

Scientific relatedness and intellectual base: a citation analysis of un-cited and highly-cited papers in the solar energy field

JingJing Zhang¹ · Jiancheng Guan¹

Received: 21 May 2016 / Published online: 11 October 2016
© Akadémiai Kiadó, Budapest, Hungary 2016

Abstract In the solar energy field, scientists publish numerous scientific articles every year. Some are highly-cited, while others may not even be cited. In this paper, we introduce two underlying scientific properties of a paper to explain this paper's highly-cited or un-cited probability: scientific relatedness and intellectual base. We utilize two main network techniques, knowledge element coupling network (concurrency-based) and paper citation network (citation-based) analyses, to measure scientific relatedness and intellectual base, respectively. What's more, we conduct descriptive analyses of un-cited and highly-cited papers at the country, organization and journal levels. Then we map knowledge element co-occurrence networks and paper citation networks to compare the network characteristics of un-cited and highly-cited papers. Further, we use article data in the solar energy field between 2004 and 2010 to examine our hypotheses. Findings from Ordered Logit Models indicate that when the scientific relatedness of a paper is high, this paper is more likely to be un-cited, whereas less likely to be highly-cited. The paper with higher intellectual base has a higher possibility to be highly-cited, whereas a low possibility to be un-cited. Overall, this paper provides important insights into the determinant factors of a paper's citation levels, which is helpful for researchers maximizing the scientific impact of their efforts.

Keywords Scientific relatedness · Intellectual base · Citation analysis · Solar energy

Introduction

Citation is a widely well-known and accepted indicator of an article's scientific impact (Schubert and Glänzel 1983). Scientific articles serve to disseminate their findings through inviting other researchers to cite their articles (McClellan 2003). Thus, citations indicate

✉ Jiancheng Guan
guanjianch@ucas.ac.cn

¹ School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, People's Republic of China

that articles make use of the resource, ideas and results of the cited articles, and the citation is considered as the extent to which the cited paper is useful to other studies. Specifically, in any field, we can observe that a handful of articles are frequently and highly-cited, and a large number of articles are cited a few times. Many articles may not even be cited at all (Redner 1998). Therefore, a growing number of researchers are looking for factors that affect the citation impact of a paper (Ahmed et al. 2004; Oppenheim and Renn 1978; Burrell 2003).

Scholars suggested that the possibility of being un-cited or highly-cited depends on many influencing factors. Knowledge elements or topics reflect scientific content and domain to some extent (Lee and Jeong 2008). Citing references or patterns provide intellectual bases for the current paper, which can reflect scientific base of the paper (Takeda and Kajikawa 2008). The characteristics of scientific content and scientific base of a paper (i.e., scientific relatedness and intellectual base in this study) are related to the knowledge production (Hammarfelt 2010). Based on the previous perspectives, we introduce two underlying scientific properties of a paper to explain the paper's highly or un-cited probability: scientific relatedness and intellectual base.

On the one hand, we propose the citation of a scientific article is strongly affected by its scientific relatedness. In this paper, we use keywords in the publication as the indication of knowledge elements, for keywords sum up the core content and represent the element distribution (Zhang et al. 2015; Muñoz-Leiva et al. 2012). Scientists usually scan the keywords of a paper to identify whether this paper is related to their work rather than finishing the whole abstract or article (HaCohen-Kerner 2003). The basic assumption of scientific citation is that, the utilizations of scientific articles, as they are represented by citations to some specific documents, are indications of scientific impact (Garfield 1979; Persson 1994). Citations can be seen as scholarly influence and value (Merton 1979). Thus, scientific relatedness between a paper's knowledge elements and other topics is related to the paper's scientific impact. However, previous work investigating the influence of knowledge elements on a paper's citation has mostly focused on the surface features of its elements, such as the number of keywords (Akhavan et al. 2016). To our knowledge, no research tried to analyze the effect of scientific relatedness on the citation impact. The aim of this paper is trying to fill the above gap by examining the effect of scientific relatedness at the paper level. We utilize knowledge element coupling network to calculate the scientific relatedness between a focal element and other elements. Coupling analysis, using knowledge element co-occurrence network, offers a promising tool for investigating knowledge structures and development patterns. The basic idea of knowledge element coupling is to connect knowledge elements according to their co-occurrence in a common article. The scientific relatedness of a knowledge element indicates its closeness and associations with other knowledge elements. The higher the scientific relatedness, the more frequently repeated co-occurrence relationships exist between two knowledge elements. Since the analysis unit of this study is a paper, which may contain several knowledge elements, we average the scientific relatedness values of all of that paper's knowledge elements as the paper's scientific relatedness value.

On the other hand, we argue that the intellectual base, which a paper derives from citing prior articles' perceptions, is also a significant determinant of its citation. Reference patterns of an article shed important insights into its knowledge domain's intellectual base. Intellectual base of a paper is its most fundamental characteristic. The basic assumption of scientific citation is that, the utilizations of scientific articles, as they are represented by citations to some specific documents, are indications of scientific subject matter (Garfield 1979; Persson 1994). Citations can be seen as scholarly influence and value (Merton 1979).

The citations of cited articles serve the useful function of transferring knowledge, and are considered as the symbol of rewarding researchers by approving their intellectual property rights (Merton 1988). We construct the paper citation network and use the quality, status or prestige of the paper's references in the paper citation networks as the indication of intellectual base. Paper citation network analysis, constructing citing-cited matrixes in which the columns denote citing articles and the rows represent cited articles, is a powerful quantitative tool for mapping the intellectual base. Paper citation network analysis uses documents as the analysis units, and represents the knowledge flow. Based on prior research, the cited documents constitute the intellectual base of the citing articles (Persson 1994; Chen 2006). To be specific, we utilize paper citation network to calculate the out-degree centrality value of all references of a paper. Since the analysis unit of this study is a paper, which may contain several references, we average the centrality values of all of that paper's references as its intellectual base value. The higher intellectual base is, the higher status of the references a paper has.

To sum up, our main objective is to investigate the influence of scientific relatedness and intellectual base on the article citation levels in the solar energy field. There are two main reasons why we choose such a field as our empirical context. On one hand, solar energy plays a very crucial role in solving our environmental problems and providing energy need in the world (Lewis and Nocera 2006). It has been a priority field of scientific research in recent years (Zhang et al. 2015). On the other hand, publication and citation practices depended on the features of the field that articles belong to (Price 1970; Narin 1976). Thus, choosing a single field as our research setting can allow us to control for the potential confounding influences of field-level characteristics. By analyzing solar energy article data from Web of Science from 2004 to 2010, we try to identify top countries/regions, organizations and journals with higher ratios of un-cited or highly-cited papers. Further, we map and compare the paper citation and knowledge element co-occurrence networks of un-cited or highly-cited papers. Finally, we utilize knowledge element co-occurrence network to calculate scientific relatedness and paper citation network to measure intellectual base of an article. We believe our findings will provide some value into the influence factors of scientific articles citations, and offer valuable guidance for researchers shaping their future research directions.

