



# Predicting potential knowledge convergence of solar energy: bibliometric analysis based on link prediction model

Yueran Duan<sup>1</sup> · Qing Guan<sup>2</sup>

Received: 13 December 2019 / Accepted: 3 February 2021 / Published online: 8 March 2021  
© Akadémiai Kiadó, Budapest, Hungary 2021

## Abstract

The innovation and development of emerging technology mostly depend on the way of knowledge convergence defined as the blurring of previously distinct domain-specific knowledge. This paper aims to explore the potential motivation of knowledge convergence and find the law of knowledge convergence, taking the solar energy field as an example. We established Keywords co-occurrence networks of solar energy literature in 2008–2017, and then link prediction is introduced to study the structural mechanism of knowledge convergence. We found that: (1) the common neighbor index better characterizes the knowledge convergence pattern in the knowledge networks among four similarity indicators. (2) The keywords co-occurrence network could effectively mine the structural characteristics of knowledge convergence; (3) the convergence cycle of knowledge in the field of solar energy was about 4 years; (4) keywords with higher betweenness centrality or eigenvector centrality easily generated knowledge convergence; (5) a literature knowledge convergence prediction model is proposed based on these results; and (6) the prediction results showed that scholars should pay attention to six basic issues including energy storage, efficiency, cost, ecological effect, application scenarios, and hybrid photovoltaic systems. This work can provide guidance not only for scholars to grasp the research direction and to generate more innovations but for the government to formulate the policies of government funding.

**Keywords** Solar energy · Knowledge convergence · Complex network · Link prediction · Text mining · Bibliometric analysis

## Introduction

Convergence can be defined as at least two distinct entities moving toward union (Chan and Miyazaki 2015). Practically, the generation of creativity to actual industrial production needs to go through the process of knowledge convergence, technology convergence, and finally industrial convergence (Zhou et al. 2019). The research of convergence is helpful to reveal the innovation characteristics and development trends of emerging fields (Jeong et al.

---

✉ Qing Guan  
guanqing35@126.com

<sup>1</sup> School of Economics and Management, China University of Geosciences, Beijing 100083, China

<sup>2</sup> School of Information Engineering, China University of Geosciences, Beijing 100083, China

2018). In recent years, technological convergence which brought industrial transformation (Sick et al. 2019; Boussemart et al. 2020) and economic and financial problems (Lim et al. 2018; Stucki and Woerter 2019) has become the focus of scholars' research. However, as the basis of technological convergence, the generation and source of ideas should also be concerned and studied. In the academic field, such convergence is manifested as knowledge convergence which can be defined as the blurring of previously distinct domain-specific knowledge. Specifically, scholars apply different fields and different types of knowledge entities to the same problem to complete a research paper (Chan and Miyazaki 2015). The blurring of boundaries between different knowledge entities has resulted in many meaningful innovations, especially in emerging technology fields. A quintessential case is the fields of solar energy. In early research on solar energy related technologies, most scholars studied battery materials (Dennler et al. 2008; Morfa et al. 2008), cost (Fthenakis et al. 2009), energy storage (N'Tsoukpoe et al. 2009; Nocera 2009), and other single problems. Later, scholars gradually considered several issues for research. For example, research that combined material research and development with cost (Wong and Ho 2010); research on the combination of photochemistry and solar energy harvesting (Schulze and Schmidt 2015); research on the economic analysis of solar technology (Armaroli and Balzani 2016; Shaner et al. 2016); and research on solar energy application scenarios, such as desalination (Sharon and Reddy 2015) and building integrated photovoltaic (Chae et al. 2014). The research on the field of solar energy is gradually detailed and the number of topics is increasing. How to converge different knowledge entities to innovate and promote interdisciplinary communication and development has become a focus that scholars pay close attention to. Therefore, this paper attempts to take the solar energy field as an example to find out the basic characteristics of knowledge convergence rules.

In the study of the innovation process, some scholars have adopted the method of case study for solar energy technology and development, and analyzed the current situation of solar energy development in certain regions and relevant policies (Dincer 2011; Ferreira et al. 2018; Sahu 2015). Some other scholars have focused on the technological innovation system, and put forward suggestions on the implementation of relevant policies on solar energy by studying the innovation diffusion method and path of solar energy technology (Kebede and Mitsufuji 2017; McEachern and Hanson 2008). These empirical analyses generally require the author to accumulate a more comprehensive professional knowledge and have certain subjectivity. It fails to realize the quantitative analysis of development in a certain field. Therefore, some scholars try to understand the global trend of solar energy research (de Paulo and Porto 2017; Du et al. 2014; Sampaio and Gonzalez 2017) and the cooperation mode of research institutions (Sanz-Casado et al. 2014) from the perspective of literature. The analysis of literature content can effectively, systematically, and objectively analyze the development status of this field, but the definition and quantitative criteria of the research object are different due to the cognition of the researchers themselves, which also reduces the objectivity of the result analysis. However, keywords cannot only quickly understand the key points of literature, but also identify the internal relations of relevant technologies, problems, and other knowledge entities (Li et al. 2016). Therefore, this paper considers using keywords in literature to explore the trend and rule of text fusion, and to find the evolutionary mechanism behind knowledge convergence.

In order to make a full sample analysis and knowledge convergence prediction of the literature in a specific field, this paper chose the complex network theory and used a link prediction model to make predictions. The method of complex network is one of the common bibliometric analysis methods (Zhu and Guan 2013), which can quickly analyze huge literature materials and master research hotspots and trends. There are some common

research perspectives, such as the citation network and the co-citation network (Pilkington and Meredith 2009), the collaborative network (Li et al. 2018), and the keyword co-occurrence network (Li et al. 2016). In the research of convergence, many scholars apply different modeling methods such as patent citation network (Kose and Sakata 2019) and co-classification network (Aaldering and Song 2019; Yun and Geum 2019) to explore the patterns and laws of technology convergence, and some scholars have applied citation networks in the field of solar energy to study the influence of scientific relevance and knowledge basis on the citation level of articles in the field of solar energy (Zhang and Guan 2017) and the path of solar technology development through the open innovation (de Paulo and Porto 2017). Link prediction model provides a method to predict the relationship between agents and their relations from an information science perspective based on the topological network. In recent years, link prediction methods based on the similarity between nodes have attracted more attention (Fahimnia et al. 2015; Lu and Zhou 2011). Compared with node attributes, the network structure similarity information of nodes is easier to obtain and is more reliable. Studies on link prediction mainly focus on the improvement of link prediction methods (Clauset et al. 2008; Liben-Nowell and Kleinberg 2007; Zhou et al. 2009), biological networks (von Mering et al. 2002; Yu et al. 2008), and recommendation systems (Kossinets 2006; Li et al. 2015). Among them, in the bibliometric study, some scholars have made predictions on the knowledge convergence, knowledge translation, and the potential trend of cooperative relations in the certain science field respectively from the perspectives of co-word networks (Choudhury and Uddin 2016), citation networks (Zhou et al. 2018), and cooperative networks (Kim and Diesner 2019). Gradually, link prediction model is applied to the research of the convergence of knowledge networks (Aaldering and Song 2019; Park and Yoon 2018). These studies confirmed that this method can be widely used in various types of networks and can effectively predict potential links. However, less attention has been paid to finding the convergence structural features of knowledge entities with the link prediction of keywords co-occurrence networks. If different knowledge entities appear in the same article, we assume that these knowledge entities converge to the article and are linked in pairs. Therefore, the keyword co-occurrence network is constructed accordingly. Therefore, this paper applies the link prediction method based on node similarity to explore whether the co-occurrence structure of keywords in solar energy literature can describe the characteristics of knowledge convergence. Additionally, by analyzing the network structure and statistical characteristics of keywords co-occurrence networks in solar energy literature, we try to find the potential motivation behind knowledge convergence and predict the trend of knowledge convergence.

The purpose of this paper is to explore the potential motivation of knowledge convergence and find the law of knowledge convergence. At the same time, a new literature knowledge convergence prediction model is proposed and applied to the solar energy field to predict the knowledge convergence trend. To describe the connection between knowledge entities in the field of solar energy, we established keywords co-occurrence networks by taking keywords from literature from 2008 to 2017 as nodes. Secondly, we used an evaluation index to evaluate the feasibility of the model with four similarity indicators respectively to find the knowledge convergence structural pattern in knowledge networks. Thirdly, the prediction results of knowledge convergence were compared with the real according to the evaluation to find the evolution rules of knowledge convergence. Fourthly, to find the knowledge convergence prediction cycle, this paper defines the prediction success rate index and analyzed the results. Fifthly, we tried to analyze the topological indices of hot keywords to find the rules of the topological relationships among them in the network. Finally, based on the existing rules, we made a

real prediction on the overall keywords co-occurrence network and obtained the trend of knowledge convergence in the field of solar energy, so as to provide guidance for researchers understanding of the whole field and the future research direction while putting forward suggestions for the determination of the direction of government funding. The main contribution of our work is to explore the potential relationship between agents in knowledge networks and propose a novel knowledge convergence prediction model.

The rest of this paper is arranged as follows: “Data and methodology” section introduces the data sources and relevant methods of this paper; “Results and analysis” section evaluates the accuracy of the algorithm and analyzes the experimental results; “Conclusion and discussion” section further discusses the experimental results. The main research idea of this paper is shown in Fig. 1.

