on the original knowledge structure can be measured by the degree of structural change. Dahlin and Behrens (2005) assessed the structural differences of patent-citation networks to identify scientific breakthroughs, noting that the greater the difference, the higher the likelihood of a breakthrough, and they verified this through an analysis of tennis-racket patents. Petzold et al. (2019) described the process of generating transformative influence over time, based on the process perspective. Scientific breakthroughs impose particular influences on the knowledge structure of science and this observation is useful for predicting scientific breakthroughs (Min, Bu, Wu, et al. 2021). Emergent technologies appear in a technology environment already structured on pre-existing technologies (Arthur, 2007; 2009). The effects of an innovation depend both on its technological impact, and on how it adapts to the prevailing trajectory. Funk and Owen-Smith (2017) explored the impact of new technologies on prior technologies and the entire technological trajectory using network analysis. Chen et al. (2009) found that potentially important scientific discoveries mostly appeared where scientific knowledge from different fields intersects. Chen et al. (2015) determined the novelty of new knowledge carriers (i.e., single research papers) on the resultant change rate of network modularity, the centrality of the entire knowledge network, and changes in the numbers of connections created by the newly published papers. Changes in knowledge structures that represent novelty were used to discover transformative innovations. Min, Bu, Wu, et al. (2021) propose a citing-structure perspective to observe how breakthrough research progresses from changes in the knowledge structure.

Winnink (2017) and Winnink et al. (2019) discovered that the impact of scientific breakthroughs can be observed in the short term. They paid attention to the period within 3 years of a paper's publication and focused on the dynamic impact of publications upon the scientific community. They found that the diffusion of scientific breakthroughs was usually rapid, and the network of cited papers showed that close cooperation among a group of researchers working in the same subject area had a cumulative effect (Winnink et al., 2016). The combination of key papers, their bridging researchers, and the resultant interdisciplinary influences can, therefore, be used to track the development of potential "Charge" breakthroughs. The changes in the number of knowledge-network communities stem from the changes of influential subjects (Winnink & Tijssen, 2014).

The evolution of network structures

'Weak ties' refers to the early characteristics of interdisciplinary and technological integration. In a network structure, a weak tie refers to the type of node relationship of which the strength is lower than a given threshold. Thus, it is often considered to be the opposite of a strong tie. Granovetter (1983) demonstrated that a strong tie could maintain a relationship within an organization, whereas a weak one could link different groups and organizations for information transfer. Weak ties, in turn, can promote structural cohesion in larger networks. Many frontier topics initially appear as weak and interdisciplinary relationships (Wei et al., 2016), which can be used to describe the initial features of interdisciplinary and technological integration. Therefore, changes in strong and weak ties can both be used to capture the growth of emerging-research areas (Xu et al., 2020). Effective methods must be found to assess weak ties to improve the accuracy of early recognition using high-order network models (Liu et al., 2021). Zhang (2011) detected the evolution of weak ties in networks from whole- and individual-network perspectives and identified possible research frontiers. Wei et al. (2016) did the same for interdisciplinary information-science topics using a co-word network that adopted weak-tie analysis. Yoo and Won (2018) explored the dynamic behavior of technological innovation through the weak ties between keywords and predicted future technological trends in the field of nanomaterials. Min et al. (2019) proposed a citing-structure perspective for observing how breakthrough research unfolds among the knowledge-structure variations, finding that, compared with less ground-breaking papers, high-breakthrough papers were more influential regarding network structure. Xu et al. (2019) and Xu et al. (2021) analyzed the

Table 2
Structural entropy indicators.

Type	Principle	Limitation
Degree distribution entropy (Costa et al., 2007)	Node degree distribution	Local entropy
Entropy of redundancy degree distribution (Solé & Valverde, 2004)	Node redundancy	
"Wu" structural entropy (Tan & Wu, 2004)	The edges number of the node	
"Cai" structural entropy (Cai et al., 2011)	The differences between node and edge	The nonlinear characteristics in complex systems ignored.
Betweenness entropy (C. Wang et al., 2015)	It is based on betweenness centrality of nodes.	Based on resistance to destruction, not necessarily applicable to knowledge networks
Nonadditive entropy (Zhang, 2017)	It uses the nodes betweenness as the nonadditive adjustment parameter of the overall complexity of the network	Ignore the consideration of edges in the network
Dynamic network structural entropy (Guan, 2014)	Measurement of structural entropy of dynamic network based on power law growth and selective connection or exponential growth and selective connection	It is needed to determine the growth model of the network in advance
K-order structural entropy (Huo, 2019)	he total number of nodes that is reachable to a node within K steps	The law of structural entropy with the change of K and the association between Hd and minH need to be further explored.

uncertainty of future emerging research topics by analyzing the structural changes in the topic network to determine whether the topic became increasingly important throughout the entire knowledge community.

Knowledge networks and structural entropy theory

We argue the proposition that (radical) changes in entropy signal breakthroughs. Scientific breakthroughs change the structure of the citation network of publications, but there may be other mechanisms that result in entropy changes that are not ‘breakthroughs’. We focus on 1) the scientific knowledge network, 2) the relationship between the state of the knowledge network and structural entropy theory, 3) structural entropy indicators and 4) on using link prediction to predict the evolution of the structure of the knowledge network.

The development of scientific knowledge is a dynamic and nonlinear process (Li & Xu, 2007), and “breakthrough” refers to a qualitative change in nonlinear developments. The breakthrough process can be regarded as a knowledge system mutating by the advent of new knowledge units. When new knowledge units enter the network, they are amplified rapidly through nonlinear chain reactions, leading to knowledge mutation (Wan, 2017). To uncover emerging trends in a research field using a process perspective, the structural change in the knowledge network must be understood before and after breakthroughs.

Entropy is traditionally used to measure the mutation of a physical quantity. Thus, structural change (mutation) can be reflected in the structural entropy value. Clausius used entropy to measure changes in a system’s degree of order (Clausius, 1865). Increasing entropy indicates that the system is approaching a more chaotic and disordered state, and vice versa (Ping, 2014). Entropic change detection has been used in many complex systems for alerting purposes, such as for power faults (Cai et al., 2015), information security alerts (Dong et al., 2019), and system structure changes (Ma et al., 2014). One category is based on entropy information-theory (Costa et al., 2008) e.g., information-search, target, hidden-information, received-information, and exchanged-information. The entropy theory shows the enormous potential in the proposal of new forecasting approaches (Popkov, 2019; Vanhoucke & Batselier, 2019). The other focuses on the structural characteristics of networks (Solé et al., 2004) according to network connection distributions, and includes degree-distribution entropy, network-structural entropy, and redundancy-degree entropy.

By definition, a research field’s knowledge network contains all related knowledge topics and their interrelationships. Structural entropy is then systematically used to observe the changes in such relationships at different stages. The more ordered the network structure, the smaller is the structural entropy and vice versa. The uniformity status of a network implies an equal probability of node-to-node linkages, and this equality indicates the highest structural entropy accompanied by high uncertainty (Guan, 2014; Huo, 2019). Thus, the emergence of a scientific breakthrough can be regarded as the process of an ordered knowledge state transitioning from a lower to a higher level (Ping, 2014). During this process, some topics accelerate the increase of positive entropy, and others accelerate the increase of negative entropy. Monitoring the dynamics of structural entropy over time can lead to significant revelations about the underlying processes in the system (Almog et al., 2019).

Luo et al. (2013) proved that the structural entropy growth rate could reflect the structural changes of a network through simulation experiments. During the initial stages of network evolution, the structural entropy growth rate is relatively high. As the network scale gradually increases, the network structure stabilizes, and the increase in structural entropy gradually diminishes. This means that the growth rate can also reflect the structural features of the complex network. The more stable the network, the less the entropy growth rate increases. Additionally, the faster the network changes, the faster the rate of entropy increases. Hence, scientific

Table 3
Similarity measures used in link prediction methods.

Methods	Similarity measures
Weighted common neighbor (WCN)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} w_{xz} + w_{zy}$
Weighted Salton	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{\sqrt{s(x)s(y)}}$
Weighted Jaccard	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{s(x) + s(y) - w_{xy}}$
Weighted Sorenson	$s_{xy} = 2 * \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{s(x)s(y)}$
Weighted Hub Promoted Index (WHPI)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{\min\{s(x), s(y)\}}$
Weighted Hub Depressed Index (WHDI)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{\max\{s(x), s(y)\}}$
Weighted Leicht-Holme-Newman (WLHN)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{s(x)s(y)}$
Weighted Preferential Attachment (WPA)	$s_{xy} = s(x)s(y)$
Weighted Adamic-Adar (WAA)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{\log(1 + s(z))}$
Weighted resource allocation (WRA)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{w_{xz} + w_{zy}}{s(z)}$
Weighted local path (WLP)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} (w_{xz} + w_{zy}) + \varepsilon * \sum_{(i,j)\in l_{x\to y}} (w_{xi} + w_{ij}) * (w_{ij} + w_{jy})$
Weighted Resource allocation along local path (WRALP)	$s_{xy} = \sum_{z\in\Gamma(x)\cap\Gamma(y)} \frac{(w_{xz} + w_{zy})}{s(z)} + \varepsilon * \sum_{(i,j)\in l_{x\to y}} \frac{(w_{xi} + w_{ij}) * (w_{ij} + w_{jy})}{s(i) * s(j)}$

Note: $\Gamma(x)$ is the neighbor of node x, $k_x = |\Gamma(x)|$ is the degree of node x, k_z is the node degree of the common neighbor node z; s_{xy} is the similarity, w_{xz} is the weight of the connection between nodes x and z, and s(z) is the strength of node z.

breakthroughs bring significant changes to knowledge networks This will be reflected in the structural entropy that can be used to measure the overall structural change in the knowledge network to identify the generation of scientific breakthroughs. Luo et al. (2021) used the topic-terms co-occurrence network to characterize the knowledge network, employed structural entropy to measure the state of the knowledge network, and explored whether topics that have a greater impact on the topology of the knowledge network are potential scientific breakthrough topics.