Theory and hypotheses

This part can be divided into two parts: scientific relatedness and paper citation; intellectual base and paper citation. We argue that a paper with a lower scientific relatedness (or a higher intellectual base) has higher quality, consequently it is more likely to be highly-cited, whereas less likely to be un-cited. Conversely, paper with a higher intellectual base has higher quality, consequently it is more likely to be highly-cited, whereas less likely to be un-cited. The hypotheses are explained in the following subsections.

Scientific relatedness and paper citation

Knowledge elements with a high scientific relatedness have frequent repeated co-occurrence relationships with other elements, while knowledge elements with a low scientific relatedness indicate that they have weak and occasional interactions with others (Guan and Liu 2015). We propose that papers with a high scientific relatedness are less likely to be

highly-cited but more likely to be un-cited. There are several reasons can explain this. First, strong links among knowledge elements transfer a great deal of redundant information (Karsai et al. 2014; Guan and Yan 2016). Thus, a paper with high scientific relatedness tends to rely heavily on repeated interactions with previous knowledge, and thus have cognitive lock-in problems (Simard and West 2006). The paper may just have access to redundant knowledge information, resulting in the stifling of innovation (Uzzi 1997) and fewer citations. Secondly, priority and exclusiveness are key rewarding rules in innovation activities (Hagstrom 1965). To be specific, scientific reward is commonly given exclusively to knowledge workers who outdo others in generating new knowledge (Wang et al. 2014). As such, from the perspective of originality of scientific papers, when a paper has a high level of scientific relatedness, knowledge elements in this paper had frequently co-occurred with some specific knowledge elements previously. In this case, because of the rewarding principle of “first-to-file priority” (Friedman et al. 1991), the combinatorial potentials between these knowledge elements may have been seriously exhausted. Innovation generally emerges by combination or recombination of knowledge elements (Weitzman 1998; Asheim and Coenen 2005; Guan and Yan 2016). Papers have high scientific relatedness, indicating that the combinatorial opportunities of their knowledge elements are exhausted, which likely results in un-citation of these papers. Third, a paper with a high scientific relatedness experiences more constraints inhibiting it from exploring new ideas, for its knowledge elements are easily suffered from knowledge inertia (Cheon et al. 2015). Specially, search is generally local and performed around the associative knowledge elements. When a paper has high scientific relatedness, knowledge elements in this paper have frequently repeated co-occurrence relationships. The paper is likely to be locked in redundant knowledge information and vulnerable to the knowledge inertial tendencies of local and associative search. Besides, the uncertainty, cost and risk of innovation will further enhance the knowledge inertial behaviors, consequently hindering the novelty of the paper. Novelty is a potentially important indicator of article quality (Tahamtan et al. 2016). Articles with higher quality are easier to be highly-cited, whereas unoriginal low-quality articles is more likely to be un-cited (Bornmann et al. 2012). We therefore contend that

Hypothesis 1: A paper with a lower scientific relatedness has higher quality, consequently it is more likely to be highly-cited, whereas less likely to be un-cited.

Intellectual base and paper citation

Intellectual base refers to the cited articles, which are commonly considered as composing the research front (Hammarfelt 2010; Persson 1994). When researchers do their studies and write their articles, they will read, understand and consequently cite some previous work as their theory and knowledge base. For example, if article A cites article B, some scientific knowledge of B will be integrated into article A. As such, we can say an article’s citing texts represent its intellectual base (Persson 1994). Scholars argued that intellectual base is associated with knowledge creation and citation pattern (Hammarfelt 2010). The paper with high value of intellectual base is more likely to be highly-cited, whereas less likely to be un-cited for the following reasons. One the one hand, by citing previous work, current articles establish connections between their concepts, theory and ideas (Small 1981) and these connections represent the current articles’ cognitive structure and the intellectual background of the articles involved. Cited papers can exert their intellectual and cognitive influences on citing papers. The aggregation of cited references of the papers can be

deemed as representing the theoretical and empirical supports or resources for building authors' argument. A paper with a high level of intellectual base indicates that it is built on the information of the most influential articles, and it develops its scientific argument with a solid and well-developed theoretical base in a promising direction or domain. When scholars do systemic research on a topic, they will find that this paper well integrate previous work and is highly related to key literature in this field. Thus, this paper cannot be neglected by scholars and is likely to be highly cited. Contrarily, a paper with a low level of intellectual base may focus on a domain which researchers don't pay much attention to, without a sound and good theoretical base. On the other hand, some research has pointed out that researchers can see further when they are standing on giants' shoulders. Put differently, highly-cited scientific papers are strongly based on prior highly-cited works (Bornmann et al. 2010). Based on previous successful research, current articles can obtain needed knowledge and information easily and efficiently, avoid detours and mistakes, and provide an in-depth view. Thus, there is a high possibility for them to be highly-cited, whereas a low possibility for them to be un-cited.

Hypothesis 2: A paper with a higher intellectual base has higher quality, consequently it is more likely to be highly-cited, whereas less likely to be un-cited.

Data and methodology

Data and context

To complete our research goals, we analyzed a sample of solar energy articles during the year 2004–2010 in Web of Science database (WoS). These articles were acquired by searching articles containing solar energy related keywords. These keywords were obtained from prior bibliographic studies (Sanz-Casado et al. 2013). Specifically, the retrieval profile of solar energy field is as follows: 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*”). There are total 44,684 unique solar energy articles obtained from this time period. Our analysis sample ends in the year 2010, although we collected article information until the end of the year 2015. The main reason why we stop sampling in the year 2010 is that a paper's citations are often low in the early years after its publication. Consequently, we may not have sufficient time to observe the citations of the articles that published after 2010, because articles take average 5 years to be steadily cited and we should take the time-lag effect of citations into account (Li et al. 2013). For each article, we extracted its data details, including the title, abstract, published year, keywords, references, journal, authors, organizations of authors, funding information and WoS category, citations. Missing data was deleted, and keywords that had same meanings, but had different expressions were merged into one word.

We constructed a dataset for our regression analysis, in which one observation was one article in one specific year. Following earlier research (Wang 2016), we measured the independent variables—scientific relatedness and intellectual base between the year $t - 4$ to t , and measured the dependent variable—citations of the focal article published in the year t . For instance, when the dependent variable is the citations of a paper published in 2008, we considered all the articles during 2004–2008 to construct the knowledge element co-occurrence network and paper citation network. Using these networks, we computed the

