

# Scientific relatedness in solar energy: a comparative study between the USA and China

Jingjing Zhang · Yan Yan · Jiancheng Guan

Received: 12 May 2014 / Published online: 27 November 2014  
© Akadémiai Kiadó, Budapest, Hungary 2014

**Abstract** In this paper, we make use of keywords in scientific articles in solar energy during the period 2000–2013 to investigate scientific relatedness at the topic level (i.e. relatedness between topic and topic) and the country level (i.e. relatedness between topic and country). The bibliometric analyses show that both publications and knowledge topics exhibit significant rise, and China has exceeded the USA and developed into the largest scientific producer after 2010. We determine the degree of relatedness by means of the topics co-occurrence network and explore the evolving dynamic processes of scientific relatedness which indicates decreasing patterns in the two countries. The results also highlight differences between the research directions in the USA and China: in the USA “energy efficiency and environment” prove more developed, while in China “solar power” shows more central. This study assesses the extent to which the scientific relatedness exerts influence on the literature productivity at the country level. We find negative relationships between scientific relatedness and publications in both of countries. Our work has potential implications for the future policies with respect to the innovative research in the solar energy field.

**Keywords** Scientific relatedness · Co-occurrence · Innovation output · Solar energy

## Introduction

Knowledge production and scientific change are crucial to economic growth (Solow 1956). The rapid scientific development has substantially stimulated knowledge convergence which affects the progress of innovation (Kodama 2005). With that trend, knowledge and scientific relatedness has obtained great attentions (Joo and Kim 2010; Makri et al. 2010).

---

J. Zhang · Y. Yan · J. Guan (✉)  
School of Management, University of Chinese Academy of Sciences, Beijing 100190,  
People's Republic of China  
e-mail: guanjianch@ucas.ac.cn; guanjianch@fudan.edu.cn; guanjianch@buaa.edu.cn

The relatedness of scientific topic with other topics presents quite different characteristic. For instance, some scientific knowledge is diverse enough to interact with other knowledge (Geum et al. 2012), while some knowledge elements have low combinatorial potential. Knowledge productivity of each country is unbalanced (Guan and Ma 2007), and the geographical distribution of knowledge elements is also uneven. For example, some scientific studies of developed countries involve vast knowledge elements and own plenty of scientists (Livingstone 2010). The distribution of knowledge productivity has changed over the past decade due to the global trend and international collaborations (Zhang et al. 2010). The traditional centers like North America have been exceeded by some Asian countries (Matthiessen and Schwarz 2010).

The USA and China are most the same in the geographical area (9,629 million, 9,600 million, respectively) and the population of the USA is about a quarter of China's population. China's economy has been growing so fast in recent years and its GDP has surpassed Japan after 2010 and occupies the second position only to the United States. What's more, the average annual GDP growth rate of China was 9.9 % during 1952–2013 (National Bureau of Statistics of China, <http://www.stats.gov.cn/>). Some scholars considered that China's economy will be equal with that of the USA in size in about 10 years (Virmani 2005). China is also the second largest nation in energy consumption. Energy demand pressures promote the government to review its energy and economic policies to support the sustainable development (Li et al. 2007). Both the USA and China are trying to make impressive strides in harnessing sources of solar energy (Fthenakis et al. 2009). For example, China has the largest solar panels producer (Bradsher 2010) and the United States shows significant research strengths in solar energy field (Hassan et al. 2014). The growth of the scientific and technological capacities of China can be expected to reinforce its economic development (Zhou and Leydesdorff 2006). The USA and China are two nations which are prominent scientific producers. China has significantly increased its publication output in solar energy in recent years which is close in terms of publication output with the USA. A comparative study in solar energy in two prominent nations can contribute to the scientific research and professional technology of this field.

This raises some questions: How the relatedness between knowledge element and others evolves with time? What's the difference between the relatedness in the USA and China? To what extent does the scientific relatedness exert influence on the literature productivity? Previous studies have showed that technological relatedness can explain the entries and exits of cities which are specialized in those knowledge (Rigby 2013), and relatedness is a main driving force behind technological change (Boschma et al. 2013). However, few scholars have focused on the scientific relatedness (Boschma et al. 2014). There is little empirical evidence on the impact of scientific relatedness on the innovation output, especially at the country level.

The goals of this article are to compare the evolution processes of scientific relatedness in the USA and China and explore the impact of relatedness on scientific output in the period 2000–2013. Specifically, we thoroughly analyze the relatedness between knowledge and knowledge, and after that we investigate the relatedness between knowledge and country. Then, this paper assesses the extent to which the scientific output of knowledge topic is dependent on its relatedness. In order to measure the main indices, we use keywords to discuss the co-occurrence relationships among knowledge topics and explore the dynamic evolution of scientific topics. Evolutionary patterns of relatedness are also compared between the USA and China. Analyzing the relatedness at the country level aims to examine contrasts between the scientific knowledge in the USA and China. More

specifically, we can compare the role how the relatedness affect the output in the two countries.

The rest of the paper is organized as follows. “[Literature review](#)” second section reviews related literature on relatedness and innovation theories which lead to our hypotheses. Afterwards, in Sect. “[Data and context](#)”, we describe our data and compare the knowledge topics in the USA and China through constructing co-occurrence networks. Section “[Relatedness between topic and topic](#)” analyzes the relatedness between topic and topic and explore the dynamic evolution of relatedness. Section “[Relatedness between topic and country](#)” investigates the relatedness between topic and country. Section “[Analyses: relatedness and output](#)” empirically tests our hypotheses and explains the results. While Sect. “[Conclusion](#)” presents the conclusion and discussion.

## Literature review

The production of innovation usually demands knowledge from multiple scientific fields (Schumpeter 1934). Attentions on the relationship between knowledge and innovation output raises a question: What kind of influence will the relatedness of knowledge exert on the innovation output. The relatedness between the topic and other topics could bring about combinatorial opportunities (Wang et al. 2014). However, if the combinatorial capacity of the knowledge topic with other topics is too strong, the potential of their creative convergence could be largely exhausted (Kim and Kogut 1996). Relative to the knowledge topic with high relatedness, the knowledge which has weak relatedness tends to have an advantage in scientific performance. First, the knowledge transfer among knowledge topics which have high relatedness is repetitive most of the time. Content transferred among such knowledge may also represent homogenous and unuseful information. While as the information transmitted among the low relatedness knowledge topics demonstrates non-redundant and useful (Hansen 1999), which makes such knowledge topic exposed to varied topics (Granovetter 1973). Different perspectives and approaches can extend innovative expertise (Coser 1975) which is crucial to innovation. Second, due to the low similarity of knowledge topics with low relatedness (Granovetter 1983), knowledge exchange opportunities are provided to produce ideas collision and innovative spark (Jerome 2013), which is helpful to stimulate innovation in different scientific backgrounds and areas. Existing studies suggest the technology portfolio of actors should not be too related (Granstrand et al. 1997; Hussinger 2010). Diversification of related knowledge can be helpful for researchers to understand emerging technological opportunities and keep up with rapid technological developments (Hussinger 2010). What’s more, knowledge topic with weak relatedness suffers less constrains, which presents strong autonomy. Relatively weak relatedness will bring fresh knowledge, making innovation easier. Thus, a knowledge topic owning low relatedness with other knowledge topics tends to be strong in innovation output.