## Data and methodology

### Data

The data of this paper is from Web of Science, which is recognized as one of the most authoritative index tools of science and technology literature in the world and can provide the most important research results in the field of science and technology. The data was downloaded on October 12, 2018. With “solar energy” as the subject, the literature in the Web of Science Core Collection from 2008 to 2018 was retrieved. A total of 6496 pieces of data was collected. Keywords, as a literature core content and key technology overview, can more directly mine the core topic and development trend of the field. Therefore, this paper chooses the keywords co-occurrence network to research the literature in the field of solar energy.

For the initial data, similar keywords are found and synonyms are normalized to reduce unnecessary nodes effectively. There are some studies on solar energy such as solar wind, solar flare, etc. in the literature, which are not related to new energy and irrelevant to the research content of this paper. So, it is necessary to conduct de-redundancy operations for such articles.

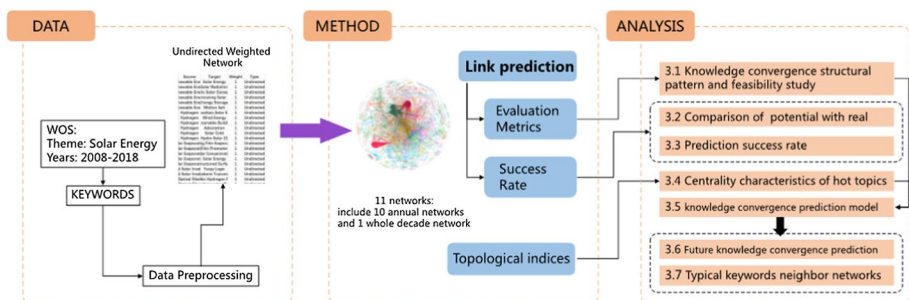


Fig. 1 The main research idea of this paper

## Methodology

### The construction of complex networks

The complex network is applied to construct the keywords co-occurrence network in the field of solar energy. The keywords existing in the same paper are considered to be inter-related, so all the keywords in the same paper can be connected in pairs. Taking keywords as nodes, keywords appearing in the same paper as edges, and occurrences as weights, we can construct an undirected weighted keywords co-occurrence network. The construction process of the network is shown in Fig. 2.

Taking set  $V = \{v_1, v_2, \dots, v_i\}$  as the nodes in the network, namely the keywords in the papers. The mathematical expression of the network can be defined as the following matrix  $K$ . The matrix  $K$  is the set of co-occurrence relations among all keywords in the set  $N$ .

$$K = \begin{bmatrix} k_{11} & \dots & k_{1i} \\ \vdots & \ddots & \vdots \\ k_{i1} & \dots & k_{ii} \end{bmatrix} \tag{1}$$

and

$$k_{ij} = \begin{cases} 0 & \text{if keyword } i \text{ and keyword } j \text{ do not appear in the same article} \\ m & \text{if keyword } i \text{ and keyword } j \text{ appear in } m \text{ airticls simultaneously} \end{cases} \tag{2}$$

According to the needs of this research, this paper established keywords co-occurrence networks from 2008 to 2017 based on the situation of keywords co-occurrence in each year, which is called the *keywords co-occurrence subnetwork*. At the same time, the network of all keywords co-occurrence from 2008 to 2017 is established, which is called the *overall keywords co-occurrence network*. Also, the CN values of all the unknown edges predicted from 2008 to 2017 are taken as weights to establish the unknown edge network, which is called the *CN prediction network*.

### Link prediction model

Link prediction method has been proved by many scholars to be well applied to the prediction of inter-agent relationships (Lu and Zhou 2011). According to the structural

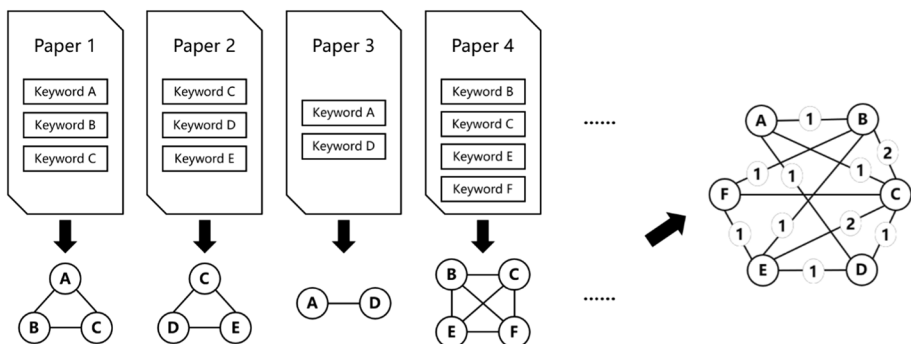


Fig. 2 The construction of the keywords co-occurrence network

characteristics of the keywords co-occurrence network, this paper chooses link prediction method to predict the connection of potential keyword links.

(1) Overview of link prediction model

For any undirected network  $G(V, E)$ ,  $V$  is the set of nodes and  $E$  is the set of edges. The number of nodes in the network is  $N(N = |V|)$ , and the number of edges is  $M(M = |E|)$ . Let  $U = N(n - 1)/2$ , that is,  $U$  is the universal set of all nodes in the network. Based on the common neighbors of nodes, each pair of nodes  $(v_x, v_y)$  without edges is assigned a score value named  $s_{xy}$ . Since  $G$  is an undirected network,  $s_{xy} = s_{yx}$ . The unconnected node pairs are sorted from largest to smallest by score value, and the node in the front is most likely to have edges in the future.

In order to test the accuracy of the algorithm, the known edge  $E$  is generally divided into two parts: the training set  $E^T$  is used to calculate the score value of the node pair of the known edge, and the test set  $E^P$  is used to test the accuracy of the algorithm. Obviously,  $E = E^T \cup E^P$  and  $E^T \cap E^P = \emptyset$ . In general, we choose 10% known edges in the network as test edges randomly, delete these edges from the network, and predict the deleted edges based on the information of the remaining 90% edges.

According to the description of the link prediction model above, Fig. 3 shows a simple calculation sample of link prediction based on common neighbor (Guan et al. 2016). The related similarity indices and evaluation index calculation method will be introduced in detail in the following text.

(2) Similarity index based on local information

Using the similarity between nodes to calculate the score value of node pairs is a kind of link prediction method. The greater the similarity between two nodes, the greater the possibility of a link between them (Lu and Zhou 2011). There are many methods to describe the nodes' similarity. The simplest similarity index based on local information is the common neighbor index, which has the advantage of low computational complexity and is suitable for large-scale networks. There are other more complex methods

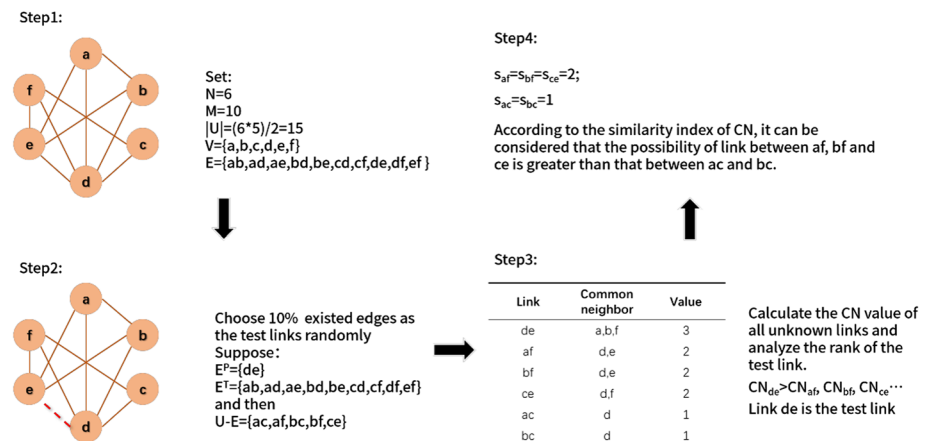


Fig. 3 A calculation sample of link prediction algorithm based on common neighbor

such as based on random walk or likelihood analysis, which consider more information and have high algorithm complexity, but the result performance is not necessarily better than the algorithm based on local information (Lu et al. 2009). Different similarity indices based on local information reflect different knowledge convergence preferences. In order to find the attachment mechanism of knowledge entities convergence, we choose four indicators to measure the knowledge convergence behaviors from different perspectives. They are common neighbor index (CN), Jaccard index (JACCARD), preferential attachment index (PA), resource allocation index (RA).

The basic assumption for applying CN index in link prediction is that two unconnected nodes are more inclined to join if they have more common neighbors. We use CN as an abbreviation of the value of common neighbors below.

CN index is defined as (Lu et al. 2009):

$$s_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)| \tag{3}$$

where  $\Gamma(x)$  represents the neighbor set of node  $v_x$ ; similarly,  $\Gamma(y)$  represents the neighbor set of node  $v_y$ .

JACCARD index considers the percentage of keywords common neighbors in their all neighbors. Because larger neighbors of keyword pairs might generate more common neighbors, JACCARD weakens the influence of this situation. Jaccard index Jaccard index.

JACCARD index is defined as (Jaccard 1901):

$$s_{xy}^{JACCARD} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \tag{4}$$

PA index only considers the number of keywords neighbors, which reflects the preference of convergence.

PA index is defined as (Barabasi and Albert 1999):

$$s_{xy}^{PA} = k_x k_y \tag{5}$$

where  $k_x$  represents the degree of node  $v_x$  and  $k_y$  represents the degree of node  $v_y$ , as well.