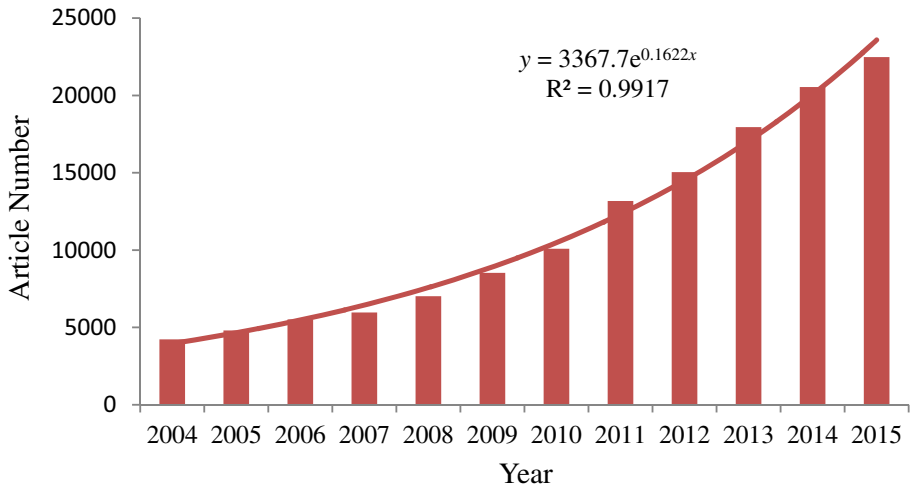


Fig. 1 Year distribution of papers in the solar energy

scientific relatedness of every knowledge element and the intellectual base of this paper. Finally, our regression sample contained 24,697 observations in total. The distribution of solar energy articles by year is displayed in Fig. 1. The fitting curve of the number of solar energy articles represents the exponential growth pattern. The solar energy scientific research grows slowly in the preliminary phase and it grows rapidly in recent years.

Measure

Dependent variable

Citation level As we aim to assess the influence factors of un-cited and highly-cited papers in the field of solar energy, our focus is on three levels of citations (1: un-cited; 2: normal cited; 3: highly-cited). As such, the dependent variable in our study is categorical and we use an order logit approach. It is relatively simple when we identify un-cited papers in the year t . The target article is an un-cited publication if it has received zero citation until now. For every paper published in the year t , we determine it is a highly-cited publication if its citation is over the top ten percentile point of citations of all the publications in the year t (Bornmann et al. 2011; Lewison et al. 2007; Tijssen et al. 2002). For example, if the top ten percentile value for citations of all article published in year 2010 is 60, the articles published in 2010 with more than 60 citations belong to the highly-cited articles. Other publications that are cited but less than top ten percentile value for all papers' citations are identified as normal cited papers.

Figure 2 plots the distribution of the citation frequencies of the solar-energy articles during 2004–2010. This figure clearly shows that the distribution of articles' citation frequencies is fitted by a power law. There is a relatively small number of articles that can occupy absolute scientific impact advantages, and majority of articles in the solar energy field receive fairly low citations. Further, we find that the un-cited papers always accounts for around 4.5 % of total solar-energy papers in a specific year.

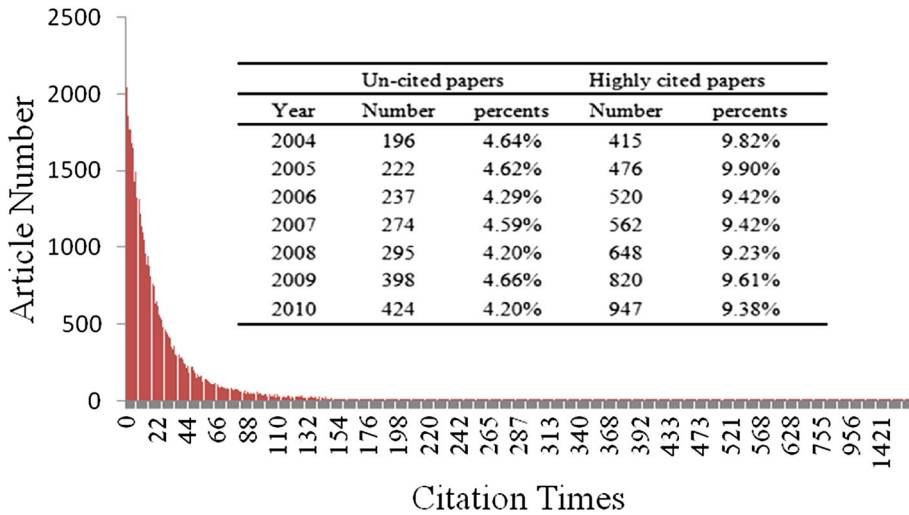


Fig. 2 Citation distribution of papers in the solar energy (2004–2010)

Independent variables

We carry out social network analysis to investigate the knowledge element co-occurrence network and paper citation network. The study of knowledge element co-occurrence network enables us to measure scientific relatedness, and the analysis of paper citation network enables us to obtain the intellectual base of the paper.

Scientific relatedness To arrive at the scientific relatedness measure, we first constructed knowledge element co-occurrence network in year $t - 4$ to t . Such networks document the relatedness of knowledge elements via publications’ keywords, where nodes are knowledge elements and a link shows the fact that two knowledge elements have been included in one article together. Knowledge element co-occurrence networks are thus undirected and weighted networks, and the weight of the link between nodes i and j means the number of papers that contain both i and j .

After that, we utilized the knowledge element co-occurrence networks to acquire co-occurrence matrixes. Following prior studies, we used Jaccard Index (Leydesdorff 2008; Real and Vargas 1996) to normalize the element values of knowledge element co-occurrence matrixes. We calculate a scientific relatedness value for every pair of knowledge elements using the formula (1) (Leydesdorff 2008; Zhang et al. 2015). To be specific, we computed the scientific relatedness R_{ij} between knowledge elements i and j through normalizing the co-occurrence times between them:

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

where occ_{ij} is the co-occurrence times of knowledge element i and element j . occ_i means the total number of occurrence times of knowledge element i , which is the sum of values in the i th row of knowledge element co-occurrence matrixes. occ_j means the total number of occurrence times of knowledge element j , which is the sum of values in the j th column of

co-occurrence matrixes. As such, scientific relatedness indicates how knowledge elements are closely associated with one another.

The next step was to average the dyadic-level scientific relatedness to an aggregate scientific relatedness measure R_i for knowledge element i :

$$R_i = \frac{\sum_{i \neq j}^n R_{ij}}{n}, \quad (2)$$

where n means the total number of knowledge elements that have co-occurrence relationships with knowledge element i . Thirdly, since this study's observation is an article, which usually includes several knowledge elements, we identified the focal paper's knowledge portfolio that contains its all knowledge elements. The average of the scientific relatedness values of the knowledge elements in the knowledge portfolio is the paper's value of scientific relatedness.

$$\text{Scientific relatedness}_{\text{paper}} = \frac{\sum_{j=1}^m R_j}{m}, \quad (3)$$

where m represents the number of knowledge elements in the focal paper's knowledge portfolio.

Intellectual base The calculation of intellectual base of a paper has three main steps. First, using the citation relations between papers, we constructed an article citation network in year $t - 4$ to t . Citation networks consist of directed and un-weighted graphs having the research papers as vertices. To be specific, there is an edge from the vertex i to the vertex j if the paper i is cited by the paper j . New papers can only cite the previous papers, and thus the directed edges in citation networks follow chronological orders. Citation networks give us insight into the knowledge transfer processes, and the directions of edges indicate that knowledge moves from paper i to paper j (Kazi et al. 2016).