At the country level, relatedness between topic and country measures the closeness between knowledge topic and country. We also think that relatedness between topic and country will have a negative impact on scientific output, which is due to two reasons. On the one hand, although knowledge production shows path dependence (Garud and Karnoe 2013), such path dependence limits the learning of new knowledge. Relatively high relatedness between topic and country indicates that the topic is cognitively close to the country and the percentage of knowledge topics associated with the specific topic is high in that country (Boschma et al. 2014). As such, knowledge topic in that country may reach a

point where its scientific value has been exhausted (Carnabuci and Bruggeman 2009) suggesting the costs may ultimately exceed the benefits (Phelps et al. 2012). The exploration space of a topic which has high relatedness with the country is limited and under restrictions. We can predict that further output on that topic in that country is no longer fruitful. On the other hand, it is an inevitable tendency that knowledge production shows globalization (Hassan and Haddawy 2013). This international collaboration will promote the process of international communication and knowledge dissemination (Ma et al. 2014) which narrow the cognitive gap among nations. For this reason, the scientific output of the topic owning low relatedness with the country will increase with great potential. Based on the above discussion and arguments, we will propose our hypotheses:

**Hypothesis:** Scientific relatedness has a negative relationship with scientific output.

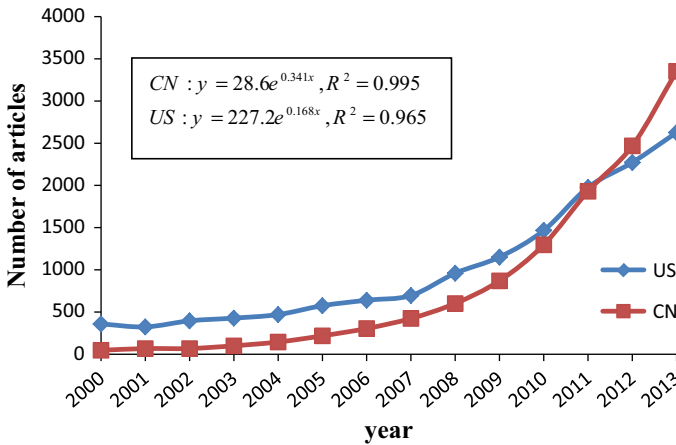
**Hypothesis (a):** The relatedness between the topic and other topics negatively affects its scientific output.

**Hypothesis (b):** The relatedness between the topic and country negatively affects its scientific output.

## Data and methodology

### Data and context

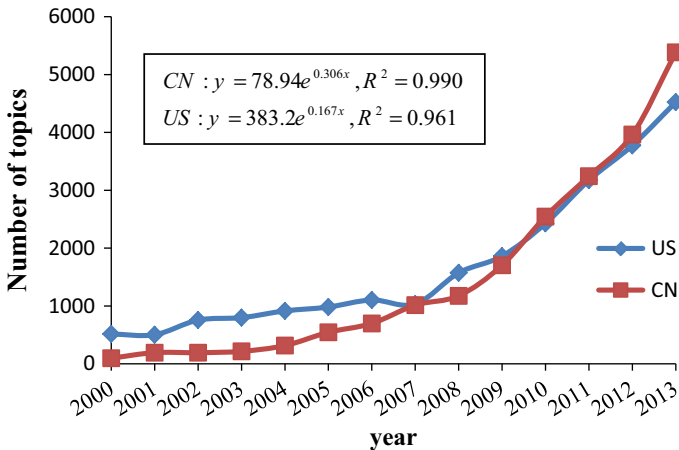
We draw on data sets of articles to analyze the USA and China research innovation networks in the period spanning 2000–2013. Though the research of solar energy has a long history, it has not been seen as a priority area of research until more recently (Hassan et al. 2014). The period 2000–2013 involving relatively new information and plentiful empirical data can be used to capture recent research trends and identify the current top collaborative authors. What's more, we found that China has only 31 papers, yet the USA has 333 papers in this field in 1999. On the one hand, few scientific documents may lead to sparse networks. On the other hand, this big gap which is found between the United States and China before 2000 may result in incomparability. Mainly due to that China shows significant increase in scientific documents and the gap with the United States is narrowed after 2000, our study interval is 2000–2013. Two scientific data sets downloaded from the Web of Science (WoS) form the core of our study. The database has been utilized extensively in early research (Zhu and Guan 2013). Searching terms which are adopted in this paper to retrieve solar energy data records, are developed by Sanz-Casado et al. (2013), and “Appendix” displays the retrieval profile for solar energy. Many research fields utilize bibliometric methods to analyze the research trend from publication output, keywords, subject category, etc. (Mallik and Mandal 2013). The total number of articles identified in the WoS is 70,562 over the period 2000–2013. The sample includes 156 countries. The database is mainly centered in the United States, which is in number one position responsible for 14,342 publications. China, in the second position, is associated with 11,896 articles published. Figure 1 plots on the number of the USA and China articles since 2000. Both of countries show significant growths in research productivity. As clearly displayed by Fig. 1, the USA occupied an absolutely leading position and China took the second largest scientific producer before 2011. Glancing at the figure, the exponential growths in solar energy articles in both countries stand out. China has a relatively



**Fig. 1** Growth in solar energy scientific articles

high scientific productivity in solar energy (CN:  $y = 28.6e^{0.341x}$ ,  $R^2 = .995$ ) and has exceeded the USA after 2011 (US:  $y = 227.2e^{0.168x}$ ,  $R^2 = .965$ ).

A large proportion of articles for each of the countries have more than one topic. The convergence of knowledge, contributes to the growth in research output for the USA, China and the world. The keywords are related to the topic (Cataldi et al. 2010) in the set of solar energy publications. Keywords summarize the core content of literature which can reflect the theme distribution and characteristics in a particular field (Glänzel and Thijs 2012). The use of keywords in the set of solar energy publications provides us with an indication of the cognitive developments of topics within the field (Rip and Courtial 1984). Keywords can be used to map knowledge structure (Su and Lee 2010), which are useful and important for the readers and authors. Accurate and concise keywords can arouse the reader’s interest and attention, which are good for the citations of the article, so almost every author writes the keywords seriously and accurately. We also considered the classification of WoS to construct the co-occurrence network. But the less number of classifications ( $N = 200$  here) urges us to give up the idea. We used data from WoS to track the country’s scientific activities. The dataset contains details about articles, including published dates, countries, and keywords. Keywords by WoS have been considered valid proxies to trace scientific development. Scientific researchers categorize keywords carefully which are usually anchored on a subject matter. Changes in a country’s keywords over time can be clearly discerned which makes the internal cognitive structure visible (Mutschke and Haase 2001). In this study, we use a key word as proxy for a topic. We clean the dataset for the reason that different words may represent the same concept, such as “solar cells” and “solar cell”. Figure 2 shows the dynamic processes of topic numbers for the USA and China. As displayed in Fig. 2 on the whole, the number of topics of the USA per year is far more than China before 2010. Nevertheless, the gap between the USA and China has remarkably narrowed year by year. Both of the USA and China perform the impressively exponential growth pattern (US:  $y = 383.2e^{0.167x}$ ,  $R^2 = .961$ ; CN:  $y = 78.94e^{0.306x}$ ,  $R^2 = .990$ ). However, the USA retains growth in a relatively stable situation and has lost its leading position after 2010 which is occupied by China. China demonstrates a great speed and strength of the catch-up process. Note that the growth in