Different methods for measuring structural entropy have applicable scenarios and concerns, that all derive from the different solutions of the discrete probability set from step one (Table 2).

The non-additive characteristic determines that the entropy of the system, $S_q(A+B)$, meets the condition of Eq. (1) when it consists of two independent subsystems, A and B, with different values of q, corresponding to super-additive (q<1), additive, and sub-additive states (q>1) (Cao & Wang, 2005). The knowledge network is a type of complex network, and the impact of each new node and edge on the network is non-additive and non-linear, producing the effect of 1+1>2. When entropy is applied to information theory (Zhang, 2017).

$$S_q(A+B) = S_q(A) + S_q(B) + (1-q)S_q(A)S_q(B) \tag{1}$$

Material and methods

Material

This study selected genetically engineered vaccines (GEVs) as the experimental field. Research into GEVs is expected to be the source of innovative vaccines. The scientific impact is confined to academic impact, ignoring societal, economic, and other effects. Therefore, instead of analyzing patents and other information journal papers were chosen as the empirical data. Scientific papers from the GEV field were collected to provide the analysis datasets. The Web of Science database was selected to search scientific papers. To select scientific publications in the WoS database, we used the following query:

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

The data was collected in December 2019 and the publication range was from 1 Jan 1940 to 31 Dec 2018. The resulting dataset consisted of 4,196 scholarly publication records.

Methods

Using scientific knowledge networks to explore scientific dynamics

Science can be seen as an evolving system of diverse basic units of science that are tightly linked and dynamically coupled. Networks can represent the collective, self-organized emerging structures in science, and allow the linking of structural properties to dynamic processes (Börner et al. 2009). The scientific knowledge network is a complex system, including three types of networks: ontology networks, knowledge subject networks, and knowledge carrier networks (Wang & Zhao, 2014). Fortunato et al. (2018) described scientific development as a complex, self-organizing, and evolving network comprised of scholars, research projects, papers, and ideological systems. The ontology network is constituted by the co-occurrence of topic terms, the subject network is constituted by the co-occurrence of knowledge subjects such as countries, institutions, and authors, and the carrier network is constituted by the citation relationship of journals, documents, and authors.

Leydesdorff et al, (1993) who used journal-journal citations to attribute journals to specialties tracked the changes in the disciplinary structure of science by using the differences among the multi-variate analyses for the various years. Valverde et al (2007) applied patent citations to study the topology and evolution of technology innovation networks and noted that the scaling behavior exhibited by the network is consistent with a preferential attachment mechanism together with a Weibull- shaped aging term. Bettencourt et al (2009) noted that, as a field develops, it undergoes a topological transition in its collaboration structure from a small, disconnected graph to a much larger network, and analyzed this qualitative change in network topology using several quantitative graph theoretical measures. Adams, (2012, 2013) analyzed papers from the past three decades and found that the best science arose from international collaboration as well as that new collaboration patterns are changing the global balance of science.

Using link prediction to predict the evolution of a network

Table 4
Discrete probability set and non-extended parameter.

Object	Discrete probability set	Nonadditive parameter
node	Degree using node i:p (i)=$\frac{\mathrm{Degree(i)}}{\sum_{i=1}^{n}\mathrm{Degree(i)}}$	qi=1-(v(max)-v(i)) v(max) is the maximum value of betweenness of nodes in the network, and v(i) is the betweenness value of node i
edge	Weights using edge j:p (j)=$\frac{\mathrm{Weight(j)}}{\sum_{j}^{n}\mathrm{Weight(j)}}$	qj=1-(l(max)-l(j)) l(max) is the maximum betweenness of edges in the network, l(j) is the betweenness value of the edge j

Table 5
Calculation algorithm of structure entropy.

Item	Function
node	$N_q^{'}=-k\sum_{i=1}^{N}\frac{(p_i^{q_i})-p_i}{1-q_i}$
edge	$L_q^{'}=-k\sum_{j=1}^{N}\frac{(p_j^{q_j})-p_j}{1-q_j}$
overall	$S_q^{'}=N_q^{'}+L_q^{'}$

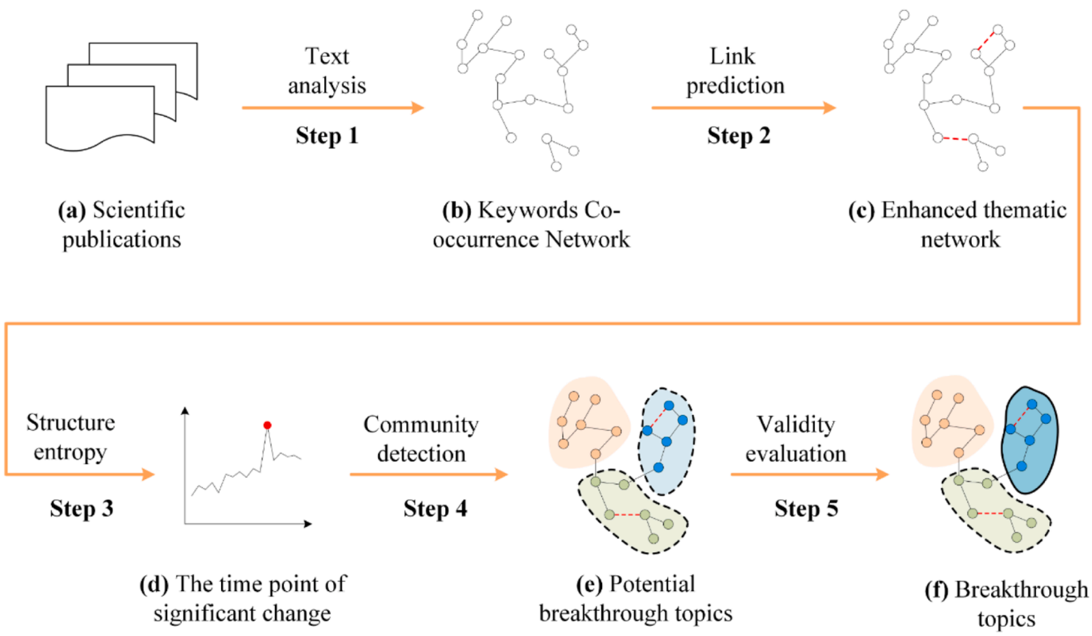

Fig 1. Technique flow chart. Note:Dot nodes in b, c, e and f represents for scientific topic terms, solid black line in b, c, e and f represents the extant connection of the scientific topic terms, the red dotted line represents links from linkage prediction. Topic encompassed in the dotted line circle are potential breakthrough topics (e) and in the solid line is the validated breakthrough topics (f).

Link prediction can be used to forecast the potential connections through the nodes and edges of an extant network. In the field of scientific research, link prediction is often used to predict knowledge structures and e.g. has been applied in knowledge-transfer research (Li, 2020). The topic terms that constitute the future topics are often present in the network in the form of weak linkages. Therefore, link prediction can reinforce the semantic association of topic terms in networks to enhance the accurate prediction of future topics.

Most current link prediction methods were developed for unweighted networks (Zhou, 2015), and only a few consider weights (Lü, 2010). The current widely used link prediction methods are frequently based on various network similarities. These methods can be roughly divided into three categories: similarity algorithms based on local node information represented by a common neighbor; path-based similarity algorithms represented by a local path; similarity algorithms based on a random walk represented by the average commute time; and SimRank (Huang, Zhu, & Zhang, 2019).

Different link prediction methods apply to different types of networks. Empirical training tests are required to select the applicable link prediction algorithm in practical applications. Table 6 lists the extant representative link prediction methods (Lü & Zhou, 2010; Meng, Ke, & Yi, 2011; Yue, Xu, & Wang, 2020) considering weighted local similarity. The meanings of the related formula symbols are listed in Table 3.

Measuring the relevant variables/parameters

First, the probability distribution for the linkage degrees of the nodes is obtained. Then, the betweenness centrality of the nodes is used to adjust the structural characteristics of the knowledge network eventually obtaining the nonadditive structural entropy. Changes in the connections between nodes contribute to the overall dynamic change. Therefore, the network complexity of the whole network must be measured each time a node is added to the network. In this study, the structural entropy index considers both node and edge distributions, as does non-additive entropy. Non-additive entropy is based on node-betweenness centrality and adjusts each node according to the overall effect. Then, it adjusts the nodes based on their contribution to the network structure. This method covers the whole network ignoring the differences in node connections. This study also considers the edges in support of node-based non-additive entropy. The measurement of the edges' structural entropy is similar to the process of measuring nodes. The metrics of both rely on non-additive entropy and are integrated into a holistic measurement index. Based on the work of Cai et al. (2011), we treat the weight of nodes and edges as equal and combine their structural entropies to obtain the 'Cai' entropy, which represents the complete network.