According to the common neighbors' degree, RA assigns corresponding capacity of knowledge transfer to common neighbors (that is, weights are assigned to them).

RA index is defined as (Zhou et al. 2009):

$$s_{xy}^{RA} = \sum_{z \in |\Gamma(x) \cap \Gamma(y)|} \frac{1}{k_z} \tag{6}$$

### (3) The evaluation indexes

Considering the algorithm itself, this paper chooses AUC (Hanley and Mcneil 1982) to measure the accuracy of the link prediction algorithm. AUC can be understood as the probability that the score value of randomly selected edge in the test set is greater than the score value of randomly selected nonexistent edge. If the score value of one randomly selected edge in the test set is greater than the score value of the nonexistent edge randomly selected, add 1 point; If the two sides are equal, add 0.5. Compare n times independently, and then AUC can be defined as (Lu and Zhou 2011):

$$\text{AUC} = \frac{n' + 0.5n''}{n} \quad (7)$$

where  $n'$  means the number of times that the score value of edge in the test set is greater than the score value of nonexistent edge, and  $n''$  means the number of times that the score value of edge in the test set is equal to the score value of nonexistent edge.

### Prediction success rate indicators

AUC and Ranking Score only measure the accuracy of the model from the perspective of the algorithm, but the accuracy of prediction of knowledge convergence was not measured. Therefore, according to the idea of AUC algorithm, this paper proposes a prediction success rate indicator. Let  $X = U - E$ , that is,  $X$  is the set of all unconnected edges. Through the calculation of link prediction algorithm based on common neighbor, each pair of unconnected edges has a CN value. Prediction success rate (PSR) can be defined as:

$$\text{PSR} = \frac{1}{\sum_{i \in X} s_i} \sum_{i \in X^p} s_i \quad (8)$$

where  $s_i$  represents the CN value of edge  $i$ , and  $X^p$  represents the edge that is predicted successfully.

### The hypernym–hyponym model

During the literature keywords processing, due to the diversity of vocabulary and the subjectivity of the author, there are many words with similar meanings. Since there is no uniform standard for keyword settings, the processing of such texts needs to measure the similarity of word meaning as a factor for the success of prediction.

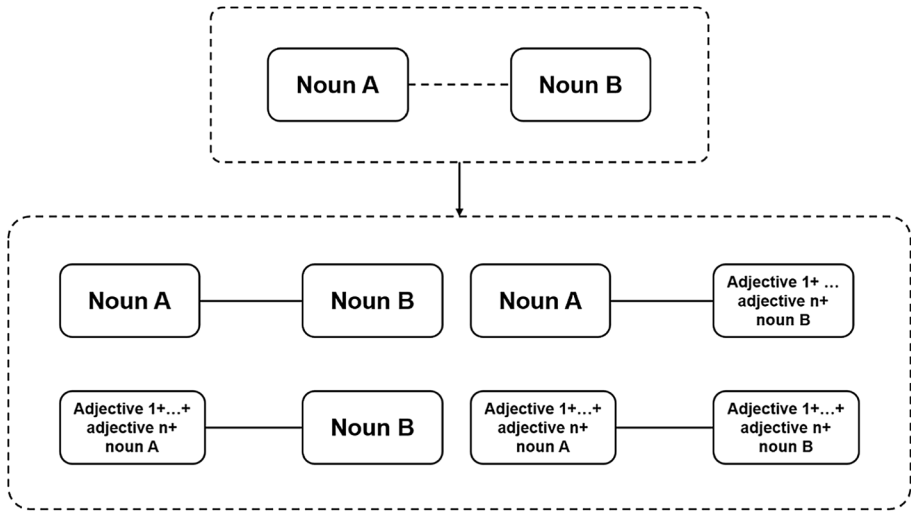
In the process of text processing, the calculation of word meaning similarity can be measured according to the degree of superposition between hypernyms hyponyms (Yang and Ping 2018). As can be seen from the characteristics of keywords, most keywords are a collection of several nouns or the structure of "adjective + noun". According to the characteristics of language, the collection of nouns, except for the central noun, all the other nouns modify the central noun and actually act as adjectives. According to this characteristic, we may consider that "noun A" is the hypernym of "adjective 1 + ... adjective n + noun A". The latter is the subdivision of the former topic. If "noun A" is predicted to have a knowledge convergence of other words, we can still regard a success of prediction in the case of the conjunction of the "adjective 1 + ... adjective n + noun A" with the same word. The four ways to predict successfully are shown in Fig. 4.

### Topological index

This paper focuses on the prediction of potential topics integration. It focuses on specific nodes and needs to analyze local topological indicators. Therefore, we choose four indicators to measure: weighted degree (WD), closeness centrality (CC), betweenness centrality (BC), and eigenvector centrality (EC).

- (1) Weighted degree (WD)





**Fig. 4** Four situations of predicting successfully

The weighted degree can describe the degree of correlation between a certain keyword and another keyword. The weighted degree of a node can be defined as (Linton and Freeman 1978):

$$S_{WDi} = \sum_{j \in N_i} w_{ij} \tag{9}$$

where  $w_i$  is the weight between  $v_i$  and  $v_j$ .

(2) Closeness centrality (CC)

The closeness centrality of a node is the sum of the shortest distance between the node and all other nodes in the graph. The closeness centrality of a node can be defined as (Brandes 2001):

$$S_{CCi} = \frac{N - 1}{\sum_{j=1, j \neq i}^N d_{ij}} \tag{10}$$

where  $d_{ij}$  represents the shortest distance between points  $i$  and  $j$ .

(3) Betweenness centrality (BC)

The betweenness centrality of a node describes the degree to which a point is "in the middle" of other "node pairs" in the graph. The greater the betweenness centrality, the more "mediating" it is in the diagram. The betweenness centrality of a node can be defined as (Goh et al. 2003):

$$S_{BCi} = \frac{2C_{ABi}}{(N-1)(N-2)} = \frac{2 \sum_j^N \sum_k^N \left( \frac{g_{jk}(i)}{g_{jk}} \right)}{(N-1)(N-2)} \tag{11}$$

where  $C_{ABi}$  represents the absolute betweenness centrality of a node,  $g_{jk}$  represents the number of existing paths between node  $j$  and node  $k$ ,  $g_{jk}(i)$  represents the number of paths between node  $j$  and node  $k$  which pass through node  $i$ .

(4) Eigenvector centrality (EC)

Connecting to a node with a high score node contributes more to the node’s score than connecting to a node with a low score. The high eigenvector score means that a node is connected to many nodes that have high scores of their own. The importance of a node depends on not only the number of its neighbors (that is, the degree of the node), but the importance of its neighbors.

The eigenvector centrality of a node can be defined as (Bonacich and Lloyd 2001):

$$S_{ECi} = \frac{1}{\lambda} \sum_{z \in N_t} S_{ECz} \tag{12}$$

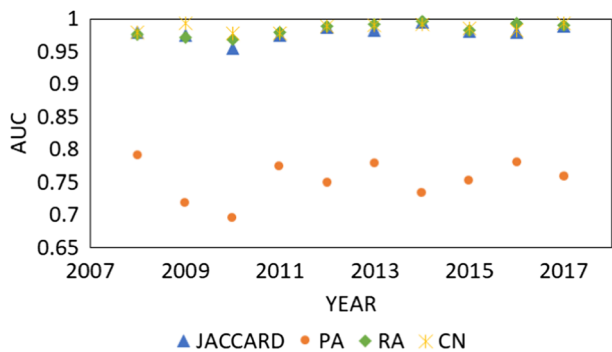
where  $N_t$  is the set of neighbor nodes which directly connected with node  $t$ , and  $\lambda$  is the number of nodes in the set of  $N_t$ .

## Results and analysis

### Knowledge convergence structural pattern and feasibility study

In this part, we apply four indicators to explore which local structural knowledge convergence pattern is the closest. We choose AUC to evaluate the feasibility of this model with different indicators. Figure 5 shows the AUC values of the links prediction model with four indicators respectively for the calendar years from 2008 to 2017. As can be seen from the figure, the AUC values of the four indicators were all greater than 0.65, especially the AUC values of CN index, JACCARD index, and RA index were greater than 0.95. Therefore, the link prediction algorithm based on local information similarity indicators has higher accuracy in this kind of knowledge networks.

**Fig. 5** The accuracy evaluation of link prediction model in 2008–2017



From these results, we can find structural mechanism of knowledge convergence. First, the accuracy of the algorithm based on PA index is obviously lower than the other three types. Although a few keywords have a large degree value in the knowledge network, it does not mean that these keywords are more likely to generate knowledge convergence. Because of this attachment pattern, knowledge convergence in knowledge networks tends to diversify and lead to innovation. Secondly, the algorithm accuracy of CN, JACCARD, and RA is roughly similar, in which CN is slightly higher. These three indicators all consider common neighbors of keywords pairs, while JACCARD index takes into account neighbors' size of keywords pairs, and RA index distinguishes the importance of different common neighbors. However, in actual researches, more common researches can also generate scholars' new ideas, while the size of relevant research scale and the difference in the heat of common researches have no obvious influence on knowledge convergence. Based on these findings, we will choose CN index to measure structural characteristics of knowledge convergence in this knowledge network.