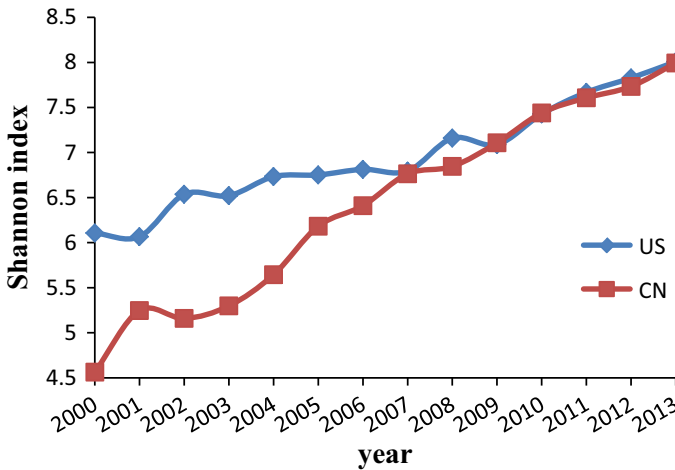


**Fig. 2** Growth of topics in solar energy field

topic numbers in the USA is exceeded by growth in China. The USA used to be the major scientific producer, but in the topic quantity in the solar energy field, China dominates.

Article output presented in Fig. 1 does not exhibit the full feature of academic status. As a comprehensive index of richness and uniformity of data, diversification can describe the distribution of the topic share (Rafols and Meyer 2010). From a RBV (resource-based view) perspective, the countries engaging in topic diversification are willing to exploit existing scientific resources to develop new capabilities which enhance innovation (Quintana-García and Benavides-Velasco 2008). Diversified countries can manage more opportunity for knowledge utilization, and their profit will increase due to topics combination arising from diversification. Given these perspectives, we seek to probe into the topic diversification of the two countries. In our study, diversification is measured by the entropy measure developed by Palepu (1985) which is calculated from articles data at the focal year by using the following equation:  $D = \sum_k^N P_k \ln\left(\frac{1}{P_k}\right)$ . Where  $P_k$  is the proportion of articles in the topic  $k$ , and  $\ln\left(\frac{1}{P_k}\right)$  is the weight for each topic  $k$ , for a country with  $N$  different topics. In order to present the dynamic change feature of topic diversification in the solar energy field, we calculate the time series data of topic share in the period 2000–2013 for the USA and China. Figure 3 portrays topic diversification of two countries, reflecting the national dynamic change. The figure reveals that the USA always has a higher diversification than China; however, the topic diversification gap is relatively smaller after 2007. After then, China closes the gap at the rapid rate.

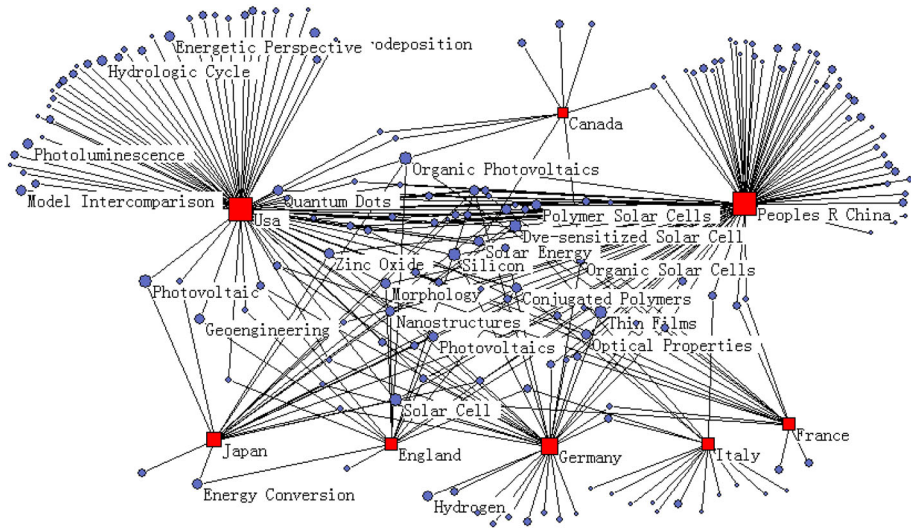
Current focal points of academic research in solar energy field are different cross nations. Social network analysis like co-word occurrence analysis, co-authorship analysis, and other indicators of scientific discovery and knowledge production, has been a useful tool for bibliometric analyses (Kumar and Jan 2014). In order to observe the differences and similar among countries, Fig. 4 visualizes crucial connections in the country-topic association network (2-mode network) in solar energy field in 2013. The visualization identifies different nodes which include country (square node) and its associated topics (circular nodes). Topic nodes whose degrees exceed 3 and edges whose weights exceed 2



**Fig. 3** Dynamic change of topic diversification

are showed in Fig. 4, and some major countries (G7countries: The United States, Britain, France, Germany, Japan, Italy, and Canada) and China are included in the figure. We just show 24 topic labels which are most frequently on articles (i.e. Solar Cells, Nanostructures, Organic Photovoltaics, Morphology .....). Based on network structure, the more massive the academic productivity of a node in the network, the larger that vertice is visualized. The USA is the largest country node in the network, China is ranks second, and Germany is third. We can find that these countries are most extensively associated to academic topics in their respective scientific activities, which demonstrates that they are the three most comprehensive of the literature producers in this network. We can identify closely linked major topics of a country by the size of nodes in each country’s sub-network. In the USA sub-network, for example, the “Solar Cells” node is largest, with the extensive association to the USA node. Furthermore, we find that Nanowires, Organic Photovoltaics, ZnO, Density Functional Theory, Saudi Arabia, Plasma Expansion, and Spacecraft Arcing, are all vastly connected to the USA. Another high density sub-network that we can identify in Fig. 4 is that China and associated topics. Dye-sensitized Solar Cells, Polymer Solar Cells, Solar Cells, Counter Electrode, Optical Properties, TiO<sub>2</sub>, Solar Energy, and Self-assembly are all highly associated with China, reflecting that they have received increasing attention in current research in China. Besides helping to identify country-associated major topics, in the center of the network (Fig. 4), one may find some topics connecting with many countries, suggesting that they attract common concerns in many countries. Take “Solar Cells”, for example, which not only connects to the USA, but also to Germany, Japan, England, and France, revealing an interrelationship between these countries. However, this is also true of some quite specific topics, which are associated with specific countries. Take Organic Dyes (whose label is not shown in Fig. 4) for example, it only connect to China, revealing a phenomenon that some specific topics acquire analyses from a single nation.