(1) Discrete probability set and the non-extended parameter

The degree of node distribution is taken as the discrete probability set for node entropy, and the weight distribution of edges is taken as the discrete probability set for edge entropy, where the weight refers to the co-occurrences between topic terms. After acquiring the discrete probability set, we normalize the betweenness centrality and take it as a non-additive parameter. The discrete probability sets are then adjusted according to this parameter (Table 4).

(2) Calculation of structural entropy

Table 5 shows the algorithm for calculating structure entropy

Approach and Data analysis

This section explains the construction method, the indicators, and how to use structural entropy for identifying and predicting breakthrough topics. First, the link prediction method is used to predict the potential co-occurrence network of topic terms. Link prediction would forecast the potential connections through the nodes and edges of an extant network, which also reinforces the semantic association of topic terms in networks. Then, structural entropy is used to screen out candidate scientific breakthrough topics based on their impact on the state of the knowledge network.

The recognition process includes five steps discussed below. Figure 1 is the technique flow chart of the method.

Step 1: Knowledge network construction using the co-occurrence of topic terms

In this study, the topic-term co-occurrence network is used to represent the knowledge network. This study assumed that significant changes in the structure of knowledge networks were significant indicators of scientific breakthroughs. By employing data cleansing and extracting topical terms using natural language processing technology, the network was prepared for setting the thresholds for topic-term filtering and to assemble the topic-term co-occurrence network.

Step 2: Semantic enhancement by link prediction

Because the topic-term co-occurrence network in this study is weighted, the weights of edges are vital considerations in the calculation of structural entropy. Thus, multiple weighted local similarity link prediction methods are applied to the network data for testing. The link prediction algorithm with the highest performance value is selected for prediction.

Step 3: Using structural entropy to identify the moment of significant change

When measuring structural entropy, each newly entered node and edge will affect the community partition of the knowledge network. When the structural-entropy value increases, the knowledge network will be in a *stuck* state between the prior and emergent paradigms. Conversely, when the value of structural entropy decreases, one of the two paradigms (prior or emerging) will be favored. When the temporal entropy value of the knowledge network enters a *steady* state, the knowledge begins to converge, indicating the emergence of a scientific breakthrough topic.

To identify breakthrough scientific topics, this study takes the time point at which the structural entropy of the entire network changes significantly as the key breakthrough moment. A significant change in the structural entropy indicates that the state of the knowledge network has also changed significantly and that a scientific breakthrough is probable. To find these key points, the structural entropy of the network is measured over time to map the temporal structural entropy function, from which the key points of state transition will be apparent. New knowledge, which causes a huge change in the network structure, may signal potential scientific

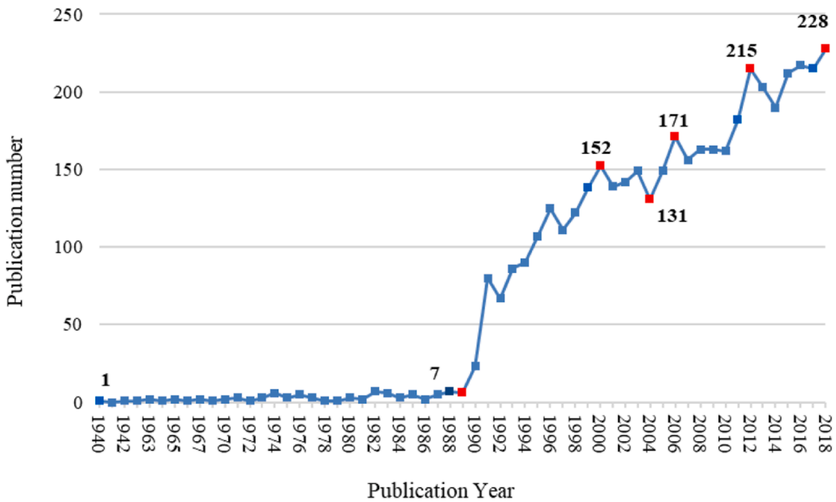

Fig 2. Scholar publication trend in genetic engineering vaccines research field.

breakthroughs.

Step 4: Identifying breakthrough topics through community detection

A temporal topic term co-occurrence network is then used to represent the changing network, and the network community division method is used to identify clearly the potential topics. The specific community detection approaches should be selected according to target fields and their data characteristics. Various topic detection methods from text have been developed and Zhang et al. (2018) review the frequently used topic extraction methods including topic models, word co-occurrence method, and the novel "kernel k-means clustering method with word embedding". All these methods could be used for identifying topics in this study.

Then, we define and calculate the structural entropy impact to narrow the range of topics to a clear field of innovation. The method of calculating topic impact by structural entropy is as follows:

$$Topicimpact = |originalstructuralentropy - newstructuralentropy| \tag{2}$$

Note the absolute value of Eq. (2), the topics that promote an increase in both positive and negative entropy are accounted for. Knowledge nodes that promote the state transition of the network will resort among the candidate breakthrough-innovation topics. Depending on the network size and the degree of topical influence, the number of candidate topics will differ. In this study, the mean value is used as the screening criterion of structural entropy. If the impact value is higher than the mean value, the topic will be selected.

Step 5: Validity evaluation

For those potential scientific-breakthrough topics, this study combines the quantitative analysis of authoritative journals and scientific awards with expert consultation to assess the proposed model's recognition results. The quantitative metrics' results are compared with the results of our proposed model to evaluate the model's validity.

As noted by Bolger & Wright, (2017), there are still no repetitive phenomena or events that can be used as a reference point for improving the accuracy of detection and forecasting. Therefore, it is almost impossible directly to compare detected and forecasted emerging technologies with realized outcomes. From our review of the literature, this situation is relatively unchanged and the point is also applicable to our case.

There are two main ways to verify the validity of the method. One is to verify after an interval, such as Apreda, Bonaccorsi, dell'Orletta, & Fantoni (2019) who re-examined the findings of a technology foresight exercise on the medical-device industry with realized technologies. Huang et al. (2019) compared recent technological activities with previous forecasts to reveal the influencing factors that led to differences between past predictions and results. They found that the data quality, forecasting methods, complexity, and dynamism of future-oriented analyses can cause differences between preceding predictions and actual performance.

The other is to perform empirical analysis and do the validation analysis contemporaneously, such as Zhang et al. (2021) who combined expert knowledge from an expert panel and empirical evidence from the literature in their empirical evaluation of candidate emerging technologies.

To verify the validity of the method, the contemporary validation approach was applied to compare the prediction results with the predictions of experts in the domain of genetically engineered vaccines. For these potential breakthrough topics, we combined authoritative journals, scientific awards, and expert consultations to validate the identification results.

Results

Data processing

Data preprocessing

After removing duplicate data, 4,374 records were retained. Figure 2 displays the statistical analysis of these data. The number of papers in the GEV field increased after 1990, until entering a steady growth state in 1993, recently reaching its publication peak. From 1989 to 1991, the field experienced especially rapid growth. From an additional examination of titles and keywords, we found that, for papers newly published in 1991, the main GEV research focused on "pseudorabies virus", "hepatitis-B vaccine", "HIV" and "HPV". Other research efforts increased, including those into "virus recombination," "antibody recombination, "cell proliferation," and others. Taking the "pseudorabies virus" topic as a notable example, in 1991 Van Oirschot invented a gene deletion vaccine using the gE-/gI-Tk three-gene deletion strain, resulting in breakthrough research (Van Oirschot et al., 1991). These activities explain the substantial increase in the number of related papers in 1991.

Considering that all the extant knowledge could become bases of breakthrough topics, we divided the data into time slices using a cumulative approach that included all topics at the current time (Appendix Table A1).

Table 6
Data in Cumulative Schematic.

Topic term	Topic term	Weight
genetic engineering	monoclonal antibodies	7
escherichia coli	genetic engineering	6
antibody fragments	escherichia coli	4
genome replication	hepatitis delta virus	4
antigenomic RNA	hepatitis delta virus	4
……	……	……

Table 7
Average of AUC.

Indexes	AUC Avg.	Indexes	AUC Avg.
WCN	0.819	WHDI	0.767
WSalton	0.848	WLHN	0.630
WJaccard	0.823	WPA	0.733
WSorenson	0.824	WAA	0.828
WHPI	0.819	**WRA**	**0.850**

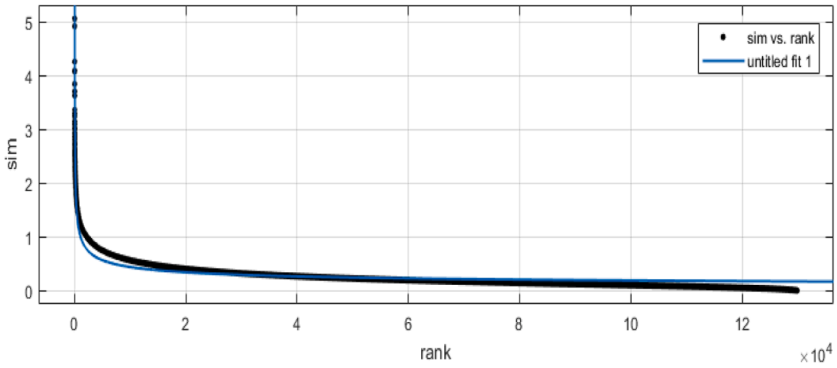

Fig 3. similarity scatter of the new knowledge connection predicted.