### Comparison of predictive keyword links with real keyword links

According to the characteristics of the commonly correlated keywords of the tested edge, we calculated the unknown edges CN values and the CN value of the unknown edge was ranked year by year. For high ranking keyword links, their topological relations indicated that the keyword links had the possibility of integration in the future, and the forms of integration were richer with more commonly correlated keywords.

Different literature has different authors, which leads to the subjectivity of keyword settings in literature. However, as the writing of literature must be based on the existing results of other literature for evidence-based innovation, the setting of keywords is not too subjective. That is due to that authors tend to use the same terminology and the normalization of some synonyms has been completed in the data preprocessing. For the existence of a large number of subtopics of a certain topic, we choose the hypernym–hyponym model for approximate processing.

With the deepening of research, a certain keyword could generate different subtopics. A typical example is shown in Table 1. In 2008, it was predicted that *efficiency* and *hydrogen* might potentially converge: *Efficiency* and *biohydrogen* appeared in the same paper in 2009; *efficiency* and *hydrogen production* appeared in the same paper in 2011. There are other similar cases listed in Table 1. Although there are some differences in the meaning between *biohydrogen* and *hydrogen*, we can consider that *biohydrogen* is a hyponym of *hydrogen*, namely a subtopic of *hydrogen*. Such a result could be regarded as the success of our prediction in 2008. The occurrence of a hyponym is a subdivision of the topic, which could better promote more refined research in the field of solar energy.

According to such rules, we selected the keyword links ranking the top five in each year's CN value for verification and analysis, as shown in Fig. 6. Light orange represents the prediction of the potential integration of the keyword links in that year and dark orange represents that the keyword links appeared in the same paper in the corresponding year, namely a successful prediction.

Three rules can be found shown in the figure by comparing the predicted results with the real ones. First of all, keyword links CN values ranking higher in 2009 and 2010 did not appear for the next few years. Through the analysis of the new energy field from 2009 to 2010, we know that the global economic crisis occurred in 2008. During the three years, the price of crude oil rose, and the country did not have the extra economic capacity to

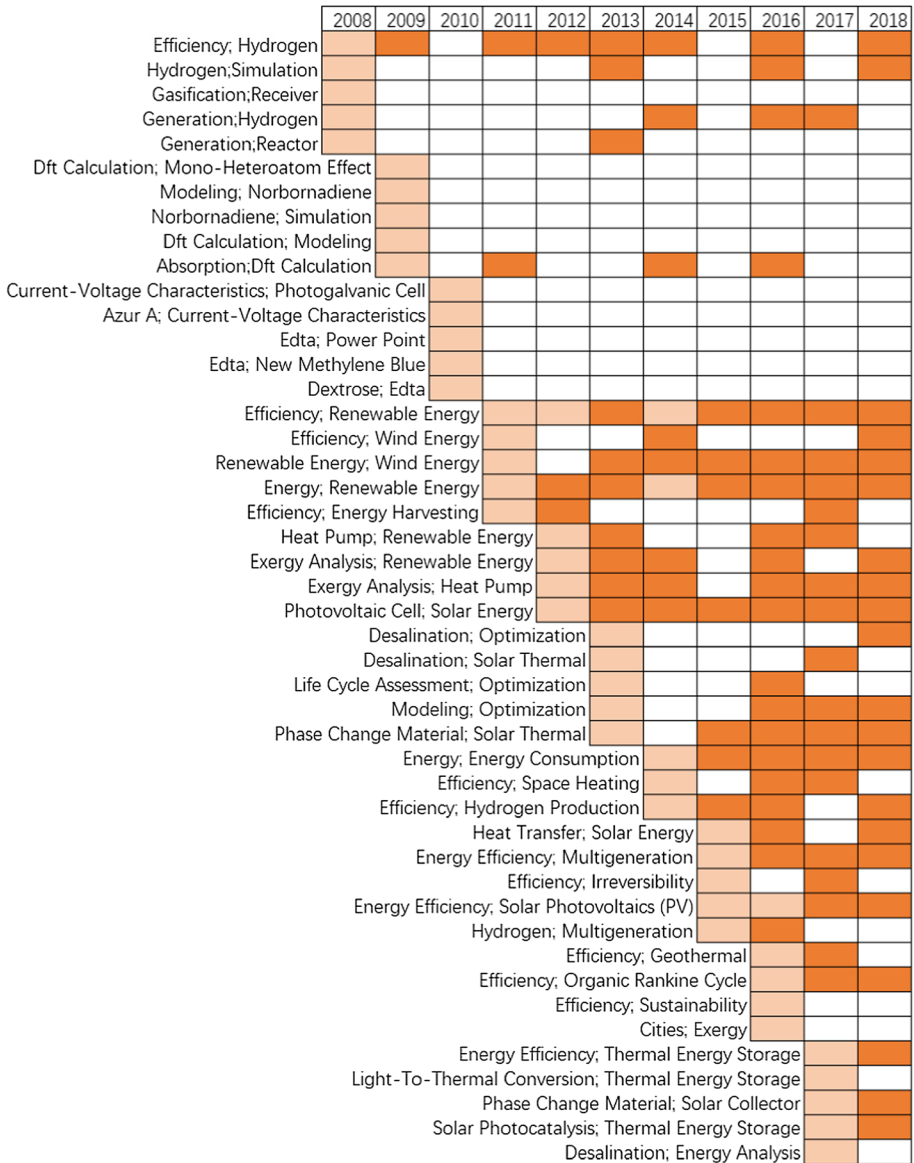
**Table 1** The links between *efficiency* and *hydrogen* in following years after prediction in 2008

Source	Target	Year
Efficiency	Biohydrogen	2009
Efficiency	Hydrogen production	2011
Electrolyze efficiency	Solar hydrogen system	2011
Efficiency	Hydrogen production	2012
Efficiency	Hydrogen production	2012
Efficiency	Hydrogen	2013
Light conversion efficiency	Photobiological hydrogen	2013
Exergy efficiency	Hydrogen production	2014
Efficiency	Hydrogen	2015
Efficiency	Hydrogen	2016
Efficiency	Hydrogen production	2016
Energy efficiency	Hydrogen storage	2016
Energy efficiency	Hydrogen system	2016
High efficiency for visible Photocatalytic hydrogen generation	Low hydrogen adsorption energy	2017
Energy efficiency	Hydrogen production	2018
Energy efficiency	Levelized cost of hydrogen	2018

develop and use new energy. Therefore, the research on new energy focused more on basic research, that is, the research on materials needed to collect solar energy. From the keywords with high ranking CN values in 2009 and 2010, it can be found that keywords related to research on new energy materials, such as *Norbornadiene*, *DFT Calculation*, *EDTA*, and *New Methylene Blue*, appeared more frequently, but after these two years, such keywords largely disappeared. Also, it is shown that scholars no longer paid any attention to these topics generally after the basic research was completed. Most of the potential keyword links predicted from 2008 also skipped 2009 and 2010 and were verified successfully in subsequent years. This also verifies that a big event like the financial crisis has a certain impact on the research direction of scholars. Secondly, except for the first three years, the keywords with higher CN values had been largely verified in the following years. This proves that keyword links based on the link prediction of common neighbors were more likely to be knowledge convergence by scholars in future studies without the impact of big events. It guides scholars to integrate research topics. Thirdly, in the process of verification, we found that if the predicted potential keyword links did not appear in the following three or four years, they would not appear in the future. This shows that scholars have certain timeliness in the research direction and field of solar energy. With the emergence of new topics, some topics will die out with time.

### Prediction success rate

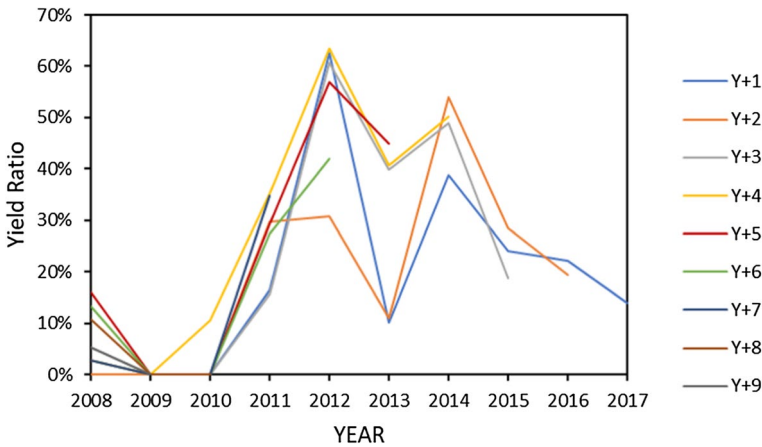
The previous section makes a comparative analysis of the relationship between the potential of knowledge convergence prediction and the real situation. Some predictions were successfully verified in the second year, some took two, or even three or more years to be



**Fig. 6** Potential knowledge convergence in 2008–2017 (rankings in the top five)

verified, and some were never verified. Then, how many years does the knowledge convergence predicted by the link prediction model usually take to be verified?

In order to solve this problem, we calculated the prediction success rate in different years after 1, 2, ... *n* years respectively, which is shown in Fig. 7. In order to unify the research, we selected all the keywords with a CN value greater than 5 and verified them year by year. As shown in the figure, the line *Y* + 4 is at the top of the figure, indicating that knowledge convergence was more likely to occur in the fourth year after the prediction.