In this study, we used keywords as important information to construct co-occurrence networks of keywords. Like patents, the keyword provides a means to organize scholarly literature intellectually (Leydesdorff et al. 2014). Figure 5 represents the largest connected component of topic co-occurrences graphically in 2013 in the USA. Nodes represent topic and lines represent keywords co-occurrence. Edges are presented in this figure if and only



**Fig. 4** 2-mode network

if their weights exceed 2, which means the number of the co-occurrence between two topics exceed 2. Topic nodes whose degrees exceed 2 are showed in Fig. 5, which means those topic nodes connect with at least three other topic nodes. The size of the nodes corresponds to the number of articles in each topic. Figure 5 shows the co-occurrence maps of the keywords analyzed in the USA. As mentioned above, the maps show the important associations between topics. We further detect scientific sub-networks using Girvan–Newman algorithm in Fig. 5, and the color of each topic indicates the subgroup to which it belongs. Girvan–Newman algorithm is a method of community identification developed by Newman and Girvan (Newman and Girvan 2004) which uses centrality indices and approaches an iterative process to find community boundaries. Figure 5 displays it clearly that four subgroups detected in the co-occurrence network in the USA. Subgroup I (red) which is mainly involved in the solar power literature includes the biggest node: solar cells. Subgroup II (black) located at the top right of Fig. 5 is mainly associated with the chemical synthesis. Subgroup III (yellow) has relatively low article production reflecting the convergence between solar energy and other renewable energy. Subgroup IV (blue) named “efficiency and environment” is that with the dense associations among its different topics.

As one observes from Fig. 6, the topics are classified into 3 subgroups in China. Subgroup I (red) is in the center of the whole network which is related to solar power literature, and it has the maximum number of topics. Subgroup II (black) corresponding to “chemical synthesis” is on the left side of Fig. 6. Subgroup III (blue) named “other energy” has few relationships with other subgroups. As indicated by Figs. 5 and 6, China has different subgroups classifications comparing with the USA. Nevertheless, their research priorities have a lot in common, such as “Dye-sensitized Solar Cells”, “Polymer Solar Cells”, “Optical Properties”, etc. The possible reasons why the subgroups classifications are different between the USA and China are as follows: on the one hand, the USA pays more attention to efficiency of energy utilization (Howarth et al. 2000) and the impact on the environment (Chu and Majumdar 2012) than China does. On the other hand, China relies on extensive method of economic growth (Fei et al. 2011). Significant investment-



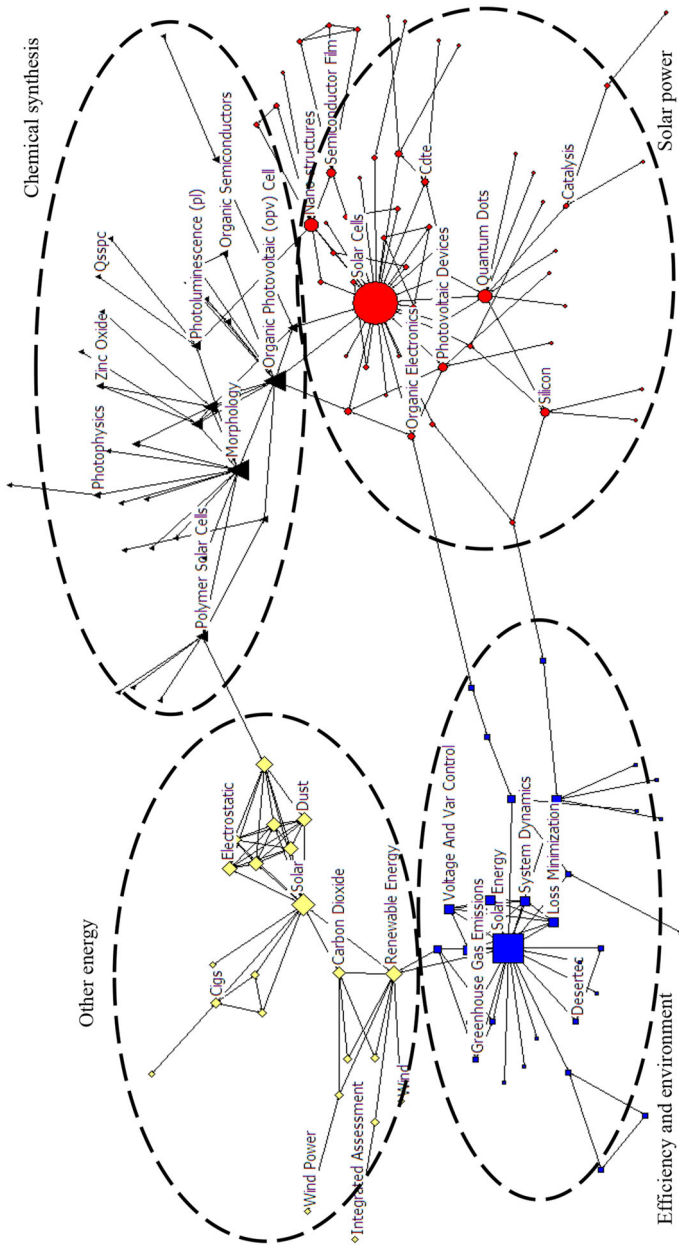


Fig. 5 Topic co-occurrence network in the USA in 2013

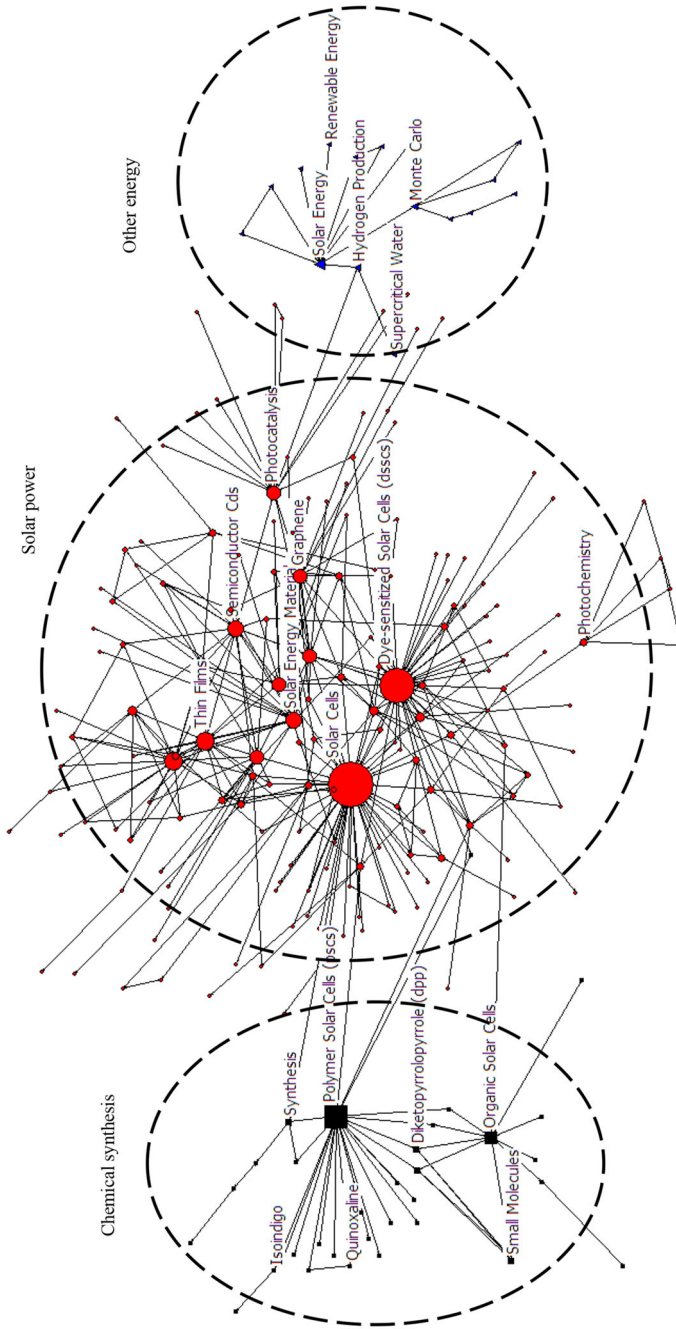


Fig. 6 Topic co-occurrence network in China in 2013

**Table 1** Top 10 high betweenness centrality topics

US	Centrality	China	Centrality
<i>Solar Cells</i>	51.36	<i>Solar Cells</i>	36.74
<i>Solar Energy</i>	31.11	Dye-sensitized Solar Cells (dsscs)	34.24
Organic Photovoltaic (opv) Cell	28.15	Photocatalysis	22.03
<i>Polymer Solar Cells</i>	25.55	<i>Polymer Solar Cells</i>	18.34
Morphology	24.64	Tio <sub>2</sub>	16.11
Efficiency	22.34	Hydrogen Production	13.16
Solar	21.75	<i>Solar Energy</i>	11.73
Renewable Energy	19.55	Graphene	9.77
Water Splitting	14.58	Thin Films	9.12
Nanowires	13.46	Semiconductor Cds	8.85