Second, we employed the network topological analysis to compute the network characteristic (e.g., out-degree centrality in this paper) of citation networks (Li et al. 2007). In a directed network, nodes have both an in-degree and out-degree centrality. The in-degree for a given node is calculated as its total number of ingoing edges, and the out-degree is calculated as its total number of outgoing edges (Koschützki and Schreiber 2008). Because links in citation networks represent knowledge diffusion, out-degree of the paper means the total number of papers that has received knowledge from it. Thus, out-degree can reflect the scientific impact of an article. We calculate the out-degree centrality of paper i through normalizing the out-degree:

$$\text{Centrality}_i = \frac{k}{n}, \quad (4)$$

where k mean the out-degree of paper i , and n represents the total number of nodes in the graph. Third, bibliometric research depicts intellectual base as the cited articles and the core scientific documents of a field (Hammarfelt 2010). Based on these, we identified the focal paper's reference portfolio that consists of all scientific articles cited by the focal paper. The average of the out-degree centrality values of the scientific articles in the focal paper's reference portfolio is its value of intellectual base.

$$\text{Intellectual base}_{\text{paper}} = \frac{\sum_{j=1}^m \text{Centrality}_j}{m}, \quad (5)$$

where m represents the number of publications in the focal paper's reference portfolio. Because this normalized measure is relatively small, we multiply the value of "Intellectual base" by 10,000 in our estimations.

Control variables

In statistical examinations of two hypotheses proposed, seven paper-level characteristics are controlled, which include *authors*, *knowledge elements*, *abstract length*, *organizations*, *funding*, *category* and *impact factor*. Variable *authors* is measured as the total number of authors who co-write the focal paper. Variable *knowledge elements* is the total number of the knowledge elements that are included in the focal paper. Variable *abstract length* refers to the written style of the focal article's abstract. Some scholars suggested that the way in which an article's abstract is written may influence the citation number of this paper. We counted the number of sentences, which are identified using a space: "." (Letchford et al. 2016). Variable *Organizations* is the total number of organizations that all authors belong to. Variable *Funding* is a dummy variable, which takes a value of one if the focal paper is supported by some funding, and takes a value of zero if the focal paper is not supported by any funding. We used WoS categories as a representation of the scientific subject fields, and we measured the variable *category* as the total number of categories that the focal paper belongs to. Some research argued that multidisciplinary research may be more highly-cited, thus we controlled *category* of the target paper in our analysis. *Impact factor* refers to the average 2-year impact factor of a journal, which is calculated as the number of citations that papers published in this journal in the two prior years, divided by the amount of articles in these 2 years. Although not without impact factor's critics (Seglen 1997), it is widely considered as an important index of the quality of a journal (Falagas et al. 2008). Thus, we controlled the impact factor of the journal where the focal article was published. This data can be directly obtained from JCR of ISI WoS.

Model

The dependent variable is the citation level of each paper, which is treated as an ordinal level measure (1: un-cited; 2: normal cited; 3: highly-cited). We aim to assess the impact of the scientific relatedness and intellectual base on the citation level of a paper. Because of this ordinal nature of our dependent variable, we utilize an ordered logit model, an appropriate model choice when the dependent variable is categorical (Amemiya 1981). Ordered logit model is one of the classes of models, which is considered as a qualitative choice model.

Results

Descriptive analyses of un-cited or highly-cited papers at different levels

Table 1 shows the top 15 countries/regions with higher ratios of un-cited or highly-cited over the period 2004–2010. We arranged Table 1 in two parts, with the left panel reporting the top 15 countries/regions with higher percentages of un-cited papers, and the right panel

Table 1 Countries/regions ranking of un-cited and highly-cited papers (>200 papers)

Un-cited papers			Highly-cited papers		
Countries/regions	Un-cited (total)	Percent	Countries/regions	Highly-cited (total)	Percent
Ukraine	46 (224)	20.54	Singapore	103 (352)	29.26
Russia	235 (1194)	19.68	Denmark	82 (290)	28.28
Poland	74 (442)	16.74	USA	2960 (11,136)	26.58
Egypt	33 (262)	12.60	Switzerland	264 (1007)	26.22
India	215 (2037)	10.55	Sweden	181 (725)	24.97
Mexico	44 (438)	10.05	Netherlands	225 (932)	24.14
Brazil	59 (621)	9.50	Taiwan (China)	234 (1157)	20.22
Argentina	23 (305)	7.54	Austria	87 (441)	19.73
Turkey	45 (631)	7.13	Belgium	102 (539)	18.92
Africa	14 (201)	6.97	England	416 (2229)	18.66
China	326 (5020)	6.49	Korea	331 (1879)	17.62
Japan	233 (3771)	6.18	Italy	275 (1599)	17.20
Norway	17 (281)	6.05	France	339 (2161)	15.69
Israel	22 (390)	5.64	Canada	161 (1057)	15.23
Greece	33 (589)	5.60	Japan	571 (3771)	15.14

reporting those with highly-cited papers. To identify productive and comparable countries/regions, we compared countries/regions only if they published more than 200 articles. The fraction of un-cited articles authored by Ukraine, which is ranked highest in the percentage of un-cited papers, is about 20.54 % in the period 2004–2010. Russia ranks second, with approximately 19.68 % un-cited papers. Besides, overall 29.26 % of articles authored by Singapore are highly-cited articles, then 28.28 % for Denmark, followed by 26.58 % for the USA.

A total of 44,684 publications are authored by 46,520 different organizations. The top 15 organizations with higher ratios of un-cited or highly-cited papers over period 2004–2010 are displayed in Table 2. We just compared organizations only if they published more than 20 articles. It can be seen that, approximately 21.43 % of the total solar energy-related publications authored by “Universidad Nacional Autónoma de México” are un-cited papers. “Tallinn University of Technology”, with about 11.11 % of the papers without citations, ranks second after “Universidad Nacional Autónoma de México”. Besides, “The University of California, Los Angeles, Department of Materials Science and Engineering” has the highest ratio of highly-cited papers (75.51 % of the total number of its papers), and is followed by “University of California, Berkeley, Department of Chemistry” (68.00 % of the total number of its papers). We also noticed that no one organization occurs in both of two lists.