**Fig. 7**  $Y+n$  years prediction success rate in the past 10 years

Therefore, researchers should pay extra attention to knowledge convergence that has been predicted over the past four years.

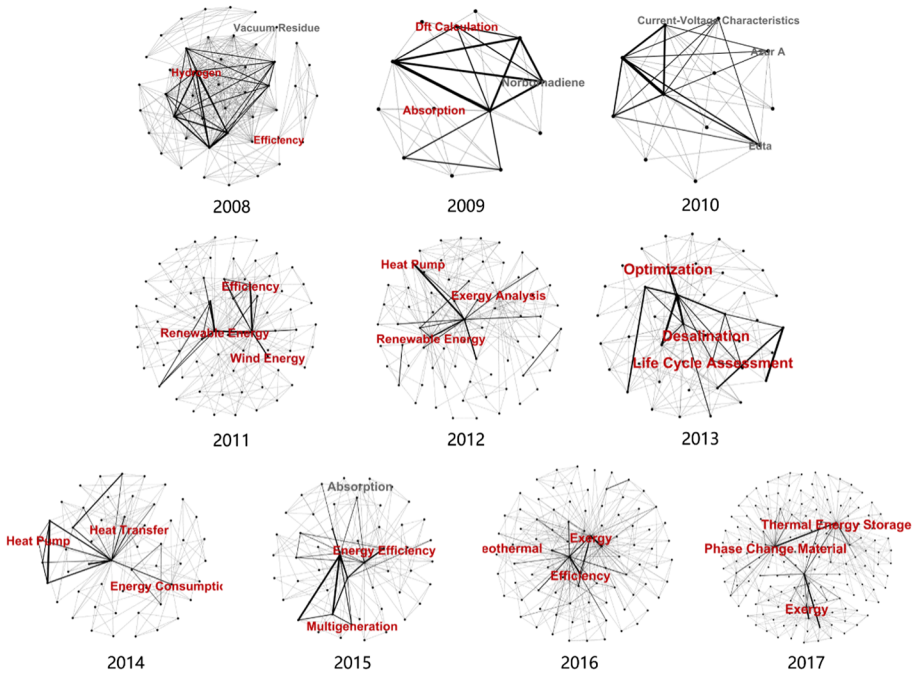
### Centrality characteristics of hot topics

Because some keywords appear for a short time, but also have more common neighbors with other keywords, the link prediction model may fail to predict. But these keywords didn't appear frequently, and the keywords with a large CN value appeared frequently in the link prediction of potential knowledge convergence every year. We tried to identify the topological characteristics of these keywords which may integrate with many keywords, so as to find the rules.

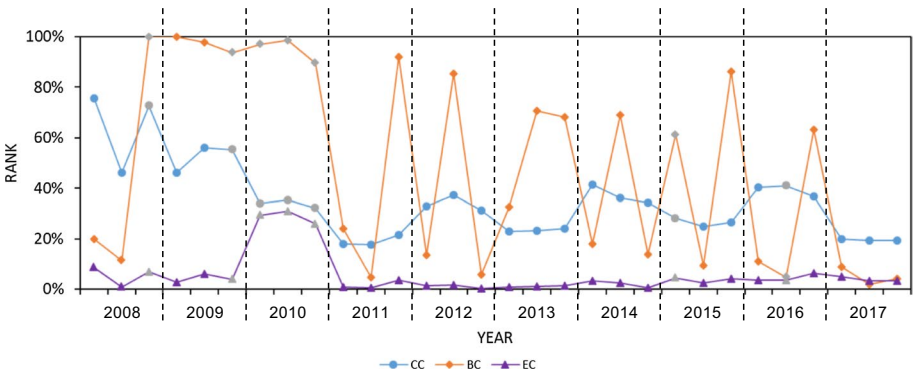
We selected the keyword links with a high CN value ranking every year, calculated the occurrence frequency of each keyword to find the top three keywords each year, and observed their distribution in the network, as shown in Fig. 8. The figure shows subnetworks of a layer of neighbors of predicted hot keywords each year. As we can see from the figure, most of the hot keywords in 2009 and 2010 were at the edge of the network, and the density of their subnetworks was low. In the past two years, especially in 2010, the link prediction of hot keywords had not been verified. In addition to these two years, hot keywords of the other years were mostly in the center of the network and the subnetwork densities were all higher.

According to the rules in Fig. 8, we calculated the closeness centrality (CC), betweenness centrality (BC), and eigenvector centrality (EC) of the keywords in each year to find the ranking of the keywords annually and to observe their distribution characteristics. Figure 9 shows the ranking percentage of the three indicators of hot keywords each year. The lower the percentage, the higher ranking, otherwise, the lower ranking.

From the figure, it can be observed that eigenvector centrality (EC) of all these keywords ranked higher. Closeness centrality (CC) ranking was between 20 and 80%, which was not characterized by prominence. The ranking of betweenness centrality (BC) had a large range and showed a state of polarization. However, after observing each year's successful prediction, the keywords with a higher ranking of betweenness centrality were easy



**Fig. 8** Distribution examples of potential knowledge convergence in keywords co-occurrence subnetwork (gray words indicate that the predicted links of the keywords are not verified)



**Fig. 9** The distribution of centrality index ranking with hot keywords from 2008 to 2017 (the grey dot indicates that the knowledge convergence prediction with this keyword was not verified in the following years)

to integrate with other keywords, while the keywords with lower-ranking were less likely to integrate, but they also had a trend.

### Literature knowledge convergence prediction model

According to the above results, the following prediction model can be summarized for literature knowledge convergence prediction, as shown in Fig. 10.

Keywords in the literature were used to construct the keywords co-occurrence network and to calculate the CN value of the nonexistent links. The links with a higher CN value have a higher possibility of the trend of knowledge convergence. In the meantime, we can construct a CN network that takes the keyword as nodes, the nonzero CN value links as edges, and CN value as weight. Then, the calculated weight degree (WD), the calculated betweenness centrality (BC) and the calculated eigenvector centrality (EC) of the higher-ranked keywords are defined as hot topics, core topics, and easily fusion topics separately. According to the three types of keywords combined with the prediction of the links with higher CN values, a new trend of knowledge convergence would be found in a certain field.

### Future potential knowledge convergence prediction

According to the link prediction analysis for the past decade, we constructed the *overall keywords co-occurrence network* based on all the data from 2008 to 2017 and use the link prediction model based on the common neighbor for this network. Therefore, we calculated the CN values of all the unknown edges. According to the results obtained, a new network was constructed for all the nonexistent edges with a CN value greater than 1, where the weight was the CN value of the edges. According to the *CN prediction network* generated by the results of link prediction, we analyzed the weighted degree (WD), betweenness centrality (BC), and eigenvector centrality (EC) of the keywords with a higher ranking in the *CN prediction network*.