driven and factor-driven growth is a specific feature of China’s economic development (Sabir and Sabir 2010). Thus, in the USA, “energy efficiency and environment” prove more developed, while in China, “solar power” shows more central.

According to the basic theory of social networks, topic nodes with high betweenness centrality are considered to represent keys of research directions (Chi and Young 2013). Topics with relatively high betweenness centrality play important roles to bridge and link with other academic topics. They are always at the hub of a network and critical for the formation of some sub-networks (Valente 2012). We rank topics in Figs. 5 and 6 according to their betweenness centrality which is multiplied by 100. Top 10 topics in the USA and China are listed in Table 1. As reported in Table 1, the highest topic is “Solar Cells” in both of countries, fully demonstrating the importance of it. The remaining topics rankings vary in two countries, but we find that 3 topics (italics and bold) are both ranked in the top 10 in the USA and China after comparing.

Relatedness between topic and topic

To measure the relatedness between topic and topic, we use Ron Bosch’s methods and compute relatedness based on topic co-occurrence analysis (Boschma et al. 2014). A simple and normalized measure is utilized to analyze the relatedness of scientific topics. Two keywords are considered to be associated with each other if they co-occur on one article. Thus, we could obtain a co-occurrence matrix in solar energy field in each period. After then, we use Jaccard index to normalize the value represents the co-occurrences (Leydesdorff 2008). According to previous studies, we define  $R_{ijt}$  as the relatedness between topic  $i$  and topic  $j$  at time  $t$  which is given in the Eq. (1).

$$R_{ijt} = \frac{occ_{ij}}{occ_i + occ_j - occ_{ij}} \tag{1}$$

We define  $occ_{ij}$  as the co-occurrences numbers between topic  $i$  and topic  $j$ . Besides,  $occ_i$  represents the total number of occurrences of topic  $i$  which can be calculated through summing up the number of  $i$ th row of the co-occurrences matrix. Similar with  $occ_i$ ,  $occ_j$  denotes the amount of occurrences of  $j$ . We also construct  $R_{it}$  as the relatedness of topic  $i$  at time  $t$  which is displayed in the following equation:

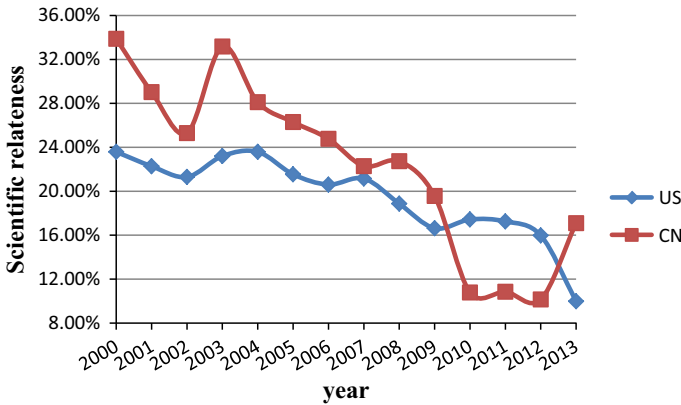


Fig. 7 Dynamic change of scientific relatedness

$$R_{it} = \frac{\sum_{i \neq j}^n R_{ijt}}{n} \tag{2}$$

where  $n$  denotes the number of topics having co-occurrence relationships with  $i$ . We further compute the scientific relatedness of the whole network at time  $t$  on the above basis. The relatedness of network  $R_t$  function given by:

$$R_t = \frac{\sum_i^m R_{it}}{m} = \frac{\sum_i^m \frac{\sum_{i \neq j}^n R_{ijt}}{n}}{m} \tag{3}$$

where  $m$  denotes the number of topics belonging to the whole network. Early studies have investigated co-occurrences of technology classes (Luan et al. 2013), co-occurrences of title words (Milojević et al. 2011). However, little scientific work analyzes the evolution of the scientific relatedness cross the time. In order to fill the gap, we plot the dynamic change of  $R_t$  over the time period of 2000–2013. As showed in Fig. 7, both the USA and China have displayed a decrease in relatedness. Overall, even the numbers of articles and topics involved in the solar energy have been rapidly increasing over time, the average scientific relatedness between topics have dramatically declined year by year. Due to emerging topics and development of science, this evolving curve represents that topic associates with a wide range of topics. However, the associations are weaker over years.

### Relatedness between topic and country

To further investigate how close a topic is to country, we construct a country-topic level variable  $R_{ict}$  through combining the relatedness  $R_{ijt}$  with the countries. Information about countries is acquired from the author addresses concluded in articles. In our study, articles are fully distributed to each author nation (Boschma et al. 2014).  $R_{ict}$  which describes the relatedness between topic  $i$  and country  $c$  at time  $t$  is given by formula (4):

$$R_{ict} = \frac{\sum_{j \in c, j \neq i} \varphi_{ijt}}{\sum_{j \neq i} \varphi_{ijt}} \tag{4}$$

where  $\sum_{j \in c, j \neq i} \varphi_{ijt}$  means the number of topics that can be found in the dataset of the country  $c$  which topic  $i$  links to at time  $t$ .  $\sum_{j \neq i} \varphi_{ijt}$  represents the total number of topics

**Table 2** Scientific relatedness between countries and topics

US			China		
Topic	Articles	$R_{ict}$	Topic	Articles	$R_{ict}$
<i>Solar Cells</i>	113	18.89	<i>Dye-sensitized Solar Cells</i>	208	44.39
<i>Photovoltaics</i>	91	39.28	<i>Solar Cells</i>	182	26.27
<i>Solar Energy</i>	67	28.36	<i>Polymer Solar Cells</i>	79	55.12
<i>Thin Films</i>	41	18.65	Tio2	57	43.24
Organic Photovoltaics	35	38.54	<i>Solar Energy</i>	52	19.58
<i>Dye-sensitized Solar Cells</i>	31	7.65	<i>Photocatalysis</i>	48	48.53
<i>Conjugated Polymers</i>	30	22.34	<i>Conjugated Polymers</i>	46	32.34
Morphology	28	32.54	<i>Thin Films</i>	45	22.47
<i>Polymer Solar Cells</i>	26	20.70	<i>Organic Solar Cells</i>	45	30.92
<i>Organic Solar Cells</i>	26	20.83	Optical Properties	44	34.53

which topic  $i$  links to at time  $t$ . For instance, if a specific topic  $i$  is related to 50 other topics in the whole dataset, and country occupies 10 topics, then the  $R_{ict}$  will be equal to  $10/50 \times 100 \% = 20 \%$ .  $R_{ict}$  can be seen as the percentage of topics found in a country, and as such, we could attribute topics to country in each year.