The publications in our sample are published in 2745 different journals. Table 3 lists the top 15 journals with the largest ratios of un-cited or highly-cited papers over period 2004–2010. We just compared journals only if they published more than 50 articles. Based on Table 3, we found that, the percentage of un-cited papers of “Optoelectronics and Advanced Materials” is highest, with about 39.22 % of the papers without citations. “Geomagnetism and Aeronomy” ranks second with about 37.70 % of the papers without citations. Further, it should be noticed that “Science” (76.81 % of the total number of its papers) outstands in the percentage of highly-cited papers, is followed by “Journal of the

Table 2 Organizations ranking of un-cited and highly-cited papers (> 20 papers)

Organizations	Un-cited papers		Highly-cited papers	
	Organizations	Percent (total)	Organizations	Percent (total)
Universidad Nacional Autónoma de México	6 (28)	21.43	The University of California, Los Angeles, Department of Materials Science and Engineering	37 (49)
Tallinn University of Technology	3 (27)	11.11	University of California, Berkeley, Department of Chemistry	17 (25)
California Institute of Technology, Pasadena	4 (42)	9.52	Stanford University, Department of Electrical Engineering	15 (24)
Ege University Solar Energy Institute	2 (21)	9.52	University of California, Berkeley, Lawrence Berkeley National Laboratory	16 (26)
NASA's Lyndon B. Johnson Space Center	2 (22)	9.09	Technical University of Denmark, Risø National Laboratory for Sustainable Energy	18 (30)
Max Planck Institute for Extraterrestrial Physics	2 (22)	9.09	Stanford University, Department of Chemical Engineering	13 (22)
University of Chinese Academy of Sciences	6 (67)	8.96	Swiss Federal Institute of Technology	20 (35)
State Key Laboratory of Fine Chemicals at Dalian University of Technology	2 (24)	8.33	Washington University, Department of Materials Science and Engineering	24 (43)
University of Stuttgart	2 (26)	7.69	Stanford University, Materials Science and Engineering	32 (58)
Swedish Institute of Space Physics	2 (26)	7.69	Georgia Institute of Technology, School of Materials Science and Engineering	12 (22)
Nanyang Technological University, School of Electrical and Electronic Engineering	4 (54)	7.41	Department of Chemistry, University of Washington	30 (57)
University of Michigan	2 (27)	7.41	Department of Electrical and Computer Engineering, University of Toronto	11 (21)
California Institute of Technology, Jet Propulsion Laboratory	8 (111)	7.21	Center for Polymers and Organic Solids, University of California, Santa Barbara	17 (35)
Observatoire de Paris	2 (30)	6.67	Risø National Laboratory, Danish Polymer Centre	12 (25)
Finnish Meteorological Institute	3 (49)	6.12	Ecole Polytechnic Fed Lausanne, Lab Photon and Interfaces	10 (21)

Table 3 Journals ranking of un-cited and highly-cited papers (> 50 papers)

Un-cited papers			Highly-cited papers		
Journals	Un-cited (total)	Percent	Journals	Highly-cited (total)	Percent
Optoelectronics and Advanced Materials	20 (51)	39.22	Science	53 (69)	76.81
Geomagnetism and Aeronomy	46 (122)	37.70	Journal of the American Chemical Society	274 (375)	73.07
Journal of Optoelectronics and Advanced Materials	20 (102)	19.61	Nature	71 (99)	71.72
Semiconductors	13 (71)	18.31	Nano Letters	158 (226)	69.91
Astrophysics and Space Science	11 (65)	16.92	Advanced Materials	164 (259)	63.32
Journal of Spacecraft and Rockets	14 (83)	16.87	Advanced Functional Materials	152 (273)	55.68
Physica Scripta	8 (51)	15.69	Angewandte Chemie International Edition	54 (97)	55.67
Acta Astronautica	20 (129)	15.50	Acs Nano	77 (145)	53.10
Advances in Space Research	58 (378)	15.34	Proceedings of the National Academy of Sciences	33 (70)	47.14
Solar System Research	9 (59)	15.25	The Journal of Physical Chemistry Letters	26 (56)	46.43
Chinese Physics Letters	9 (61)	14.75	Chemical Communications	95 (234)	40.60
Journal of the Korean Physical Society	14 (104)	13.46	Chemistry of Materials	109 (301)	36.21
Earth, Planets and Space	7 (52)	13.46	Physical Review Letters	41 (123)	33.33
Astronomy Letters	7 (57)	12.28	Macromolecules	76 (235)	32.34
European Physical Journal Applied Physics	9 (74)	12.16	The Journal of Physical Chemistry B	114 (367)	31.06

American Chemical Society” (73.07 % of the total number of its papers). Comparing the left and right panels of Table 3, we found that no one journal occurs in both of them.

Network analysis of un-cited or highly-cited papers

To visualize the citation patterns in the solar energy field, we drew paper citation networks of un-cited and highly-cited articles during 2006–2010 in Figs. 3 and 4. A node in the network shows a paper, and a direct edge implies a citation from the source reference to the target article. The red round nodes represent the cited papers, and the blue square nodes represent the citing papers. Node size is scaled to the number of citations. In Fig. 3, the bigger the red paper node, the more un-cited articles cited this paper, while in Fig. 4, the bigger the red paper node, the more highly-cited articles cited this paper. For there are too many nodes and ties in these networks, we only display the papers that are cited by un-cited papers at least 5 times in Fig. 3. In Fig. 4, we only display the papers that are cited by highly-cited papers at least 50 times. Paper citation networks of un-cited papers in the solar energy field represent a low degree of integration. Conversely, the paper citation network