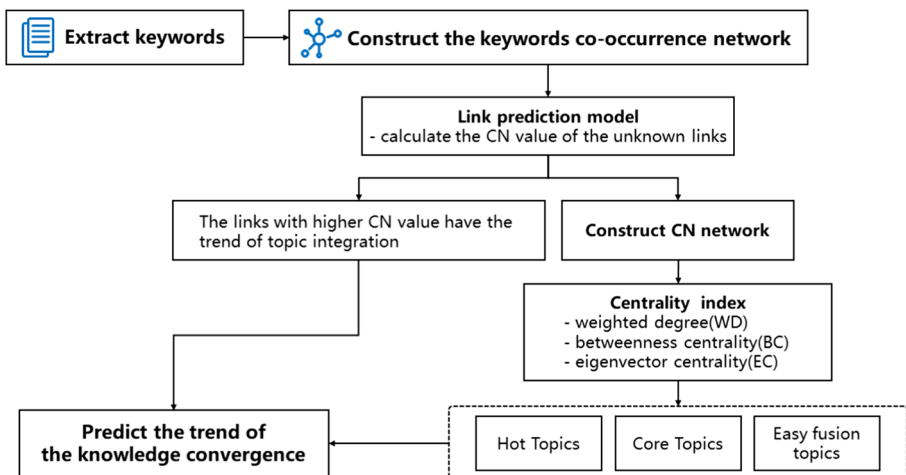


Fig. 10 Literature knowledge convergence prediction model



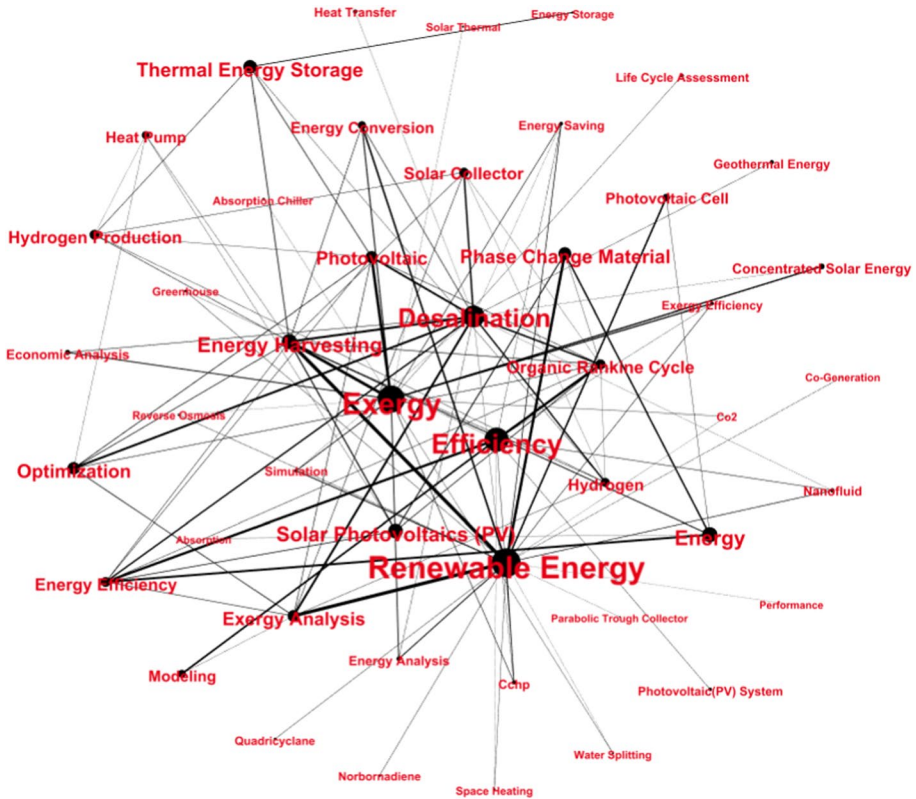
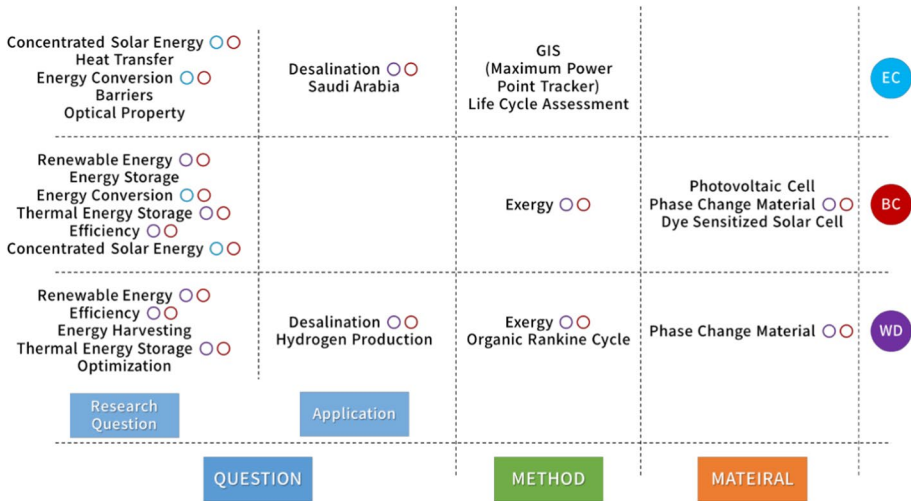


Fig. 11 The visualization of CN prediction network (CN > 200)

The unknown edge would be more likely connected in the future with the high CN value. In order to more intuitively see the potential knowledge convergence with greater possibility, we filtered out the CN value ranking the top 100 edges (that is, where the CN value of the unknown edge was more than 200), and the network is shown in Fig. 11 (where the higher weighted degree value of the keyword is shown with the larger node and word size). The emergence of the links in this network predicted the future potential knowledge convergence.

Figure 12 lists the top 10 keywords for the values of weighted degree (WD), betweenness centrality (BC), and eigenvector centrality (EC). A higher weighted degree (WD) indicated that the keyword would have more common words with other keywords and have a higher possibility of knowledge convergence. Such keywords are defined as hot topics. As we can see from the picture, *renewable energy* is in a higher position and *energy harvesting*, *thermal energy storage*, and *efficiency* are all hot keywords of knowledge convergence. In the application scenarios, *desalination* and *hydrogen production* tended to generate knowledge convergence and the technologies with a high possibility of knowledge convergence were *exergy* and *organic Rankine cycle*. In terms of materials, the research on *phase change material* was more likely to be converged with other topics.

Betweenness centrality (BC) describes the extent to which a certain keyword is at the core of knowledge convergence, and keywords with a higher mediation centrality are



**Fig. 12** List of top 10 keywords for the values of weighted degree (WD), betweenness centrality (BC), and eigenvector centrality (EC)

defined as core topics. Scholars pay close attention to the research question, *renewable energy*, *energy storage*, *energy conversion*, and *efficiency* became integrated subjects at the core of the topic. The *exergy* approach, as well as *photovoltaic cell*, *dye-sensitized solar cell*, and *phase change material*, were at the center of research integration. Among them, the problems of renewable resources, energy storage, efficiency, exergy, and the weighted value and intermediate centrality of phase change materials were all relatively high, which are the hot issues of knowledge convergence.

Eigenvector centrality (EC) describes the importance of keywords connected by a certain keyword. Keywords with high eigenvector centrality are defined as topics that are easy to converge with hot topics. Scholars research on *heat transfer*, *energy conversion*, *barriers*, and *optical property* are apt to converge with hot topics. *Desalination* and *Saudi Arabia* are also subject to hot topics. In terms of methods, *GIS technology*, *MPPT controller (Maximum Power Point Tracker)*, and *life cycle assessment* are likely to fuse with hot topics. *Energy conversion* is not only the core topic of knowledge convergence but also the integration of hot issues. The *desalination* itself is a hot issue and it could also be easily converged with other hot issues.

### Analysis of typical keywords neighbor networks

According to the *CN prediction network* and the topological characteristics of the keywords, we selected two typical subnetworks in the CN prediction network, as shown in Fig. 13. Figure 13a shows the *desalination* neighbor network, which is a practical application of solar technology. In this neighbor network, keywords of *desalination*, *solar collector*, *hydrogen*, *energy harvesting*, *exergy analysis*, *energy efficiency*, *optimization*, *photovoltaic*, and *organic Rankine cycle* are approximately fully connected. Therefore, the combination of these knowledge entities may be a trend of future research. For example, *exergy analysis* should be applied to analyze the *energy efficiency* of *desalination* in terms of *photovoltaic* technology, so as to further study how to improve the *efficiency* of

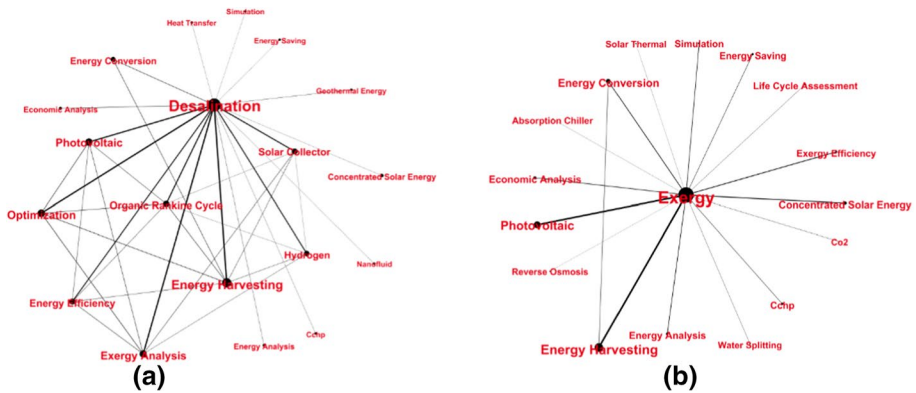


Fig. 13 Typical keywords' neighbor networks

*desalination* and reduce costs. Also, *organic Rankine cycle* can be considered for *solar collectors* to improve the *efficiency* of *desalination*. Also, *desalination* with *economic analysis*, *geothermal energy*, and *combined cooling heating and power (CCHP)* would be easy to converge in future studies. Figure 13b shows the neighbor network of *exergy*. Compared with the *desalination* neighbor network, this neighbor network is simple and has a star shape. Based on existing relevant research, scholars can generate creativity by combining *exergy* with other keywords in neighbor networks. *Exergy* is a technology that measures energy efficiency and environmental impact. By using *exergy* technology, it can be measured whether new materials or new technologies can improve efficiency or reduce environmental impact. For example, *exergy* can be used to measure the *efficiency* of the *combined cooling heating and power (CCHP)* system for urban power generation compared to other systems. Notably, *exergy* can be combined with *life cycle assessment* to more comprehensively describe the impact of technology on the environment and its power generation efficiency.

For the prediction results, scholars can find the keywords they are interested in firstly, then study the potential knowledge convergence and expand the topic based on existing research, so as to find a better solution to the problem. We can also focus on hot topics and core topics in this field, finding the direction of the potential knowledge convergence around such issues and conducting further research on this basis.

## Conclusion and discussion

### Potential motivations of knowledge convergence

In order to explore the potential motivation behind knowledge convergence, we introduced link prediction model and compared the prediction accuracy of four different prediction indexes. We find that common neighbors—two different agents that connect the same knowledge entity—are more likely to generate new research. Also, it is confirmed in the comparison of predicted and real keywords links each year. However, the comparative verification also found that the index of common neighbor (CN) alone is not completely reliable due to the changes in the external environment. Therefore, we introduced the centrality

measurement method to eliminate the false prosperity caused by the local substructure of the network at a certain period.