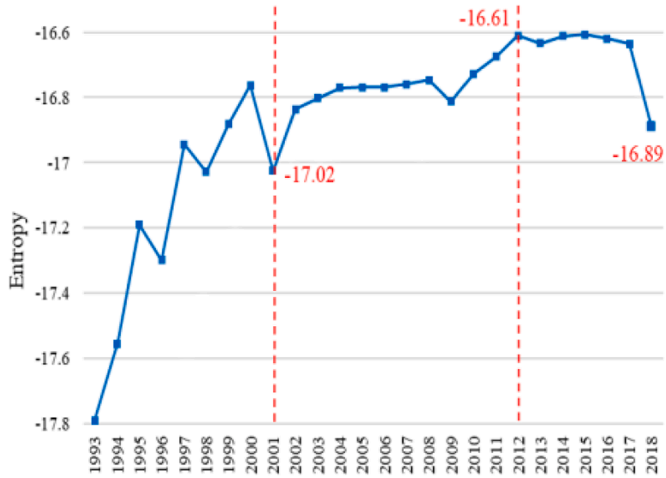

Fig 4. Changing trend of entropy value.

Construction of the topic-term co-occurrence network

First, text data were imported into Clarivate's Derwent Data Analyzer^{TM}, and the multi-word 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 from titles and abstracts using a series of term-clumping processes (Zhang et al., 2014) which is good at cleaning words and terms, and synthesizing technical synonyms, and 131,080 phrases and words were accessed. Afterward, the stop words were removed, the word forms on singular and plural issues were normalized, and manual semantic cleansing was applied to delete irrelevant words and merge synonyms. Owing to the excessive number of topic terms, only term frequencies greater than 15 were selected to obtain a total of 578 topic terms. The BibExcel bibliographic data-analysis tool was used to construct an edge matrix. BibExcel is a toolbox designed by Persson et al. (2009) to assist a user in analyzing bibliographic and any similarly formatted textual data. The number of years in which topic terms occurred and their co-occurrence times are listed in Table 6.

The list data were imported into Gephi software, and the Louvain method for community detection (Blondel et al., 2008) was applied to divide the community following the co-word network graphs from 1940 to 1993 and 1940 to 2017. The Louvain algorithm is one of the most popular for detecting communities in a network (Blondel et al, 2008), and is one of the fastest and best-performing algorithms in comparative analyses (Lancichinetti & Fortunato, 2009; Yang, Algesheimer, & Tessone, 2016). Figures A1 and A2 in

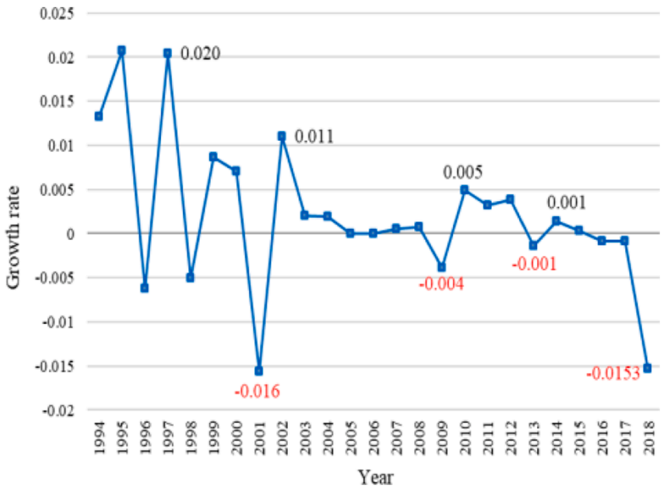

Fig 5. Changes in growth rate of structural entropy value.

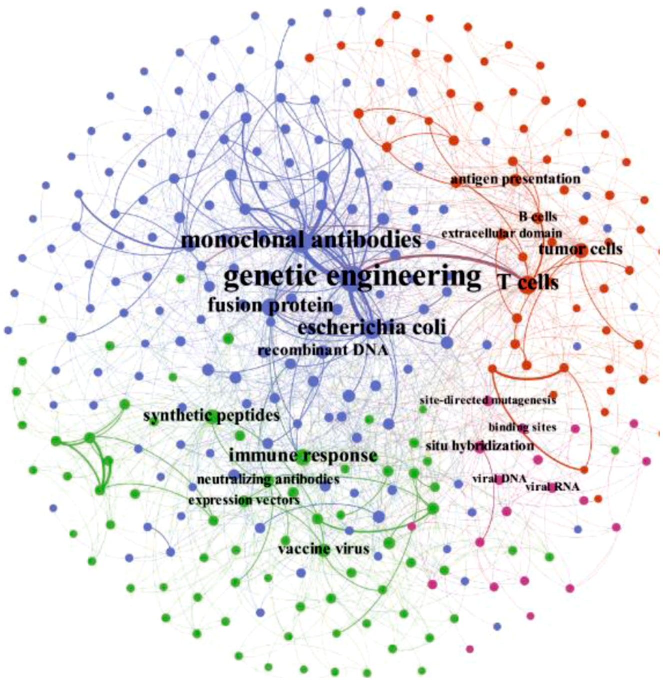

Fig A1. community division from 1940 to 1993.

the Appendix display the result of the community partition. Communities having nodes greater than 1% of the total network nodes (i.e., no fewer than 57 nodes) were reserved. The node sizes were proportional to their centrality and the colors were randomly distributed according to the community divisions. The Fruchterman–Reingold algorithm was used for the layout. In the graph, the larger the node the greater is the node degree, and the greater the edge weight, the thicker the edge.

During the first time slices, from 1940 to 1993, there were few topics and little crossover among them. The connections were mainly between nodes having relatively large degrees (e.g., "genetic engineering," "monoclonal antibodies," "Escherichia coli," and "T cells") (Fig A1). For the remaining time slices, from 1940 to 2018, the number of topics and their crossovers were extensive. The most obvious connections were between "DNA vaccine" with "plasma DNA," "immune response," "dendritic cells," "protective immunity," and "balb/c mice," and between "adaptive immunity" with "T cells," "dendritic cells," "tumor cells," and "CAR T cells" (Fig A2).

This confirms that as knowledge accumulates, the research direction expands. However, an obvious centralization is difficult to perceive. With the integration of knowledge from various research directions, it is also difficult to form an obvious division of communities.

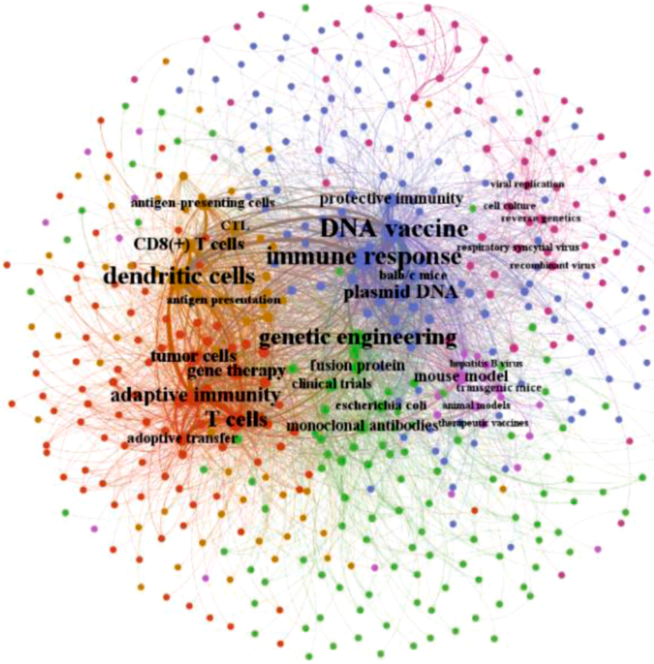

Fig A2. community division from 1940 to 2018.

Table 8
Key time points.

Year	Feature of structural entropy	Representational meaning to the knowledge network
1940-1996	At a descending position after continuous rise	Changes from in disorder to order
1940-1997	At a sharp increase after the decline	Increased disorder
1940-1998	A larger decrease	Changes from in disorder to order
1940-1999	A larger increase	Changes from in order to disorder
1940-2001	At a decreasing position before continuous increase, and the decrease is larger	In the turning point of two disordered states
1940-2002	At a sharp increase after the decline	Changes from in order to disorder
1940-2003	At a sharp decrease point after a sharp increase	Changes from in disorder to order
1940-2009	Is at a decreasing point after continuous growth	Changes from in disorder to order
1940-2010	Is at a rising point after a decrease, and the increase scale is larger	Changes from in order to disorder
1940-2013	In a decreasing state after continuous growth	Changes from in disorder to order
1940-2014	In a state of growth after decline	Changes from in order to disorder
1940-2018	A larger decrease	Changes from in disorder to order

Table 9
Predicted potential breakthrough topics.

Topic No.	Topic content	Breakthrough topic (Yes / No)
#13	Adoptive Cell Transfer Therapy (ACT)	Yes
#16	Reverse genetics system to construct live attenuated vaccine	No
#4	Research on vaccine using CpG as immune adjuvant	No
#9	Immune response to intranasal immunity of Mycobacterium tuberculosis vaccine	No
#14	Therapeutic tumor vaccine	Yes
#0	Phage display monoclonal antibody	Yes
#5	CD8(+) T cell response and cancer treatment	No
#3	Site-specific mutagenesis technology for hepatitis virus research	No
#6	Immune response to therapeutic vaccines	Yes
#2	Stem cell therapy for cancer	Yes

Implementing link prediction

The first step is to divide all knowledge nodes and connections into training and test sets at a ratio of 9:1 for the relatively large-scale dataset. The second step uses all the listed classic algorithms to perform multiple training experiments on the training set and compares the results with the test set to calculate the area under the curve (AUC) curve to evaluate the effect of the algorithm. The third

Table A1
Time slice division and paper number distribution.