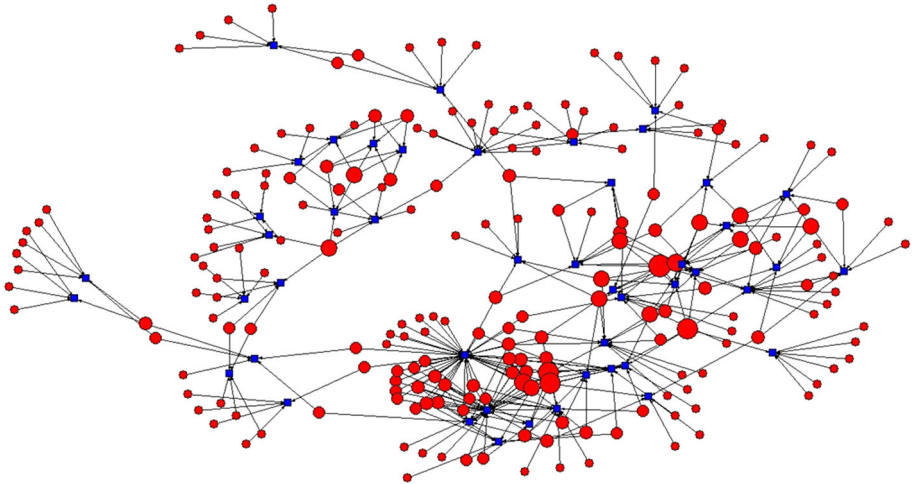


Fig. 3 Paper citation network of un-cited papers in 2006–2010

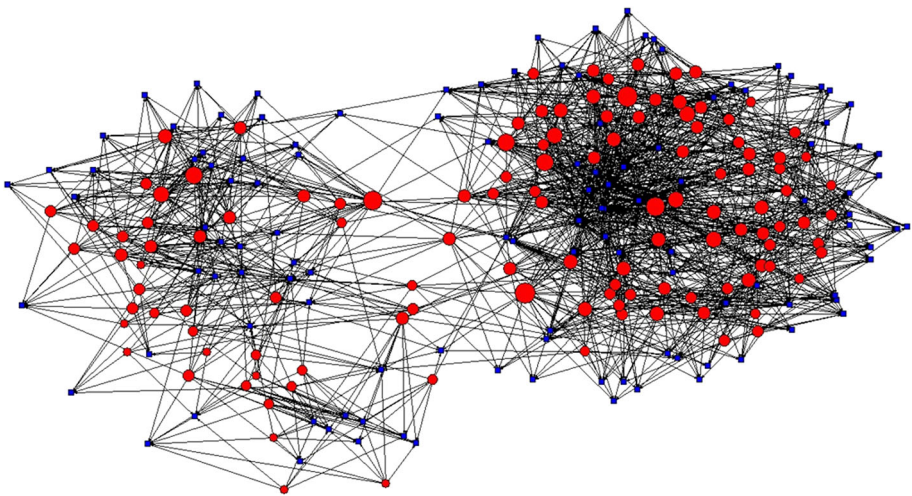


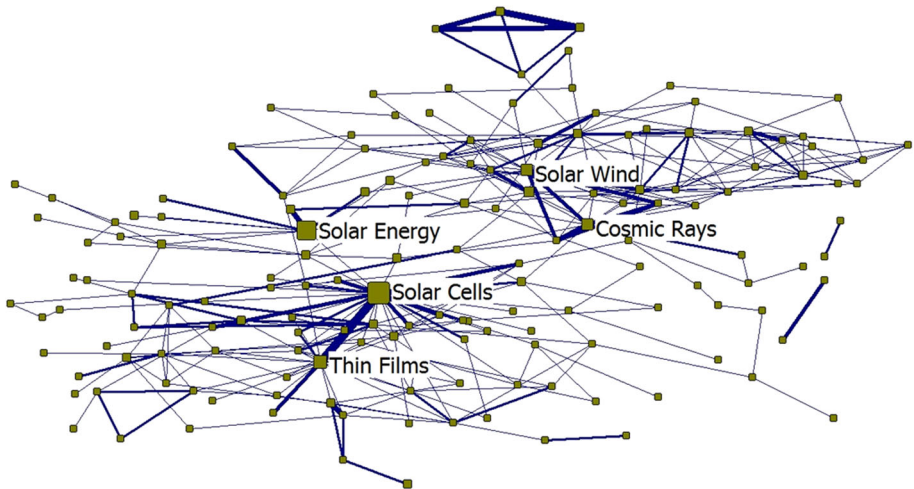
Fig. 4 Paper citation network of highly cited papers in 2006–2010

of highly-cited papers in the solar energy field represent more clustered integrated network structures.

Table 4 summarizes the network characteristics of paper citation network of un-cited papers and highly-cited papers, which includes the number of nodes, the number of ties, network density and average degree of two networks. Comparing with the paper citation network of un-cited papers, there is a big increase in the number of nodes, ties, density and average degree in the paper citation network of highly-cited papers. The top 10 papers (displayed as their DOI number) that are cited by un-cited or highly-cited papers are also listed in Table 4. The most popular papers, which are likely to be cited by un-cited papers, include “10.1038/353737a0”, “10.1021/ja00067a063”, “10.1038/35104607” and so on. The most popular papers, which are likely to be cited by highly-cited papers, seem to be

Table 4 Comparisons of paper citation networks of un-cited and highly-cited papers

	Un-cited	Highly-cited
Nodes	21,037	61,745
Ties	21,472	121,076
Density	0.00005	0.00003
Average degree	2.04135571	3.9218
Top 10 cited papers	10.1038/353737a0 (41) 10.1021/ja00067a063 (26) 10.1038/35104607 (15) 10.1126/science.1069156 (12) 10.1021/cr050149z (11) 10.1023/a:1022428818870 (10) 10.1126/science.270.5243.1789 (9) 10.1557/jmr.2004.0252 (8) 10.1021/cm049654n (8) 10.1021/cr00033a003 (8)	10.1038/353737a0 (501) 10.1038/nmat1500 (343) 10.1002/adfm.200500211 (320) 10.1126/science.270.5243.1789 (267) 10.1038/35104607 (218) 10.1021/ja00067a063 (201) 10.1038/nmat1387 (186) 10.1002/1616-3028(200102)11:1<15::aid-adfm15>3.0.co (186) 10.1038/nmat1928 (179) 10.1002/adma.200501717 (172)

**Fig. 5** Knowledge element co-occurrence network of un-cited papers in 2006–2010

“[10.1038/353737a0](#)”, “[10.1038/nmat1500](#)”, “[10.1002/adfm.200500211](#)” and so on. The value in the parentheses next to the paper indicates the number of total citations received from un-cited or highly-cited papers. For example, “[10.1038/353737a0](#)” has been cited 41 times by un-cited papers, while 501 times by highly-cited papers. We also find there are four papers (bold and italic) occur in both paper lists, and they are “[10.1038/353737a0](#)”, “[10.1021/ja00067a063](#)”, “[10.1038/35104607](#)” and “[10.1126/science.270.5243.1789](#)”.

To better understand the relatedness of knowledge elements, we depicted the knowledge element co-occurrence networks of un-cited papers (Fig. 5) and highly-cited papers

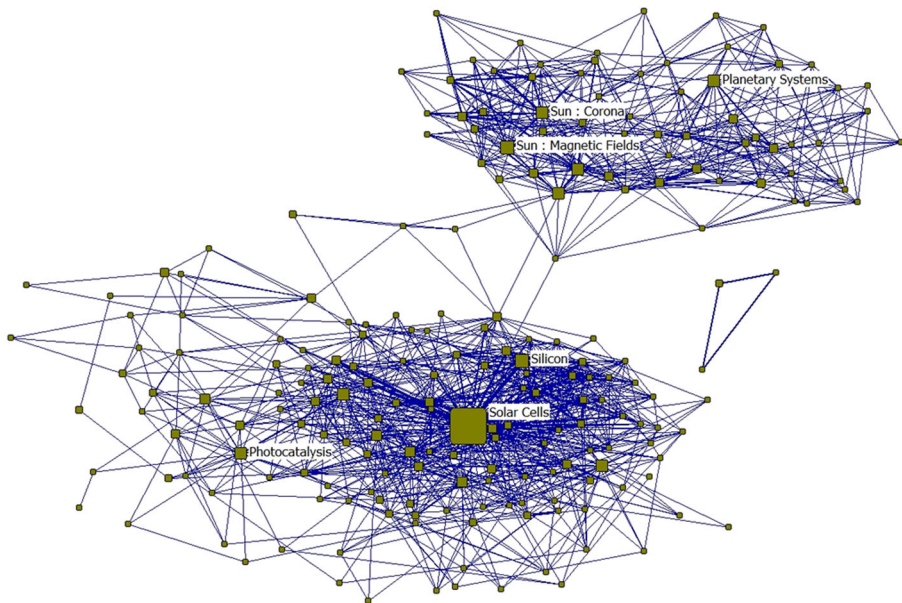


Fig. 6 Knowledge element co-occurrence network of highly-cited papers in 2006–2010

(Fig. 6) in 2006–2010. Each node means a knowledge element, i.e., article keyword. Edges represent established patterns of co-occurrence between two knowledge elements. To make figures clearer, we just reported the knowledge elements that are contained in more than three un-cited papers in Fig. 5, and displayed the knowledge elements that are contained in more than five highly-cited papers in Fig. 6. Each knowledge element is assigned a size, representing the number of papers contained it. The thickness of lines in Figs. 5 and 6 indicates the number of co-occurrence times of two knowledge elements. Figure 5 indicates that the most significant knowledge elements in un-cited papers are “Solar cells”, “Solar energy”, “Thin Films”, “Solar wind”, “Cosmic rays” and so forth. However, there is a big difference between the significant knowledge elements of un-cited papers and those of highly-cited papers. Figure 6 reveals some top productive knowledge elements, such as “Solar cells”, “Silicon”, “Photocatalysis”, “Planetary Systems”, “Sun: Corona” and so forth. Further, it is noted that knowledge element co-occurrence network of highly-cited papers in 2006–2010 is much larger and denser than that of un-cited papers.