Keywords with a high weighted degree (WD) are regarded as hot topics, which are usually the star topics studied by scholars. Keywords with high betweenness centrality (BC) are regarded as core topics. They are usually the key topics connecting two fields which are also important knowledge entities for convergence. Keywords with high eigenvector centrality (EC) are regarded as easy fusion topics, as they are easier to be combined with hot topics to generate new ideas from scholars. The recognition of these three kinds of keywords can effectively reduce the impact of marginal topics in the field at a certain period on the prediction, which can predict more accurately in knowledge convergence. Additionally, the recognition of these keywords is also of great help for scholars to understand the development and innovation of the whole field.

Due to the subjective innovation of scholars and the flexibility of keywords, a single prediction model may fail. Therefore, we proposed the literature knowledge convergence prediction model, which adds the measurement of the centrality of the main body in the network based on link prediction, so as to eliminate the influence of some external factors and make a more accurate prediction and judgment.

### Characteristics of knowledge convergence

The knowledge convergence of the literature in the field of solar energy has its unique characteristics. Due to the high accuracy of the CN-based link prediction, we can confirm that scholars tend to integrate the knowledge entities with common entities that have been studied together in previous literature in their research, which leads to knowledge innovation. Based on CN index, we can use link prediction to find more characteristics of knowledge convergence of scientific research.

Scientific research is a long process, which may last for several months, one year, three years, five years, and even decades. Therefore, the prediction of knowledge convergence also has the problem of the time cycle. By comparing the predicted results of different years with those of the following years, it was found that predicted knowledge convergence could take place in about three or four years. In one to three years some knowledge entities would converge and a peak would be reached in 4 years that would then gradually level off. If the predicted knowledge convergence does not take place after three or four years, it may be that a certain keyword has gradually died out, which is easier to identify in the actual research. Therefore, we should pay attention to the prediction of knowledge convergence of solar energy literature for four consecutive years and then find the trend of knowledge convergence in the future based on the overall prediction.

The research on the centrality index of keywords can be concluded as follows: firstly, keywords with higher eigenvector centrality were connected with more important keywords, which are easier to converge with other knowledge entities. Secondly, keywords with a higher betweenness centrality were at the core of the network. The connection of many keywords depends on these keywords. In the meantime, these knowledge entities are easier to converge with others. Finally, keywords with lower betweenness centrality were at the edge of the network, but the keywords that rely on common connection keywords can generate links, although the possibility is lower. Although it is difficult to generate knowledge convergence in the actual situation for this kind of keyword, from the perspective of prediction, it has a trend of convergence which is the important perspective that scholars need to focus on.

### Suggestions on the research direction for scholars

Through our research on the potential drivers of knowledge convergence, we apply the prediction model to the field of solar energy. We can get research trends and research priorities in this field: The research on solar energy can be divided into six issues, namely energy storage, efficiency, cost, ecological effect, application, and hybrid photovoltaic systems. The sub-issues and knowledge convergence derived from the research are shown in Fig. 14. These six questions are all the focus of solar energy research, which can produce multi-angle, multi-directional convergence. In this figure, the focus is on *technical innovation*, which directly reduces the *cost* of photovoltaic technology, while *energy storage*, *material*, and *life time* of devices and other issues directly rely on *technical innovation*. Also, the *cost* could be influenced by the location problem (*distribution*), which requires the study of *temperature* and *solar irradiance*. These problems directly affect the efficiency of power generation.

With the deepening of research, the topic is enriched and developed constantly, and scholars’ research on solar energy and related topics is refined and deepened continually. Therefore, our results are helpful for researchers to better understand the current research situation and future research direction of solar energy. Based on the prediction of the knowledge convergence of solar energy research, we have provided the following suggestions for relevant scholars to facilitate researchers to grasp the direction of research and for the government to formulate the policies of funding.

1. In the research of solar energy, energy conversion technology, energy efficiency and energy storage (battery) are the core issues of long-term concern in this field. From the practical point of view, these problems are also the key issues in the development of solar energy and the core problems that need continuous breakthroughs. Therefore, they should be paid attention to by scholars and relevant enterprises, and breakthroughs in core technologies can make solar energy better serve users’ daily life.

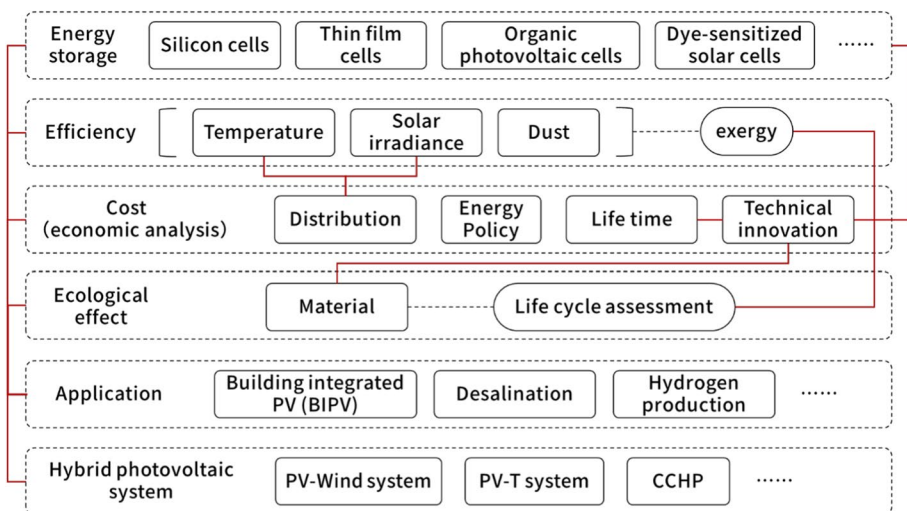


Fig. 14 Research direction and knowledge convergence of solar energy

2. Solar energy has the advantages of renewable and clean, but its acquisition cost and storage technology also limit its development. Single energy generation will be limited by many aspects. The development of renewable energy in the future requires scholars to study hybrid power generation technologies combining solar energy, wind energy, marine energy, and even traditional fossil energy. In fact, scholars have also made many contributions to the technology of hybrid power generation. For example, the Combined Cooling Heating and Power (CCHP) system combines hydropower, biological energy, wind energy, solar energy, and all other resources that can generate electricity or heat, and combines the power generation system with the Heating and Cooling system, providing a comprehensive way of energy supply for users.
3. The emergence of major technologies is conducive to promoting the knowledge flow, such as exergy, which provides a new perspective for the calculation of energy loss and energy efficiency. In future studies, exergy can also combine the topics of life cycle assessment and economic analysis to carry out cost analysis on various photovoltaic technologies. In addition, exergy can also be combined with CCHP to evaluate the energy loss and cost accounting of the comprehensive energy supply method proposed by the scholars in the whole life cycle. Also, the organic Rankine cycle is proposed to apply solar energy technology in seawater desalination scenarios, and could even be used to convert solar energy into electricity.
4. Scholars should pay attention to interdisciplinary research. For example, the application of the life cycle assessment method combines the economic cost and environmental cost in the solar energy conversion process with the research of traditional technologies, effectively promoting the development of solar energy towards a cleaner and more economical direction.
5. Both technology-driven knowledge convergence and interdisciplinary knowledge convergence greatly improve scholars' ability to solve problems. While focusing on hot issues, scholars should think more about how to integrate different factors of production and technological methods to solve problems more practically. The flow and convergence of knowledge in a wider range can more effectively promote innovation and social progress.

The above is the discussion on the potential motivation and characteristics of knowledge convergence and relevant suggestions for scholars based on the predicted results in the field of solar energy. Further research should incorporate the direction of knowledge convergence into the link prediction model and give more consideration to the time factor of topic evolution. In addition to the research on knowledge convergence, the dying topic and emerging topic are also important guiding factors for scholars' research, so as to improve our understanding of the historical dimension of the whole research field and make more accurate predictions of future knowledge convergence.

**Acknowledgements** The authors would like to express their gratitude to Jianhe Guan, Yixiong Zhang and Kaiming Wang who helped a lot during this work, including processing data and providing analysis suggestions.

**Author contributions** Duan Yueran organized and wrote the paper. Guan Qing provided several suggestions and revised the paper.

**Funding** This research is supported by grants from the National Natural Science Foundation of China (Grant No. 42001236, 71991481, 71991480), Beijing Outstanding Talent Training Foundation (Grant No. 2018000020124G151), and the Fundamental Research Funds for the Central Universities (Grant No. 2-9-2018-079).

**Availability of data and material** The data of this study is downloaded from Web of Science, which is a public literature database. It can be downloaded by any paying user.

**Code availability** In this paper, the experiment on network structure is completed by the software Gephi, and the link prediction model is completed by the custom coding in MATLAB.