Table 2 shows scientific relatedness between the USA and China and top 10 topics appeared which most frequently on articles in 2013 ( $R_{ict}$  is multiplied by 100). For instance, the degree of scientific relatedness between the country-the USA and the topic “Solar Cells” is 18.89, which implies that researchers from the USA are concentrated on 18.89 % of the topics that are related to “Solar Cells”. Comparing with China, we find that 8 topics (italics and bold) are both ranked in the top 10 in the USA and China and there are some significant differences between the two countries. For example, the scientific relatedness between the country- the USA and the topic “Dye-sensitized Solar Cells” is 7.65, whereas the value is 44.39 in China representing that the researchers from China are specialized in more topics related to “Dye-sensitized Solar Cells” than the USA.

**Analyses: relatedness and output**

We collect our sample by downloading all solar energy articles from WoS which contain details about publish dates, authors, and locations to construct topic co-occurrence matrices as discussed above. In order to achieve a high precision in analyzing networks and get a fine-grained analysis of relatedness dynamics (Gulati et al. 2012), we use one-year windows from 2000 to 2013 to construct co-occurrence networks. The total dataset finally includes 156 countries and 64,658 scientific topics. The sample involved in the USA includes 23,933 topics and there are 21,284 topics in China dataset. Our goal is to empirically investigate the impact of relatedness on innovation output. Our analysis unit is topic, and we regress the number of articles containing this topic in the next year on its scientific relatedness.

The dependent variable  $Output_{t+1}$  is the number of articles containing this topic in the given year ( $t + 1$ ). The explanatory variable  $R_{it}$  means the average scientific relatedness of topic  $i$  with other topics in year  $t$ . Another important explanatory variable is  $R_{ict}$  which shows the relatedness of topic  $i$  with country  $c$  in year  $t$ . We control some factors which can

**Table 3** Means, standard deviations, and correlations (US)

	Mean	SD	Output	Stock	Structure holes	Centrality	$R_{it}$
Output	.64	2.55					
Stock	1.40	2.08	.83**				
Structure holes	1.44	.24	.30**	.33**			
Centrality	.001	.005	.66**	.73**	.25**		
$R_{it}$	.19	.12	-.08**	-.08**	-.76**	-.04**	
$R_{ict}$	.79	.31	-.33**	-.23**	-.25**	-.18**	.10**

\*  $p < .05$ ; \*\*  $p < .01$

influence the dependent variable in our analysis. Productivity of topic is controlled by a key variable: Stock which is constructed by the total number of articles in which a topic occurs on a given 1 year period. To control for the impact of network structures on the innovation output, we include variables: centrality and structural holes. The centrality of the topic in the co-occurrence network is measured by Pajek software. Due to that betweenness centrality can measure the ability of topic controlling the resource; we use betweenness centrality as the centrality measure (Kumar and Jan 2013):

$$C_B(n_i) = \frac{\sum_{j < k} g_{jk}(n_i) / g_{jk}}{(g-1)(g-2)} \quad (5)$$

where  $g_{jk}$  is the amount of the shortest paths from node  $j$  to  $k$ ,  $g_{jk}(n_i)$  shows the number of paths that pass  $i$  and  $g$  is the number of nodes in the network.

We compute structural holes using Burt's measure (Burt 1992). The index CO displays the extent to which all of  $i$ 's ego network is invested in its associations with  $j$ . The CO index is computed as following equation:

$$CO_i = \sum_j \left( p_{ij} + \sum_{q, q \neq i, q \neq j} p_{iq} p_{qj} \right)^2 \quad (6)$$

where  $p_{ij}$  is the ratio of  $i$ 's relationships invested in linking  $j$ . Due to the fact that the value of  $CO_i$  is perhaps bigger than 1 (Lee 2010), we subtract  $CO_i$  from 2 to obtain the structural holes  $S_i$  (Wang et al. 2014; Guan et al. 2014) which is shown in Eq. (7).

$$S_i = 2 - CO_i. \quad (7)$$

Table 3 provides some descriptive statistics and the correlation of the variables included in our analysis of the USA. We can obtain that control variables—stock, Structure holes, Centrality are all positively related to the output, which is consistent with early study.

We also analyze some descriptive statistics and the correlation of the variables included in our analysis of China. The result displayed in Table 4 is similar with Supporting the prior literature review,  $R_{it}$  is negatively related to topic's article output in both the USA and China ( $r = -.08, p < .01$ ;  $r = -.15, p < .01$ , respectively), and a negative correlation is found that  $R_{ict}$  is negatively related to its output ( $r = -.23, p < .01$ ;  $r = -.21, p < .01$ , respectively).

As showed in Tables 3 and 4, there are no obvious bivariate correlations among explanatory variables. We also find that the highest VIF (variance inflation factor) is 3.01, which suggests that multicollinearity is not a concern (O'Brien 2007). The dependent

**Table 4** Means, standard deviations, and correlations (China)

	Mean	SD	Output	Stock	Structure holes	Centrality	$R_{it}$
Output	.91	3.65					
Stock	1.52	2.62	.91**				
Structure holes	1.41	.24	.33**	.36**			
Centrality	.001	.005	.71**	.77**	.29**		
$R_{it}$	.15	.11	-.15**	-.15**	-.75**	-.07**	
$R_{ict}$	.76	.33	-.28**	-.21**	-.27**	-.18**	.10**

\*  $p < .05$ ; \*\*  $p < .01$

variable in our study is count variable and takes non-negative integer values (the number of articles). We find our dependent variables suffer from overdispersion. The mean of dependent variable is .64 and the SD is 2.55 in the USA, and the mean of the dependent variable is .91 and the SD is 3.65 in China. We use negative binomial models which can accommodate over dispersion (Cameron and Trivedi 2013). According to the results of the Hausman test ( $p < .001$ ), we decide to utilize fixed effects models which are appropriate for the data (Hausman 1978). Table 5 shows the results of the Negative Binomial model with fixed effects. Model 1 included all control variables of the USA, and model 2 adds  $R_{it}$  and  $R_{ict}$ . After that, we put all control variables of China into model 3, and model 4 adds  $R_{it}$  and  $R_{ict}$ . There has been no significant difference between the USA and China. As we expected, models 1 and 3 with control variables indicate that stock and structure holes are positively associated with article output. However, centrality is negatively associated with article output non-significantly. In models 2,  $R_{ict}$  has a significant negative impact on the innovation output ( $\beta = -.903, p < .01$ ), and compared with that, the coefficient for  $R_{it}$  is negative and non-significant ( $\beta = -.181, p > .05$ ). In models 4, both  $R_{ict}$  and  $R_{it}$  exhibit the negative and significant influence ( $\beta = -.557, p < .01$ ;  $\beta = -3.328, p < .01$ , respectively). As such, our results support that higher scientific relatedness leads to lower scientific output.