Time slices	Number of publications	Time slices	Number of publications
1940-1993	342	1940-2006	2068
1940-1994	432	1940-2007	2224
1940-1995	539	1940-2008	2387
1940-1996	664	1940-2009	2550
1940-1997	775	1940-2010	2712
1940-1998	897	1940-2011	2894
1940-1999	1035	1940-2012	3109
1940-2000	1187	1940-2013	3312
1940-2001	1326	1940-2014	3502
1940-2002	1468	1940-2015	3714
1940-2003	1617	1940-2016	3931
1940-2004	1748	1940-2017	4146
1940-2005	1897	1940-2018	4374

step repeats 100 independent experiments and calculates the average AUC value. The link prediction algorithm having the best performance is selected as the superior index. The pre-screening algorithm is applied to the full dataset in the fourth step.

The AUC is used to evaluate the accuracy of link prediction. The higher the value, the better the link prediction outcome (Zhang & Ma, 2015). Table 7 shows the results of calculating the average AUC of this study.

This study uses the average value of the AUC for judging the effectiveness of the algorithm. The data shows that the WRA algorithm performs best, thereby concurring with the research of Zhang et al. (2021) where link prediction was used for a co-word network. After using the WRA algorithm to predict the potential new associations of the knowledge network, there were 129,937 possible new knowledge connections, the similarity distribution of which is shown in Fig 3. The vertical axis in the figure represents the similarity value (sim) of the newly added edge. According to the link prediction approach, the higher the value, the greater the possibility of linkage between network nodes, in this study, it is the topic-term connection. The horizontal axis represents the serial number (id) of the newly added edge. The serial number is the ranking in descending order of similarity value. We used Matlab to perform power-curve fitting and the similarity scatter of the new knowledge connection predicted in Fig 5 shows a long-tailed distribution approximately satisfying sim=12.97*id^-0.3679 and passing the significance test ($P<0.05$).

The magnitude of the similarity value of the link prediction between the topic terms only represents the possibility of the two future knowledge associations. Thus, even if the similarity value between the topic terms is nonzero, it does not mean that the two will definitely have associations in the future. To reduce the influence of false-noise connections, this study predicted the similarity value distribution characteristics based on the co-occurrence link of the new keywords in the GEV field and used a similarity of 1.3 corresponding to the inflection point of the power-law distribution as the threshold. Finally, 889 strong edges equal to this value were separated and merged with the original keyword co-occurrence network to form a predictive knowledge network.

It remains difficult to predict directly the weight of the new knowledge connection using link-predicting methods. In this study, the similarity value of topic terms was substituted as the weight value to calculate the structural entropy referred to in Zhao's research (Zhao et al., 2015). We conducted a Pearson correlation analysis on the weight and similarity value of the existing knowledge connection. The Pearson correlation coefficient between the two was 0.71, indicating a strong positive correlation between the weight and similarity. This shows that the similarity of topic terms can be used as a substitute variable for weights to calculate structural entropy.

Key time points when structural entropy changed significantly

Fig 4 shows the structural entropy at each time slice. From 1993 to 2018, the overall structural entropy showed an increasing trend for the GEV field. This indicates that the entire knowledge entity and links may have been in "disordered" or "uniform" states (i.e., the non-central features). For the overall trend, the abundance of knowledge obviously surged, and various research directions were involved. The research angle was more diversified, and the centrality characteristic was gradually weakened. Throughout the overall change trend of structural entropy and growth rate, we could screen out the time point at which the structural entropy changed significantly (Fig 5). Table 8 lists the key time points when structural entropy changed significantly.

Screening candidate scientific-breakthrough-topics

After the link prediction analysis, the newly generated network consisted of 563 topic terms and 28,868 knowledge connections. To control the number of divided topics and ensure a good division effect, the Fluid C method (Parés et al., 2017) was used to divide the topical community of complex networks. This algorithm comprehensively considers the number of divided communities and their qualities. Finally, the Fluid C community division algorithm was selected to divide the newly generated co-occurrence network into 19 topics based on the degree of module value and the average number of words contained in each topic. Then, the influence of each topic was measured according to Equation (2), and the top-10 topics having the highest values of "topic impact" were selected as the main analysis objects. Expert judgments were then obtained for comparison. Table 9 shows the expert evaluation results. Five topics were evaluated by experts as potential breakthroughs with great future development prospects. In order of breakthrough-potential degree, they are #13, #14, #0, #6, and #2.

Empirical validation

For these five topics (#13, #14, #0, #6, and #2), we combined authoritative journals, scientific awards, and expert consultations to validate the identification results. As mentioned, expertise and search results were used for the evaluation. Two experts were consulted; a scientific researcher engaged in long-term vaccine research and a science and technology intelligence analyst with experience in drug discovery and developments with chemical, biological, and medicinal intelligence. The experts evaluated each topic on its breakthrough potential using the following four indicators:

Does this constitute the discovery of a new phenomenon or mechanism of action?

Will new findings and breakthroughs continue to appear in this field in the next 3–5 years?

Are the results likely to be applied in the future and do they have great prospects?

Is this topic receiving continuous attention from scientific research/funding institutions/R&D companies?

Because the results returned by the experts during these consultations were inconsistent, we reviewed relevant materials and conducted another round of consultation. Ultimately, three rounds were needed. Specifically based on expert opinions, we further elaborated five topics that may represent further scientific breakthroughs. Furthermore, we also compared the results of the empirical analysis with qualitative evidence in the field of GEV, including authoritative reviews of prospective articles, annual scientific breakthroughs, etc. The considerations that served as evidence to identify the five topics as scientific breakthroughs are discussed below.

Topic #13: Adoptive Cell Transfer Therapy (ACT).

ACT is applied to induce, modify, and amplify the body's immune cells in vitro and to re-transport to the patient the selected effector cells with specific and efficient tumor-killing capacity (Rosenberg & Restifo, 2015). This is a new anti-tumor method with good clinical application prospects with curative effects. ACT includes lymphokine-activated killer cells, chimeric antigen receptor (CAR) modified T cells, CAR-natural-killer cells, T-cell-receptor modified T-cells, and cytokine-activated killer cells (CIK), tumor-infiltrating lymphocytes, induced dendritic cells (DC), DC-CIK, etc. (Wang et al., 2018). ACT has high specificity and low physiological side effects. It is now gradually shifting from basic experiments to clinical applications and has achieved significant results in the treatment of various malignant tumors and hematological diseases. Therefore, it is expected that there will be further progress in the future.

Topic #14: Therapeutic tumor vaccine

Therapeutic tumor vaccines provide active tumor immunotherapy, which uses the patient's immune system to induce continuous anti-tumor immunity to treat and prevent tumor recurrence or metastasis. Current therapeutic tumor vaccine studies generally refer to specific tumor vaccines that specifically attack and destroy tumor cells without damaging normal cells while inhibiting advanced cancer or recurrent tumors from progressing. This therapy has made some progress since it was proposed, and a variety of tumor vaccines have been produced, including cell, nucleic acid, peptide, and dendritic types (Li & Zheng, 2019). However, the variability of cancer and the acquisition of tumor antigens still restrict the therapy's further development. According to official US statistics, since 2017, 105 therapeutic tumor vaccines have undergone clinical trials (Du, 2018). Therefore, it is expected this method will have plentiful research investments that may help overcome the above issues in the future.

Topic #0: Phage display monoclonal antibody.

This technology was established in 1985 and was first used for antibody preparation in 1990. In 2002, the first monoclonal antibody developed with phage display technology was certified by the US Food and Drug Administration (Hou, Hu, & Cui, 2018) and was assessed as a Nobel Prize-level technological innovation in 2018 (Qiang, Zhang, Wei, & Hou, 2010). When using this technology to construct an antibody library, cell fusion steps are omitted and the tedious procedure of repeated subcloning is avoided. This technology can directly obtain antibody genes, facilitate the further construction of various genetically engineered antibodies, and assist some difficult production antibodies (e.g., weak immunogens, toxic antigens, and humanized antibodies). This technology has a short cycle and adds value in bacteria. Therefore, it is suitable for the large-scale industrial production of antibodies. Presently, there are many single-chain antibody therapeutic applications (Xia, 2018), and further breakthroughs are expected.

Topic #6: Immune response to therapeutic vaccines.

Therapeutic vaccines refer to products that are natural, artificially synthesized, or expressed by genetic recombination technology in a body infected with pathogenic microorganisms or suffering from certain diseases. The vaccine induces specific immune responses to treat or prevent disease-related deterioration. Therapeutic vaccines belong to the field of specific active immunotherapy, which provides a new way to treat severe and chronic diseases caused by viral infections. In particular, some studies have insisted that we pay attention to neo-antigen tumor therapeutic vaccines in the future (Zhang, Luo, Yan, & Wu, 2020).

Topic #2: Stem-cell therapy for cancer.

After in-depth research, scientists found that stem-cell therapy is expected to be used to treat a variety of human diseases, including those that are incurable (e.g., Alzheimer's disease, diabetes, and cancer) (Boya, 2018). For the future development of stem-cell research, using induced pluripotent stem cells to prepare vaccines specifically to identify and lock cancer cells is expected to introduce new methods of cancer treatment.