## Compliance with ethical standards

**Conflict of interest** We have no potential conflict of interest or competing interests.

## References

- Aaldering, L. J., & Song, C. H. (2019). Tracing the technological development trajectory in post-lithium-ion battery technologies: A patent-based approach. *Journal of Cleaner Production*, *241*, 118343.
- Armaroli, N., & Balzani, V. (2016). Solar Electricity and solar fuels: Status and perspectives in the context of the energy transition. *Chemistry-A European Journal*, *22*(1), 32–57.
- Barabasi, A. L., & Albert, R. (1999). Albert, r.: emergence of scaling in random networks. *Science*, *286*(5439), 509–512.
- Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, *23*(3), 191–201.
- Boussemart, J. P., Leleu, H., Mensah, E., & Shitikova, K. (2020). Technological catching-up and structural convergence among US industries. *Economic Modelling*, *84*, 135–146.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, *25*(2), 163–177.
- Chae, Y. T., Kim, J., Park, H., & Shin, B. (2014). Building energy performance evaluation of building integrated photovoltaic (BIPV) window with semi-transparent solar cells. *Applied Energy*, *129*, 217–227.
- Chan, S. K., & Miyazaki, K. (2015). Knowledge convergence between cloud computing and big data and analysis of emerging technological opportunities in malaysia. In *Picmet '15 Portland International Center for Management of Engineering and Technology*, pp. 1501–1512.
- Choudhury, N., & Uddin, S. (2016). Time-aware link prediction to explore network effects on temporal knowledge evolution. *Scientometrics*, *108*(2), 745–776.
- Clauset, A., Moore, C., & Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature*, *453*(7191), 98–101.
- de Paulo, A. F., & Porto, G. S. (2017). Solar energy technologies and open innovation: A study based on bibliometric and social network analysis. *Energy Policy*, *108*, 228–238.
- Dennler, G., Scharber, M. C., Ameri, T., Denk, P., Forberich, K., Waldauf, C., et al. (2008). Design rules for donors in bulk-heterojunction tandem solar cells-towards 15 % energy-conversion efficiency. *Advanced Materials*, *20*(3), 579.
- Dincer, F. (2011). The analysis on photovoltaic electricity generation status, potential and policies of the leading countries in solar energy. *Renewable & Sustainable Energy Reviews*, *15*(1), 713–720.
- Du, H. B., Li, N., Brown, M. A., Peng, Y. N., & Shuai, Y. (2014). A bibliographic analysis of recent solar energy literatures: The expansion and evolution of a research field. *Renewable Energy*, *66*, 696–706.
- Fahimnia, B., Tang, C. S., Davarzani, H., & Sarkis, J. (2015). Quantitative models for managing supply chain risks: A review. *European Journal of Operational Research*, *247*(1), 1–15.
- Ferreira, A., Kunh, S. S., Fagnani, K. C., De Souza, T. A., Tonezer, C., Dos Santos, G. R., et al. (2018). Economic overview of the use and production of photovoltaic solar energy in brazil. *Renewable & Sustainable Energy Reviews*, *81*, 181–191.
- Fthenakis, V., Mason, J. E., & Zweibel, K. (2009). The technical, geographical, and economic feasibility for solar energy to supply the energy needs of the US. *Energy Policy*, *37*(2), 387–399.
- Goh, K. I., Oh, E., Kahng, B., & Kim, D. (2003). Betweenness centrality correlation in social networks. *Physical Review E*, *67*(1), 017101.
- Guan, Q., An, H. Z., Gao, X. Y., Huang, S. P., & Li, H. J. (2016). Estimating potential trade links in the international crude oil trade: A link prediction approach. *Energy*, *102*, 406–415.
- Hanley, J. A., & Mcneil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (Roc) curve. *Radiology*, *143*(1), 29–36.
- Jaccard, P. (1901). étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, *37*, 547–579.

- Jeong, D., Lee, K., & Cho, K. (2018). Relationships among international joint research, knowledge diffusion, and science convergence: the case of secondary batteries and fuel cells. *Asian Journal of Technology Innovation*, 26(2), 246–268.
- Kebede, K. Y., & Mitsufuji, T. (2017). Technological innovation system building for diffusion of renewable energy technology: A case of solar PV systems in Ethiopia. *Technological Forecasting and Social Change*, 114, 242–253.
- Kim, J., & Diesner, J. (2019). Formational bounds of link prediction in collaboration networks. *Scientometrics*, 119(2), 687–706.
- Kose, T., & Sakata, I. (2019). Identifying technology convergence in the field of robotics research. *Technological Forecasting and Social Change*, 146, 751–766.
- Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks*, 28(3), 247–268.
- Li, H. J., An, H. Z., Wang, Y., Huang, J. C., & Gao, X. Y. (2016). Evolutionary features of academic articles co-keyword network and keywords co-occurrence network: Based on two-mode affiliation network. *Physica A: Statistical Mechanics and its Applications*, 450, 657–669.
- Li, J., Xiong, J. G., & Wang, X. J. (2015). The structure and evolution of large cascades in online social networks. *Computational Social Networks, Csonet, 2015*(9197), 273–284.
- Li, Y., Li, H. J., Liu, N. R., & Liu, X. Y. (2018). Important institutions of interinstitutional scientific collaboration networks in materials science. *Scientometrics*, 117(1), 85–103.
- Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7), 1019–1031.
- Lim, S., Kwon, O., & Lee, D. H. (2018). Technology convergence in the Internet of Things (IoT) startup ecosystem: A network analysis. *Telematics and Informatics*, 35(7), 1887–1899.
- Linton, C., & Freeman. (1978). Centrality in social networks conceptual clarification. *Social Networks*.
- Lu, L. Y., Jin, C. H., & Zhou, T. (2009). Similarity index based on local paths for link prediction of complex networks. *Physical Review E*, 80(4), 046122.
- Lu, L. Y., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6), 1150–1170.
- McEachern, M., & Hanson, S. (2008). Socio-geographic perception in the diffusion of innovation: Solar energy technology in Sri Lanka. *Energy Policy*, 36(7), 2578–2590.
- Morfa, A. J., Rowlen, K. L., Reilly, T. H., Romero, M. J., & van de Lagemaat, J. (2008). Plasmon-enhanced solar energy conversion in organic bulk heterojunction photovoltaics. *Applied Physics Letters*, 92(1), 1551.
- Nocera, D. G. (2009). Chemistry of Personalized Solar Energy. *Inorganic Chemistry*, 48(21), 10001–10017.
- N'Tsoukpoe, K. E., Liu, H., Le Pierres, N., & Luo, L. G. (2009). A review on long-term sorption solar energy storage. *Renewable & Sustainable Energy Reviews*, 13(9), 2385–2396.
- Park, I., & Yoon, B. (2018). Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. *Journal of Informetrics*, 12(4), 1199–1222.
- Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management-1980-2006: A citation/co-citation analysis. *Journal of Operations Management*, 27(3), 185–202.
- Sahu, B. K. (2015). A study on global solar PV energy developments and policies with special focus on the top ten solar PV power producing countries. *Renewable & Sustainable Energy Reviews*, 43, 621–634.
- Sampaio, P. G. V., & Gonzalez, M. O. A. (2017). Photovoltaic solar energy: Conceptual framework. *Renewable & Sustainable Energy Reviews*, 74, 590–601.
- Sanz-Casado, E., Lascurain-Sanchez, M. L., Serrano-Lopez, A. E., Larsen, B., & Ingwersen, P. (2014). Production, consumption and research on solar energy: The Spanish and German case. *Renewable Energy*, 68, 733–744.
- Schulze, T. F., & Schmidt, T. W. (2015). Photochemical upconversion: present status and prospects for its application to solar energy conversion. *Energy & Environmental Science*, 8(1), 103–125.
- Shaner, M. R., Atwater, H. A., Lewis, N. S., & McFarland, E. W. (2016). A comparative technoeconomic analysis of renewable hydrogen production using solar energy. *Energy & Environmental Science*, 9(7), 2354–2371.
- Sharon, H., & Reddy, K. S. (2015). A review of solar energy driven desalination technologies. *Renewable & Sustainable Energy Reviews*, 41, 1080–1118.
- Sick, N., Preschitschek, N., Leker, J., & Broring, S. (2019). A new framework to assess industry convergence in high technology environments. *Technovation*, 84–85, 48–58.
- Stucki, T., & Woerter, M. (2019). The private returns to knowledge: A comparison of ICT, biotechnologies, nanotechnologies, and green technologies. *Technological Forecasting and Social Change*, 145, 62–81.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., et al. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417(6887), 399–403.



- Wong, W. Y., & Ho, C. L. (2010). Organometallic photovoltaics: A new and versatile approach for harvesting solar energy using conjugated polymetallaynes. *Accounts of Chemical Research*, *43*(9), 1246–1256.
- Yang, Y. H., & Ping, Y. (2018). An ontology-based semantic similarity computation model. In *2018 IEEE international conference on big data and smart computing (Bigcomp)*, pp. 561–564.
- Yu, H. Y., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., et al. (2008). High-quality binary protein interaction map of the yeast interactome network. *Science*, *322*(5898), 104–110.
- Yun, J., & Geum, Y. (2019). Analysing the dynamics of technological convergence using a co-classification approach: A case of healthcare services. *Technology Analysis and Strategic Management*, *31*(12), 1–18.
- Zhang, J. J., & Guan, J. C. (2017). Scientific relatedness and intellectual base: a citation analysis of un-cited and highly-cited papers in the solar energy field. *Scientometrics*, *110*(1), 141–162.
- Zhou, T., Lu, L. Y., & Zhang, Y. C. (2009). Predicting missing links via local information. *European Physical Journal B*, *71*(4), 623–630.
- Zhou, W., Gu, J. Y., & Jia, Y. F. (2018). h-Index-based link prediction methods in citation network. *Scientometrics*, *117*(1), 381–390.
- Zhou, Y., Dong, F., Kong, D. J., & Liu, Y. F. (2019). Unfolding the convergence process of scientific knowledge for the early identification of emerging technologies. *Technological Forecasting and Social Change*, *144*, 205–220.
- Zhu, W. J., & Guan, J. C. (2013). A bibliometric study of service innovation research: based on complex network analysis. *Scientometrics*, *94*(3), 1195–1216.