### Conclusion

In this study, we comparatively understand the scientific relatedness of the USA and China between 2000 and 2013, taking a network approach in the solar energy field based on the panel data from WoS database. Both differences and similarities are found in the scientific research of the two countries. The results of our analysis generate the following findings.

While the literature productivity of the USA is the highest before 2011, China has shown an incremental increase and exceeds the USA in 2011, becoming a new major scientific contributor to the solar energy field. Both of countries show an exponential growth process in the numbers of knowledge topics. The USA demonstrates an obvious low growth rate in topic numbers, while China displays a noticeable rise so as to have developed into the largest scientific topic producer since 2010. Although the topic diversification of China is always lower than the USA, the gap between the USA and China is smaller over the recent years.

**Table 5** Results of regression models

Model	1 US	2	3 China	4
Stock	.027** (.003)	.027** (.003)	.028** (.002)	.020** (.002)
Structure holes	1.063** (.726)	1.138** (.091)	1.463** (.068)	.604** (.093)
Centrality	-.172 (1.699)	.949 (1.534)	-1.610 (1.699)	-.168 (1.541)
$R_{ict}$		-.903** (.071)		-.557** (.076)
$R_{it}$		-.181 (.237)		-3.328** (.230)
Cons	-.985** (.124)	-.616** (.180)	-1.341** (.118)	1.040* (.185)
Log likelihood	-6,552.927	-6,467.972	-5,978.780	-5,827.576
Wald chi2	584.48	785.10	1,198.92	1,526.30

\*  $p < .05$ ; \*\*  $p < .01$

We apply social network method to extract country information and keywords from the database to visualize a 2-mode network, helping us to identify country-associated major topics. Word co-occurrences network is representative of research priorities which is widely used in capturing key knowledge topics. Comparative different structures of word co-occurrences networks are detected in the USA and China. Both of countries show the important associations with “solar energy”, “chemical synthesis”, “combining with other energy”. However, the USA displays the concerns on the “Efficiency and environment”.

Moreover, we construct keywords co-occurrences networks to analyze the relatedness between knowledge topic and topic, and then we make use of relatedness between knowledge topic and topic to develop the variable: relatedness between knowledge topic and country. As such, there are two kinds of scientific relatedness involved in this paper: (1) relatedness between topic and topic; (2) relatedness between topic and country. We also investigate the dynamic change of scientific relatedness over the time. A dramatically declined trend of scientific relatedness is found. We then investigate how close a topic is to country through analyzing the relatedness between topic and country. We obtain that 8 same topics are both ranked in the top 10 in the USA and China.

Furthermore, previous studies have used patent and literature documents to the influence of scientific relatedness on emergence and disappearance of topics (Boschma et al. 2014; Rigby 2013) and investigated the relationship between the knowledge relatedness and innovation outcome at the firm level (Tanriverdi and Venkatraman 2005). Our study investigates the impact of scientific relatedness on the innovation output in the solar energy field during the year 2000–2013. As expected, we find empirical evidence that scientific relatedness tend to negatively affect the scientific performance both in the USA and China. Our study contributes to the innovation and knowledge research.

The theory contribution of our research as follows. First, we contribute to the literature of scientific relatedness at the topic level (i.e. relatedness between topic and topic) and the country level (i.e. relatedness between topic and country). Secondly, we study on what extent the scientific relatedness exerts influence on the literature productivity at the country level. Finally, we make a better understanding of the research directions in the USA and China. Our work has potential implications for the future innovative research in the solar energy field.

Our work may help policymakers and researchers about the further scientific development at the country level. This is an urgent issue in solar energy sciences due to importance of energy development. Our results show that researchers should take into



account the relatedness of the topics. At country level, the scientific relatedness between the topic and the country negatively affect the innovation performance in the next period. Thus, nation government should provide a policy direction and pay attention to support the topics which have low scientific relatedness with this country. These topics are cognitively far to the country, and they have great potential to be developed in this country. In addition, our results suggest that each country should clear directions for research and focus on the research topics which have low scientific relatedness instead of imitating other nations. Researchers should focus on the scientific topics which owning low relatedness in order to yield more scientific output.

The current study has several strengths. We differentiated two kinds of scientific relatedness and operate them in different ways. We empirically tested the relationship between scientific relatedness and innovation performance which fills the gaps in the previous studies. There are also some limitations in our study. We should note that our empirical setting is solar energy field; future studies should collect data from wider fields to check the consistency of results. In addition, we just investigate the relationship between scientific relatedness and the number of articles; future studies could probe the deep mechanism between them and the origin of scientific relatedness.

**Acknowledgments** This study is supported by a grant from National Natural Science Foundation of China (No. 71373254). The authors are very grateful for the valuable comments and suggestions from the anonymous reviewers and Editor-in-Chief of the journal, which significantly improved the quality and readability of the paper.

## Appendix: Retrieval profiles for solar energy

TS = (“solar energy\*” OR “solar radiation” OR “solar cell\*” OR “solar photovoltaic\*” OR “solar power” OR “solar heat\*” OR “solar plant\*” OR “solar concentrate\*” OR “solar thermal” OR “solar collect\*” OR “solar technolog\*”) AND PY = (2000–2013) Refined by: Document Type = (ARTICLE) AND [excluding] Web of Science Categories = (HORTICULTURE OR PLANT SCIENCES OR FORESTRY) Databases = SCIEXPANDED,SSCI,CPCI-S,CPCI-SSH Timespan = 2000–2013 Lemmatization = On.

## References

- Boschma, R., Balland, P.-A., & Kogler, D. F. (2013). Relatedness and technological change in cities: The rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010: Paper Presented at the Annual Meeting of the Association of American Geographers, Los Angeles.
- Boschma, R., Heimeriks, G., & Balland, P.-A. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy*, 43(1), 107–114.
- Bradsher, K. (2010). On clean energy, China skirts rules. *New York Times*, 9, A1.
- Burt, R. S. (1992). The social structure of competition. *Networks and organizations: Structure, form, and action*. Boston: Harvard Business School Press.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data*. Cambridge: Cambridge University Press.
- Carnabuci, G., & Bruggeman, J. (2009). Knowledge specialization, knowledge brokerage and the uneven growth of technology domains. *Social Forces*, 88(2), 607–641.
- Cataldi, M., Di Caro, L., & Schifanella, C. (2010). *Emerging topic detection on twitter based on temporal and social terms evaluation*. Proceedings of the Tenth International Workshop on Multimedia Data Mining, July 25, 2010. Washington, D.C., USA, pp. 1–10.
- Chi, R., & Young, J. (2013). The interdisciplinary structure of research on intercultural relations: A co-citation network analysis study. *Scientometrics*, 96(1), 147–171.
- Chu, S., & Majumdar, A. (2012). Opportunities and challenges for a sustainable energy future. *Nature*, 488(7411), 294–303.