Discussion

Our research gives insight into the dynamics of science because the structural entropy method regards networks as complex systems from a holistic and dynamic perspective and considers network growth through the lens of dynamic evolutionary processes. Moreover, structural entropy can measure the changes in both the strong and weak ties between topics in the domain knowledge network. Thus, it can function as an early signal of domain topic changes and help to identify a scientific breakthrough during its early stages. The structural entropy method considers the evolutionary processes of gestation, propagation, development, and mutation—it is a dynamic perspective. Therefore, the change in structural entropy can be attributed to the evolution of the network and can be used to recognize emerging topics that significantly impact the relevant knowledge-network structure. These can then be highlighted as potential breakthrough topics.

Furthermore, the structural changes in complex networks depict the non-linear characteristics of the complex knowledge networks and identify the key time-points of the transition states of these knowledge systems. The point at which the structural entropy of the whole network changes significantly can be considered as the key moment of a scientific breakthrough. Furthermore, the time slices of major changes in the domain knowledge network can be analyzed to identify the time slices before and after the formation of potential scientific breakthroughs. This facilitates the prediction of potential future scientific breakthroughs by aiding further investigation into the scientific advancements around this time slice.

The results of this study also contribute to the identification the future research topics since the terms that identify such topics are often already present in the knowledge network in the form of weak linkages. Thus, breakthrough topics can be identified by applying semantic enhancement. This is where link prediction becomes important because it can reinforce the semantic association of topic terms in networks and enhance accuracy when predicting future topics.

Our proposed method can lead to the identification of a scientific breakthrough during its early stages and even help to establish forward-looking predictions. Therefore, this method can certainly play a key role in identifying and predicting scientific breakthroughs. Such early-stage identification is relevant as it provides decision-makers with sufficient time to react. Allowing research organizations, governments, and enterprises to adjust their scientific strategies rapidly, allocate reasonable research-and-development resources, and seize developmental opportunities.

Conclusion

A structural entropy measurement index for knowledge-network assessment that is based on network structures and their nonadditive characteristics, was constructed from the perspective of changing knowledge networks. This research attempts to combine link prediction and structural entropy methods to predict scientific breakthrough topics. Through the temporal-value change of structural entropy in a network, changes can be monitored to identify potential breakthrough topics. An empirical analysis of GEV studies is provided and the methodological results are compared with published scientific breakthroughs and expert evaluations. Ultimately, the proposed model's recognition capability was verified, receiving good expert support. We found that monitoring the dynamics of structural entropy over time can reveal significant changes in the knowledge network, where the community structure of the network changes continuously, and some of the identified breakthroughs were future-pending or imminent showing that the method can indeed identify potentially major breakthroughs.

However, this proposed identification method does not guarantee the accuracy of our research method. A scientific breakthrough may still take a long time to manifest and develop. The change in a knowledge structure is only one representation of this phenomenon, and many other impact factors ought to be considered in future works.

Limitations of this study

First, although network-change measurement can begin with the analysis of its structure, the structural change will not be equal to the change of topic content. Additionally, the scientific-knowledge network includes many topics and connections, and the structure is highly complex. Structural entropy can only measure changes in the network linkages through the knowledge links. As such, this research does not reveal the reasons for the changes.

Second, the topic network in this study was divided into annual time slices. However, according to the known relationship between structural entropy and knowledge networks, the measurement of structural entropy changes requires a more finely grained time segmentation. The finer the granularity, the richer the information to be monitored. In summary, to monitor changes in structural entropy more accurately, the construction of the knowledge network should be based upon the emergence of new knowledge or new connections.

Third, because the composition of the actual knowledge network involves more elements, the topic-terms co-occurrence network is a proxy and not equivalent to the knowledge network. This article only used the GEV field as a testbed. The feasibility of the verification method and its generalizability is yet to be verified. Moreover, the data sources of this article were mainly based on scientific papers and lacked analyses of science- and technology-related topics.

Future work

Based on the findings of this study we foresee three future research activities: 1) a revised assessment, 2) focusing on more effectively capturing the early weak signals of scientific breakthroughs, and 3) paying attention to the relevant dynamic evolutionary processes.

Revised assessment

In the future, we plan to undertake revisiting assessment research by the end of 2023, similar to Huang et al. (2019). This date is five years after 2018 by when there will be a complete and independent data range, covering five consecutive years from 2019 to 2023. Therefore, the accuracy of multidimensional indicators can be tested to further reveal the influencing factors that led to differences between past predictions and actual performance. Further improvements will accordingly be made to our multidimensional indicators.

Focusing on capturing the early weak signals of scientific breakthroughs more effectively.

This study based the calculation of structural entropy of the knowledge network on high-frequency topic terms. However, weak ties with low frequency were not included. In their early stages, breakthrough topics often appear as weak ties among other topics. These weak ties differ greatly from strong ties but can accurately reflect the nature of an emerging subject. Ignoring this type of data could result in the omission of important information. Unlike strong signals (e.g., high-frequency terms or highly cited documents), the information contained in weak signals is more diverse and can better reflect the knowledge structure of the overall discipline. There are many scientific discoveries and scientific breakthroughs in real-time and attending to the early signs of change would be a monumental task. However, it is worth investigating this type of improvement for the future. The limitation of having fewer signals to measure during the early stages of emerging topics, could be overcome using the Kalman filter (Sheen & Wang, 2017) technique since more available data improves the results' reliability.

Considering the dynamics of evolving scientific breakthroughs.

The dynamic mechanism of technological development is extremely complicated. Coccia reveals that the evolution of science is due to expanding human life-interests the increasing realization of which constitutes the progress that has characterized civilization for millennia (Coccia and Bellitto, 2018), due to manifold historical factors (Coccia, 2020b), such as the social contexts of nations, new technologies, new discoveries, economic growth, democratization, military and political tensions between competing powers to advance scientific and technological superiority, new challenges between superpowers for sustaining global leadership and other social and scientific events and processes.

The characterizations and methods of quantifying scientific breakthroughs, here referring to the dynamic evolution of scientific breakthroughs still require further exploration. This evolution proceeds incrementally through quantitative and qualitative changes. Most extant recognition methods only focus on static signals at certain points in time but cannot effectively observe the knowledge diffusion of an entire topic area. As a complex system, the development process of the knowledge network is dynamic and changeable. It will be necessary to interpret the process of knowledge development from a dynamic perspective to adequately analyze the mechanisms of scientific breakthroughs. Therefore, future breakthrough-identifying research should consider the introduction of time-series analyses and growth models (e.g., preferential attachment) to analyze the laws governing the generation, development, and demise of topics.

Acknowledgments

We appreciate the constructive advice and suggestions from Professor Shuo Xu at the College of Economics and Management, Beijing University of Technology. This article is an outcome of the Taishan Scholars Youth Expert Program of Shandong Province (202103069). The "Study on the Recognition Method of Innovative Evolving Trajectory based on Topic Correlation Analysis of Science and Technology" (No.71704170) is supported by the National Natural Science Foundation of China, China Postdoctoral Science Foundation-funded project (2016M590124).

References

Adams, J. (2012). The rise of research networks. *Nature, 490*(7420), 335–356.
Adams, J. (2013). The fourth age of research. *Nature, 497*(7451), 557–560.
Almog, A., & Shmueli, E. (2019). Structural entropy: monitoring correlation-based networks over time with application to financial markets. *Scientific reports, 9*(1), 1–13.
Andersen, H., Barker, P., & Chen, X. (2006). *The cognitive structure of scientific revolutions*. Cambridge: Cambridge University Press.
Apreda, R., Bonaccorsi, A., dell'Orletta, F., & Fantoni, G. (2019). Expert forecast and realized outcomes in technology foresight. *Technological Forecasting and Social Change, 141*, 277–288.
Arthur, W. B. (2007). The structure of invention. *Research Policy, 36*(2), 274–287.
Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. New York: Simon and Schuster.
Bettencourt, L. M., Kaiser, D. I., & Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks. *Journal of Informetrics, 3*(3), 210–221.
Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment, 2008*(10), P10008.
Börner, K., & Scharnhorst, A. (2009). Visual conceptualizations and models of science. *Journal of Informetrics, 3*(3), 161–172.
Bolger, F., & Wright, G. (2017). Use of expert knowledge to anticipate the future: Issues, analysis and directions. *International Journal of Forecasting, 33*(1), 230–243.
Börner, K., Boyack, K. W., Milojević, S., & Morris, S. (2012). An introduction to modeling science: basic model types, key definitions, and a general framework for the comparison of process models. In A. Scharnhorst, K. Börner, & P. van den Besselaar (Eds.), *Models of science dynamics* (pp. 3–22). New York: Springer.
Börner, K., Glänzel, W., Scharnhorst, A., & van den Besselaar, P. V. (2011). Modeling science: Studying the structure and dynamics of science. *Scientometrics, 89*(1), 347–348.
Boya, S. C. B. (2018). Use stem cell therapy to treat cancer, Alzheimer's disease and other diseases. Retrieved from http://k.sina.com.cn/article_3053403605_b5ff3dd5027004wqd.html.
Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351–374.
Brad Wray, K. (2011). Kuhn and the Discovery of Paradigms. *Philosophy of the Social Sciences, 41*(3), 380–397.
Cai, M., Du, H., Ren, Y., & W. Feldman, M. (2011). A new network structure entropy based node difference and edge difference. *Acta Physica Sinica, 60*(11), 165–173.