We calculated the network characteristics of knowledge element co-occurrence network of un-cited papers and highly-cited papers in 2006–2010 to compare their different general network structure. As shown in Table 5, the values in the number of network nodes, density, clustering coefficient and average path of two knowledge element co-occurrence networks are relatively close. However, comparing with the knowledge element co-occurrence network of un-cited papers, there is a big increase in the number of tie and average degree of such network of highly-cited papers. This is indicative of the fact that knowledge elements of highly-cited papers have high combinatorial potential to connect with each other. The value in the parentheses next to the knowledge element represents the number of un-cited or highly-cited papers that contain the knowledge element. “Solar Cells” ranked first in the number of both un-cited and highly-cited papers that contain it. There are 47 un-cited papers contained the knowledge element “Solar Cells” and 151

Table 5 Comparisons of knowledge element co-occurrence networks of un-cited and highly-cited papers

	Un-cited	Highly-cited
Nodes	3383	3871
Ties	9143	16,545
Density	0.0016	0.0022
Average degree	5.41	8.55
Clustering coefficient	0.91	0.85
Average path	4.75	4.11
Top 10 knowledge elements	Solar cells (47)	Solar cells (151)
	Solar energy (36)	Silicon (42)
	Thin films (21)	Sun: magnetic fields (38)
	Solar wind (18)	Mhd (36)
	Cosmic rays (16)	Sun: corona (34)
	Sun (12)	Planetary systems (33)
	Photoluminescence (9)	Photocatalysis (33)
	Hydrogen (9)	Organic solar cells (32)
	Energy (9)	Tio ₂ (31)
	Microstructure (9)	Turbulence (31)

highly-cited papers contained it. Observing top 10 most frequent knowledge elements in un-cited or highly-cited papers, we find that only one knowledge element, i.e., “Solar Cells” (bold and italic) occurs in both knowledge element lists. It is important to note there is a relatively big difference between research topics in un-cited and highly-cited papers in solar energy. Investing in the profitable research domains is very important for the research development of solar energy, whereas researchers face huge risks and uncertainties when they choose directions. Consequently, our findings can provide some practical guidelines on the selection of solar energy research topics. We found some research topics, such as “Silicon”, “Sun: Magnetic Fields”, “Mhd”, “Planetary Systems” and so on, are more attractive and popular. While some other research topics, such as “Thin Films”, “Solar wind”, “Cosmic Rays”, “Photoluminescence” and so on, have received little attention from wind energy researchers. Combining the findings in Fig. 5, we notice that these primary research topics in un-cited papers usually have thicker lines with other research topics. Since the lines in Fig. 5 can describe the number of co-occurrence times of two knowledge elements, these research topics usually have higher scientific relatedness. Thus, this further supports our theoretical views that a paper with a higher scientific relatedness is more likely to be un-cited. Further, the results indicate that some solar-related materials have received more attentions (Du et al., 2014), such as silicon, photocatalysis, Tio₂ and organic solar panel, while some microscopic research (e.g., Cosmic Rays, Hydrogen and Microstructure) achieved fewer attentions. Thus, researchers can shift their research topics from less popular topics to highly concerned topics.

Regression analysis

Table 6 shows the descriptive statistics, variance inflation factor (VIF) and correlations of all variables in our study. All VIF values in Table 6 are <2, which is much less than the widely accepted threshold of 5.0. Thus, we can conclude that multicollinearity is not a big

Table 6 Means, standard deviations and correlations

	M	SD	VIF	1	2	3	4	5	6	7	8	9
1 Citation level	2.01	0.33	–									
2 Authors	4.58	5.80	1.94	0.06**								
3 Knowledge elements	4.77	1.78	1.00	0.06**	0.003							
4 Abstract length	11.14	5.29	1.03	0.01	0.01	–0.0003						
5 Organizations	3.17	1.97	1.98	0.08**	0.69**	–0.01	0.08**					
6 Funding	0.51	0.50	1.06	0.10**	0.07**	0.04**	–0.03**	0.10**				
7 Category	1.78	0.91	1.01	0.06**	0.01	0.01	–0.01	–0.05**	–0.04**			
8 Impact factor	2.85	2.09	1.10	0.35**	0.06**	–0.01	–0.03**	0.12**	0.21**	0.01		
9 Scientific relatedness	0.19	0.22	1.06	–0.09**	–0.03**	0.01	0.05**	–0.03**	–0.07**	0.04**	–0.19**	
10 Intellectual base	0.81	1.57	1.03	0.07**	0.04**	–0.01	–0.12**	–0.01	0.07**	0.03**	0.06**	–0.15**

* $p < 0.05$; ** $p < 0.01$

Table 7 Results of order logit regression analyses

	Model 1	Model 2	Model 3	Model 4
Authors	0.01* (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Knowledge elements	0.12** (0.02)	0.12** (0.01)	0.12** (0.02)	0.12** (0.02)
Abstract length	0.03** (0.01)	0.03** (0.01)	0.03** (0.01)	0.03** (0.01)
Organizations	0.04* (0.02)	0.04* (0.02)	0.05** (0.02)	0.05** (0.02)
Funding	0.23** (0.06)	0.22** (0.06)	0.21** (0.06)	0.21** (0.06)
Category	0.17** (0.03)	0.18** (0.03)	0.16** (0.03)	0.17** (0.03)
Impact factor	0.48** (0.01)	0.48** (0.01)	0.48** (0.01)	0.47** (0.01)
Scientific relatedness		−0.64** (0.12)		−0.54** (0.13)
Intellectual base			0.12** (0.02)	0.11** (0.02)
LR chi ²	1939	1966	1987	2006
Pseudo R ²	0.15	0.15	0.15	0.16
Year effects	Yes	Yes	Yes	Yes

* $p < 0.05$; ** $p < 0.01$

problem in our study (O'Brien 2007). The examinations for each of the hypotheses are reported in the following section.

The results of order Logit regression model for citation impact of scientific relatedness and intellectual base are presented in Table 7. Model 1 includes all control variables used in this study. Model 2 includes control variables and main effect of scientific relatedness. Model 3 includes control variables and main effect of intellectual base. Model 4 is a full model that includes all previous terms. Chi squares indicate these models can explain significant variance in papers' citation level.