- Coser, R. L. (1975). The complexity of roles as a seedbed of individual autonomy. *The idea of social structure: Papers in honor of Robert K. Merton*, pp. 237–263.
- Fei, L., Dong, S., Xue, L., Liang, Q., & Yang, W. (2011). Energy consumption-economic growth relationship and carbon dioxide emissions in China. *Energy Policy*, 39(2), 568–574.
- 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.
- Garud, R., & Karnoe, P. (2013). *Path dependence and creation*. New York: Psychology Press.
- Geum, Y., Kim, C., Lee, S., & Kim, M.-S. (2012). Technological convergence of IT and BT: Evidence from patent analysis. *ETRI Journal*, 34(3), 439–449.
- Glänzel, W., & Thijs, B. (2012). Using ‘core documents’ for detecting and labelling new emerging topics. *Scientometrics*, 91(2), 399–416.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological theory*, 1(1), 201–233.
- Granstrand, O., Patel, P., & Pavitt, K. (1997). Multi-technology corporations: Why they have ‘distributed’ rather than ‘distinctive core’ competences. *California Management Review*, 39(4), 8.
- Guan, J., & Ma, N. (2007). China’s emerging presence in nanoscience and nanotechnology: A comparative bibliometric study of several nanoscience ‘giants’. *Research Policy*, 36(6), 880–886.
- Guan, J., Yan, Y., & Zhang, J. (2014). How do collaborative features affect scientific output? Evidences from wind power field. *Scientometrics*, doi:10.1007/s11192-014-1311-x.
- Gulati, R., Sytch, M., & Tatarynowicz, A. (2012). The rise and fall of small worlds: Exploring the dynamics of social structure. *Organization Science*, 23(2), 449–471.
- Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, 44(1), 82–111.
- Hassan, S.-U., & Haddawy, P. (2013). Measuring international knowledge flows and scholarly impact of scientific research. *Scientometrics*, 94(1), 163–179.
- Hassan, S.-U., Haddawy, P., & Zhu, J. (2014). A bibliometric study of the world’s research activity in sustainable development and its sub-areas using scientific literature. *Scientometrics*, 99(2), 549–579.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 46(6), 1251–1271.
- Howarth, R. B., Haddad, B. M., & Paton, B. (2000). The economics of energy efficiency: Insights from voluntary participation programs. *Energy Policy*, 28(6), 477–486.
- Hussinger, K. (2010). On the importance of technological relatedness: SMEs versus large acquisition targets. *Technovation*, 30(1), 57–64.
- Jerome, L. W. (2013). Innovation in social networks: Knowledge spillover is not enough. *Knowledge Management Research & Practice*, 11(4), 422–431.
- Joo, S. H., & Kim, Y. (2010). Measuring relatedness between technological fields. *Scientometrics*, 83(2), 435–454.
- Kim, D.-J., & Kogut, B. (1996). Technological platforms and diversification. *Organization Science*, 7(3), 283–301.
- Kodama, M. (2005). Knowledge creation through networked strategic communities: Case studies on new product development in Japanese companies. *Long Range Planning*, 38(1), 27–49.
- Kumar, S., & Jan, J. (2013). Mapping research collaborations in the business and management field in Malaysia, 1980–2010. *Scientometrics*, 97(3), 491–517. doi:10.1007/s11192-013-0994-8.
- Kumar, S., & Jan, J. (2014). Research collaboration networks of two OIC nations: Comparative study between Turkey and Malaysia in the field of ‘Energy Fuels’, 2009–2011. *Scientometrics*, 98(1), 387–414. doi:10.1007/s11192-013-1059-8.
- Lee, J. J. (2010). Heterogeneity, brokerage, and innovative performance: Endogenous formation of collaborative inventor networks. *Organization Science*, 21(4), 804–822.
- Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton’s Cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology*, 59(1), 77–85.
- Leydesdorff, L., Kushnir, D., & Rafols, I. (2014). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*, 98(3), 1583–1599. doi:10.1007/s11192-012-0923-2.
- Li, Z.-S., Zhang, G.-Q., Li, D.-M., Zhou, J., Li, L.-J., & Li, L.-X. (2007). Application and development of solar energy in building industry and its prospects in China. *Energy Policy*, 35(8), 4121–4127.
- Livingstone, D. N. (2010). *Putting science in its place: Geographies of scientific knowledge*. Chicago: University of Chicago Press.

- Luan, C., Liu, Z., & Wang, X. (2013). Divergence and convergence: Technology-relatedness evolution in solar energy industry. *Scientometrics*, 97(2), 461–475.
- Ma, F.-C., Lyu, P.-H., Yao, Q., Yao, L., & Zhang, S.-J. (2014). Publication trends and knowledge maps of global translational medicine research. *Scientometrics*, 98(1), 221–246.
- Makri, M., Hitt, M. A., & Lane, P. J. (2010). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. *Strategic Management Journal*, 31(6), 602–628.
- Mallik, A., & Mandal, N. (2013). Bibliometric analysis of global publication output and collaboration structure study in microRNA research. *Scientometrics*, 98(3), 2011–2037.
- Matthiessen, C. W., & Schwarz, A. W. (2010). World cities of scientific knowledge: Systems, networks and potential dynamics. An analysis based on bibliometric indicators. *Urban Studies*, 47(9), 1879–1897.
- Milojević, S., Sugimoto, C. R., Yan, E., & Ding, Y. (2011). The cognitive structure of library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology*, 62(10), 1933–1953.
- Mutschke, P., & Haase, A. Q. (2001). Collaboration and cognitive structures in social science research fields. Towards socio-cognitive analysis in information systems. *Scientometrics*, 52(3), 487–502.
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690.
- Palepu, K. (1985). Diversification strategy, profit performance and the entropy measure. *Strategic Management Journal*, 6(3), 239–255.
- Pelphs, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks a review and research agenda. *Journal of Management*, 38(4), 1115–1166.
- Quintana-García, C., & Benavides-Velasco, C. A. (2008). Innovative competence, exploration and exploitation: The influence of technological diversification. *Research Policy*, 37(3), 492–507.
- Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2), 263–287.
- Rigby, D. L. (2013). Technological relatedness and knowledge space: Entry and exit of US cities from patent classes. *Regional Studies* (ahead-of-print), pp. 1–16.
- Rip, A., & Courtial, J.-P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6(6), 381–400.
- Sabir, R. I., & Sabir, R. M. (2010). Managing technological innovation: China's strategy and challenges. *Journal of Technology Management in China*, 5(3), 213–226.
- Sanz-Casado, E., García-Zorita, J. C., Serrano-López, A. E., Larsen, B., & Ingwersen, P. (2013). Renewable energy research 1995–2009: A case study of wind power research in EU, Spain. *Germany and Denmark. Scientometrics*, 95(1), 197–224.
- Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle* (Vol. 55). New Jersey: Transaction Publishers.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1), 65–94.
- Su, H.-N., & Lee, P.-C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight. *Scientometrics*, 85(1), 65–79.
- Tanriverdi, H., & Venkatraman, N. (2005). Knowledge relatedness and the performance of multibusiness firms. *Strategic Management Journal*, 26(2), 97–119.
- Valente, T. W. (2012). Network interventions. *Science*, 337(6090), 49–53.
- Virmani, A., (2005). *A tripolar century: USA, China and India*. Working Paper, No. 160. New Delhi: Indian Council for Research on International Economic Relations.
- Wang, C., Rodan, S., Fruin, M., & Xu, X. (2014). Knowledge networks, collaboration networks, and exploratory innovation. *Academy of Management Journal*, 57(2), 484–514.
- Zhang, G., Xie, S., & Ho, Y.-S. (2010). A bibliometric analysis of world volatile organic compounds research trends. *Scientometrics*, 83(2), 477–492.
- Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, 35(1), 83–104.
- Zhu, W., & Guan, J. (2013). A bibliometric study of service innovation research: Based on complex network analysis. *Scientometrics*, 94(3), 1195–1216.