Cai, Y., Cao, Y., Tan, Y., & Li, Y. (2015). Influences of Power Grid Structure on Cascading Failure Based on Standard Structure Entropy. *Transactions of China Electrotechnical Society, 30*(3), 36–43.
Cao, K., & Wang, C. (2005). Tsallis entropy and nonex tensive statistical mechanics. *Journal of Yunnan University*, (6), 60–66.
Chen, C. (2012). *Turning points: The nature of creativity*. Berlin: Springer Science & Business Media.
Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics, 3*(3), 191–209.
Chen, C., Chen, Y., Wang, X., & Hu, Z. (2015). *Turning Points: the nature of creativity*. Beijing: Science Press.
Chinese-Academy-of-Sciences. (2020). *2019 Scientific Development Report*. Beijing: Science Press.
Clausius, R. (1865). Presentation to the Philosophical Society of Zurich. Retrieved from http://web.mit.edu/keenansymposium/.
Coccia, M. (2018). General properties of the evolution of research fields: a scientometric study of human microbiome, evolutionary robotics and astrobiology. *Scientometrics, 117*(2), 1265–1283.
Coccia, M. (2020a). Theories and laws of scientific development. *Working Paper CocciaLab n. 50/2020*. CNR–National Research Council of Italy, 62489–62018, Available at SSRN: https://ssrn.com/abstract=3568913.
Coccia, M., & Finardi, U. (2012). Emerging nanotechnological research for future pathway of biomedicine. *International Journal of Biomedical nanoscience and nanotechnology, 2*(3-4), 299–317.
Coccia, M., & Wang, L. (2015). Path-breaking directions of nanotechnology-based chemotherapy and molecular cancer therapy. *Technological Forecasting & Social Change, 94*, 155–169.
Coccia, M. (2020b). The evolution of scientific disciplines in applied sciences: dynamics and empirical properties of experimental physics. *Scientometrics, 124*(1), 451–487.
Coccia, M., & Bellitto, M. (2018). Human progress and its socioeconomic effects in society. *Journal of Economic and Social Thought, 5*(2), 160–178.
Costa, L. D. F., Rodrigues, F. A., & Cristino, A. S. (2008). Complex networks: the key to systems biology. *Genetics and Molecular Biology, 31*(3), 591–601.
Costa, L.d. F., Rodrigues, F. A., Travieso, G., & Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in physics, 56*(1), 167–242.
Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy, 34*(5), 717–737.
DeepTech. (2020). MIT Technology Review"2020 "Top Ten Global Breakthrough Technologies". *Chinese Technology Business*, (3), 5–11.
Dong, H., Tang, T., & Wang, H. (2019). A 2D Structure Entropy-based Approach to Security Assessment of Communication-based Train Control System. *Acta Automatica Sinica, 45*(1), 153–162.
Du, H. (2018). Progress in International Research and Development of Therapeutic Cancer Vaccine. *Progress in Pharmaceutical Sciences, 42*(9), 685–696.
Du, J. (2017). *A study on systematic identification of "Sleeping Beauty" publications and on their awaking mechanism*. Nanjing, China: Nanjing University.
Eulaerts, O., Joanny, G., Giraldi, J., Fragkiskos, S., & Perani, S. (2019). *Weak signals in Science and Technologies - 2019 Report, EUR 29900 EN, Publications Office of the European Union*. Luxembourg: ISBN, 978-92-76-12387-3, doi:10.2760/50544, JRC118147.
Fanelli, D., & Glänzel, W. (2013). Bibliometric evidence for a hierarchy of the sciences. *PLoS ONE, 8*(6), e66938.
Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., & Barabási, A. L. (2018). *Science of science. Science, 359*(6379).
Fu, Y., & Zhang, H. (2004). Breakthrough innovation: concept definition and comparison. *The Journal of Quantitative & Technical Economics, 21*(3), 73–83.
Fukawa, I. (2006). Case studies on how to enhance the chance of technical breakthrough and (pseudo) serendipity. In *, 2. 2006 Technology Management for the Global Future-PICMET 2006 Conference* (pp. 668–675). IEEE.
Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817.
Gibbons, M., Limoges, C., Nowotny, H., Schwatzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary society*. London: Sage.
Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological theory* (pp. 201–233).
Guan, X. (2014). *Study on method of dynamic network structure entropy measurement based on BA model*. Shenyang: Northeastern University.
Hou, X., Hu, Z., & Cui, B. (2018). Recent advances of phage display techniques for drug discovery. *Acta Pharmaceutica Sinica, 53*(8), 1279–1288.
Huang, L., Zhu, Y., & Zhang, Y. (2019). Research on Identification of Emerging Topics Based on Link Prediction with Weighted Networks. *Journal of the China Society for Scientific and Technical Information, 38*(4), 335–341.
Huo, Y. (2019). *Research on Node Importance and Network Heterogeneity of Brain Network Based on K-order structure entropy*. Nanjing: Nanjing University of Posts and Telecommunications.
Jia, W., Xie, Y., Zhao, Y., Yao, K., Shi, H., & Chong, D. (2021). Research on Disruptive Technology Recognition of China's Electronic Information and Communication Industry Based on Patent Influence. *Journal of Global Information Management (JGIM), 29*(2), 148–165.
Jiang, J., Li, X., Yin, T., & Qu, X. (2017). Formation of Breakthrough Technology Innovation: Review and Outlook. *R & D Management, 29*(6), 109–120.
Joanny, G., Perani, S., & Eulaerts, O. (2019). Detection of disruptive technologies by automated identification of weak signals in technology development. proceedings of the ISSI. *International Society for Scientometrics and Informetrics* (pp. 2644–2645).
Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery, 7*(4), 373–397.
Koshland, Daniel E (2007). The cha-cha-cha theory of scientific discovery. *Science, 317*(5839), 761–762.
Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical review E, 80*(5), Article 056117.
Leydesdorff, L., & Cozzens, S. E. (1993). The delineation of specialties in terms of journals using the dynamic journal set of the Science Citation Index. *Scientometrics, 26*(1), 135–156.
Li, X, & Xu, Y. (2007). Research on Information Science Application of Complex Network Theory. *Information and Documentation Services*, (3), 9–13.
Li, F., & Zheng, S. (2019). The Current State and Prospect of Tumor Immunotherapy. *Journal of Kunming University of Science and Technology (Natural Science), 44*(5), 76–78.
Li, G. (2020). *Research on Knowledge Transfer between Enterprises Based on Link Prediction*. Jinan: Shandong Normal University.
Li, Z., Liu, C., & Luo, H. (2016). Analysis of Essence and Nurturing of Disruptive Technology: Highlighting Basic Science Researches behind Disruptive Technologies. *Global Science Technology and Economy Outlook, 31*(10), 53–61.
Liang, Z., Deng, X., & Hong, Y. (2017). From Transformative Research to Transformative Innovation: Evolution of Concept and Implication for Policy. *Science and Society, 7*(3), 94–106.
Liang, Z., Mao, J., Lu, K., Ba, Z., & Li, G. (2021). Combining deep neural network and bibliometric indicator for emerging research topic prediction. *Information Processing & Management, 58*(5), Article 102611.
Liu, Y., & X, H. (2021). A Review of Early Recognition of Breakthrough Innovations and the Weak Signal Analysis. *Library and Information Service, 65*(4), 89.
Lü, L. (2010). Link Prediction on Complex Networks. *Journal of University of Electronic Science and Technology of China, 39*(5), 651–661.
Lü, L., & Zhou, T. (2010). Link prediction in weighted networks: The role of weak ties. *EPL (Europhysics Letters), 89*(1), 18001.
Luo, P., Li, Y., & Wu, C. (2013). Complex Networks Evolution Research Using the Network Structure Entropy. *Complex Systems and Complexity Science, 10*(4), 62–68.
Luo, R. (2020). *Identification and Prediction of the Field's Scientific Breakthrough based on Entropy theory*. Beijing: University of Chinese Academy of Sciences.
Luo, R., Xu, H., & Liu, Y. (2021). Identifying scientific breakthrough topics using structure entropy: Taking the study field of genetic engineering vaccines as an experimentation. *Information Studies: Theory & Application, 44*(05), 106–114.
Luo, S., Kou, C., Jin, J., & Yuan, H. (2019). Disruptive Technology Prediction Based on Outlier Patents: Traditional Chinese Medicine Patents as an Example. *Information studies: Theory & Application, 42*(7), 165–170.
Ma, X., Wang, Y., & Ye, Y. (2014). Research on the Relationship between Chinese Energy Consumption Structure and Industrial Structure Based on Entropy Theory. *Energy of China, 36*(7), 31–35.

Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., & Marrs, A. (2013). *Disruptive technologies: Advances that will transform life, business, and the global economy, 180* pp. 17–21). San Francisco, CA: McKinsey Global Institute.
Mao, J., Ma, C., & Liang, Z. (2019). Identifying Emerging Technology: A Neural Network Based Solution. Retrieved from https://vpinstitute.org/wp-content/uploads/2019/10/MTEGTM2019-Jin.pdf.
Meng, B., Ke, H., & Yi, T. (2011). Link prediction based on a semi-local similarity index. *Chinese Physics B, 20*(12), Article 128902.
Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.
Merton, R. K., & Barber, E. G. (2004). *The travels and adventures of serendipity: A study in historical semantics and the sociology of science*. Princeton, NJ, USA: Princeton University Press.
Min, C., Bu, Y., & Sun, J. (2019). Comparing Breakthrough and Non-Breakthrough Papers from Early Citing Structures. In *ISSI 2019 conference proceedings* (pp. 1110–1115).
Min, C., Bu, Y., & Sun, J. (2021). Predicting scientific breakthroughs based on knowledge structure variations. *Technological Forecasting and Social Change, 164*, Article 120502.
Min, C., Bu, Y., Wu, D., Ding, Y., & Zhang, Y. (2021). Identifying citation patterns of scientific breakthroughs: a perspective of dynamic citation process. *Information Processing & Management, 58*(1), Article 102428.
Palomeras, N. (2003). Sleeping patents: any reason to wake up. *IESE Research Papers No D/506*.
Parés, F., Gasulla, D. G., Vilalta, A., Moreno, J., Ayguadé, E., Labarta, J., & Suzumura, T. (2017). Fluid communities: A competitive, scalable and diverse community detection algorithm. In *International Conference on Complex Networks and their Applications* (pp. 229–240). Cham: Springer.
Persson, O., Danell, R., & Schneider, J. W. (2009). How to use Bibexcel for various types of bibliometric analysis. *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday, 5*, 9–24.
Petzold, N., Landinez, L., & Baaken, T. (2019). Disruptive innovation from a process view: A systematic literature review. *Creativity and Innovation Management, 28*(2), 157–174.
Ping, E. (2014). *Research on negentropy driven development of key technologies for radical innovation*. Tianjin: Hebei University of Technology.
Popkov, Y. S. (2019). Soft Randomized Machine Learning Procedure for Modeling Dynamic Interaction of Regional Systems. *Entropy, 21*(4), 424.
Qiang, R., Zhang, N., Wei, L., & Hou, Y. (2010). Phage Display Technology and Its App lications for Tum or Research. *China Biotechnology, 29*(9), 81–86.
Rosenberg, S. A., & Restifo, N. P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. *Science, 348*(6230), 62–68.
Roshani, S., Bagheri, R., Mosleh, M., & Coccia, M. (2021). What is the relationship between research funding and citation-based performance? A comparative analysis between critical research fields. *Scientometrics, 126*(9), 7859–7874.
Savov, P., Jatowt, A., & Nielek, R. (2020). Identifying breakthrough scientific papers. *Information Processing & Management, 57*(2), Article 102168.
Schneider, J. W., & Costas, R. (2017). Identifying potential "breakthrough" publications using refined citation analyses: Three related explorative approaches. *Journal of the Association for Information Science and Technology, 68*(3), 709–723.
Seymour, S. B. (2009). Serendipity. *The North Carolina Law Review, 88*, 185–211.
Sheen, J., & Wang, B. Z. (2017). Estimating Macroeconomic Uncertainty from Surveys-A Mixed Frequency Approach. *Available at SSRN 3020697*.
Small, H. G. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. *Social Studies of Science, 7*(2), 139–166.
Small, H., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy, 43*(8), 1450–1467.
Solé, R. V., & Valverde, S. (2004). Information theory of complex networks: on evolution and architectural constraints. *Complex networks* (pp. 189–207). Berlin, Heidelberg: Springer.
Staudt, J., Yu, H., Light, R. P., Marschke, G., Börner, K., & Weinberg, B. A. (2018). High-impact and transformative science (HITS) metrics: Definition, exemplification, and comparison. *PloS one, 13*(7), Article e0200597.
Tan, Y., & Wu, J. (2004). Network Structure Entropy and Its Application to Scale-free Networks. *System Engineering Theory and Practice, 24*(6), 1–3.
ThomsonReuters. (2015). *Open future: 2015 global innovation report*. New York: Thomson Reuters.
Tria, F., Loreto, V., Servedio, V. D. P., & Strogatz, S. H. (2014). The dynamics of correlated novelties. *Sci. Rep, 4*(5890), 1–8.
Valverde, S., Sole, R. V., Bedau, M. A., & Packard, N. (2007). Topology and evolution of technology innovation networks. *Phys. Rev. E, 76*(5), Article 056118.
van Andel, P. (1994). Anatomy of the unsought finding. serendipity: Origin, history, domains, traditions, appearances, patterns and programmability. *The British Journal for the Philosophy of Science, 45*(2), 631–648.
Van Oirschot, J., Moormann, R., Berns, A., & Gielkens, A. (1991). Efficacy of a pseudorabies virus vaccine based on deletion mutant strain 783 that does not express thymidine kinase and glycoprotein I. *American journal of veterinary research, 52*(7), 1056.
van Raan. (2017). Sleeping beauties cited in patents: Is there also a dormitory of inventions? *Scientometrics, 110*(3), 1123–1156.
van Raan, A. F., & Winnink, J. J. (2018). Do younger Sleeping Beauties prefer a technological prince? *Scientometrics, 114*(2), 701–717.
Vanhoucke, M., & Batselier, J. (2019). A Statistical Method for Estimating Activity Uncertainty Parameters to Improve Project Forecasting. *Entropy, 21*(10), 952.
Wan, H. (2017). *Research on the Growth Mode of Scientific Knowledge Scale——Based on Mathematical Modeling and Simulation*. Beijing: University of Chinese Academy of Sciences.
Wang, M., & Zhao, B. (2014). Structural Features and Evolution Dynamics of Science Knowledge Network. *Journal of intelligence*, (5), 88–95.
Wang, C., Wu, X., & Luo, X. (2015). Method for invulnerability evaluation of WSNs based on betweenness centrality. *Transducer and Microsystem Technologies, 34*(5), 19–21.
Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy, 46*(8), 1416–1436.
Wang, X., Yang, X., Du, J., Wang, X., Li, J., & Tang, X. (2021). A deep learning approach for identifying biomedical breakthrough discoveries using context analysis. *Scientometrics, 126*(7), 5531–5549.
Wang, Z., Zhang, J., Fu, Z., Meng, Q., Xuan, G., Feng, R., & Yang, K. (2018). The research progress and prospect of adoptive cell therapy of tumors. *Chemistry of Life, 38*(1), 97–103.
Wei, L., Xu, H., Wang, Z., Dong, K., Wang, C., & Fang, S. (2016). Topic detection based on weak tie analysis: a case study of LIS research. *Journal of Data and Information Science, 1*(4), 81–101.
Winnink, J. (2017). *Early-stage detection of breakthrough-class scientific research: using micro-level citation dynamics*. Leiden: Leiden University (Ph.D Thesis).
Winnink, J. J., Tijssen, R. J., & van Raan, A. F. (2016). Theory-changing breakthroughs in science: The impact of research teamwork on scientific discoveries. *Journal of the Association for Information Science and Technology, 67*(5), 1210–1223.
Winnink, J., & Tijssen, R. J. (2014). R&D dynamics and scientific breakthroughs in HIV/AIDS drugs development: the case of integrase inhibitors. *Scientometrics, 101*(1), 1–16.
Winnink, J., Tijssen, R. J., & van Raan, A. (2019). Searching for new breakthroughs in science: How effective are computerised detection algorithms? *Technological Forecasting and Social Change, 146*, 673–686.
Wolcott, H. N., Fouch, M. J., Hsu, E. R., DiJoseph, L. G., Bernaciak, C. A., Corrigan, J. G., & Williams, D. E. (2016). Modeling time-dependent and-independent indicators to facilitate identification of breakthrough research papers. *Scientometrics, 107*(2), 807–817.
Wuestman, M., Hoekman, J., & Frenken, K. (2020). A typology of scientific breakthroughs. *Quantitative Science Studies, 1*(3), 1203–1222.
Xia, J. (2018). Research progress of phage display technology in the field of monoclonal human single-chain antibodies. *Biology Teaching, 43*(4), 69–70.
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., & Luo, R. (2019). A Study on the Multidimensional Scientometric Indicators to Detect the Emerging Topics. In *the 17th International Conference on Scientometrics and Informetrics*.
Xu, L., & Wang, F. (2019). Scientific Frontier Prediction Model Based on Support Vector Machine and Improved Particle Swarm Optimization. *Information Sciences, 37*(8), 22–28.

Yang, X., Wang, W., Du, Jian., & Tang, X. (2020). Identifying Feature Words Based on Abstracts and Citation Text Corpus of Breakthrough Research. *Library and Information Service, 64*(11), 125.

Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific reports, 6*(30750), 1–18.

Yoo, S. H., & Won, D. (2018). Simulation of weak signals of nanotechnology innovation in complex system. *Sustainability, 10*(2), 486.

Yoon, J., & Kim, K. (2011). Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics, 88*(1), 213–228.

Yue, Z., Xu, H., & Wang, Q. (2020). Dynamic Link Prediction of Knowledge Diffusion in Disciplinary Citation Networks Based on Local Information. *Information studies: Theory & Application, 43*(2), 84.

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.

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, J., Luo, X., Yan, P., & Wu, Y. (2020). Progress of neoantigen tumor therapeutic vaccine. *Immunological Journal, 36*(3), 264–270.

Zhang, Q. (2017). *Structure analysis of complex networks based on entropy*. Chongqing: Southwest University.

Zhang, Y. (2011). Research on the scientific front detection by low-frequency occurrence phenomenon. *Chinese Academy of Sciences*.

Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H., & Zhang, G. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics, 12*(4), 1099–1117.

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.

Zhao, J., Miao, L., Yang, J., Fang, H., Zhang, Q.-M., Nie, M., & Zhou, T. (2015). Prediction of links and weights in networks by reliable routes. *Scientific reports, 5*, 12261.

Zhou, T. (2015). Link prediction and weight prediction of weighted network. Retrieved from http://blog.sciencenet.cn/blog-3075-909760.html.