Hypothesis 1 argues that when the scientific relatedness of a paper is high, this paper is more likely to be un-cited, whereas less likely to be highly-cited. In confirmation of Hypothesis 1, a negative relationship is found between the scientific relatedness and the citation level. Models 2 and 4 in Table 7 show significantly negative term for scientific relatedness ($\beta = -0.64$, $p < 0.01$; $\beta = -0.54$, $p < 0.01$; respectively). This finding is consistent no matter whether we control the other independent variable (i.e., intellectual base) in the models. Thus, Hypothesis 1 is supported. Hypothesis 2 predicts that when the intellectual base of a paper is high, this paper is more likely to be highly-cited, whereas less likely to be un-cited. As shown in Models 3 and 4 in Table 7, the independent variable, intellectual base attains significance ($\beta = 0.12$, $p < 0.01$; $\beta = 0.11$, $p < 0.01$; respectively). Again similar results are obtained from Model 3 that only include one independent variable. Results are also similar with results in Model 4, which is a full model including two independent variables. Hypothesis 2 is thus fully supported. Table 7 also shows that several control variables are significant: probability of highly-cited is improved by the number of knowledge elements, abstract length, the number of organizations authors belong to, whether the paper is supported by funding, the number of WoS categories and the impact factor of the journal.

Conclusions and discussions

The longitudinal dataset, citation analysis and keyword coupling analysis employed in this study yield several expected findings and some primary contributions. First, we performed a comparative network analysis in two groups of papers in the solar energy field. Specifically, we visualized knowledge element co-occurrence networks and paper citation networks of un-cited and highly-cited papers during 2006–2010. We detected the similarities and differences of network characteristics, research domains and research bases in un-cited and highly-cited papers. We found that knowledge co-occurrence networks and paper citation networks of highly-cited papers represent relatively higher degree of interconnections than those of un-cited papers. Besides, the most significant knowledge elements and popular cited papers of highly-cited papers show big differences. Based on that, our study contributes to the measuring of scientific relatedness and intellectual base by employing knowledge element co-occurrence network and paper citation networks. Applying network analysis tools to study scientific relatedness and intellectual base make it possible to evaluate the knowledge search efficacy in the future research. This study goes far beyond the focus of surface features of papers. The two measures may expand our future research of the role of paper knowledge elements in citations and likely opens the door to a new scope of empirical opportunities.

Second, this paper investigates scientific articles in the solar energy field from the year 2004–2010. We especially conducted analyses of un-cited and highly-cited papers in this field on the country, organizational and journal levels. Thus, our study contributes to the research about un-cited and highly-cited papers. This study highlights the influence factors of un-cited papers, which has not received sufficient attention (Yamashita and Yoshinaga 2014; MacRoberts and MacRoberts 2010). Further, this study firstly highlights the critical roles of scientific relatedness and intellectual base in paper citations. We found that a paper with a lower scientific relatedness (or a higher intellectual base) is more likely to be highly-cited, whereas less likely to be un-cited. Scientific relatedness between knowledge elements could be a liability rather than an impetus in citation levels. Consistent with prior research (Bornmann et al. 2010), we found a paper with high intellectual base can avoid detours and mistakes and provide an in-depth view, receiving more citations from others.

Third, this paper makes both theoretical and policy contributions that stem from our research findings. In the theoretical section of the paper, we propose that when the scientific relatedness of a paper is high, this paper is more likely to be un-cited, whereas less likely to be highly-cited. Our empirical results support the theoretical expectations. We further propose that intellectual base, instead, play a converse important role of scientific relatedness on citation levels. Put differently, when the intellectual base of a paper is high, this paper is more likely to be highly-cited, whereas less likely to be un-cited. Energy is a basic human need, and essential for human and social development. Countries try to concentrate on research, development and adoption of renewable energy to cope with the severe challenges in energy issues (Guan et al. 2015). Solar energy is considered as the largest source of renewable energy supply (Timilsina et al. 2012) and this field continues to increase rapidly in recent years. Therefore, scientists, knowledge workers, organizations and even countries should value on the scientific development of solar energy, and then seize innovative opportunities of solar energy science and thus achieve and maintain high-quality scientific performance. Researchers might also consider scientific directions and aim to raise the levels of article quality. With respect to the findings conducted so far, choosing a scientific direction with which papers can obtain attentions or receive citations

more easily can be better appropriate. If scientists can analyze their research topics of a paper in the knowledge element co-occurrence networks, surely they can shift their research topics into topics with low relatedness with other research topics, thereby enhancing the citations of this paper. If not, constructing their work based on prior high-quality papers might also be a good choice because it can help a paper in standing on giants' shoulders and developing its intellectual base.

Some limitations may influence our results, which should be noticed in future studies. First, although citation levels are a common proxy for scientific impact of papers, the appropriation of the usage of citation level as a proxy is not clear. Citation network analysis is premised upon an important assumption: higher quality articles are likely to be cited frequently. However, there is no certain relationship between citation levels and article quality. For instance, some articles may be highly-cited just because they are controversial. There could be other reasonable ways to measure scientific impacts in the future. Second, our empirical data comes from a single industry—solar energy. Such sampling may limit the generalizability of our findings to other industries. Further empirical validation is needed in other contexts.

Acknowledgments This study is supported by Grants from National Natural Science Foundation of China (Nos. 71373254 and 71540034).

References

- Ahmed, T., Johnson, B., Oppenheim, C., & Peck, C. (2004). Highly cited old papers and the reasons why they continue to be cited. Part II., The 1953 Watson and Crick article on the structure of DNA. *Scientometrics*, *61*(2), 147–156.
- Akhavan, P., Ebrahim, N. A., Fetradi, M. A., & Pezeshkan, A. (2016). Major trends in knowledge management research: A bibliometric study. *Scientometrics*. doi:10.1007/s11192-016-1938-x.
- Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, *19*(4), 1483–1536.
- Asheim, B. T., & Coenen, L. (2005). Knowledge bases and regional innovation systems: Comparing Nordic clusters. *Research Policy*, *34*(8), 1173–1190.
- Bornmann, L., de Moya Anegón, F., & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. *PLoS ONE*, *5*(10), e13327.
- Bornmann, L., Mutz, R., Marx, W., Schier, H., & Daniel, H. D. (2011). A multilevel modelling approach to investigating the predictive validity of editorial decisions: Do the editors of a high profile journal select manuscripts that are highly cited after publication? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(4), 857–879.
- Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics*, *6*(1), 11–18.
- Burrell, Q. L. (2003). Predicting future citation behavior. *Journal of the American Society for Information Science and Technology*, *54*(5), 372–378.
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377.
- Cheon, Y.-J., Choi, S. K., Kim, J., & Kwak, K. T. (2015). Antecedents of relational inertia and information sharing in SNS usage: The moderating role of structural autonomy. *Technological Forecasting and Social Change*, *95*, 32–47.
- Du, H., Li, N., Brown, M. A., Peng, Y., & Shuai, Y. (2014). A bibliographic analysis of recent solar energy literatures: The expansion and evolution of a research field. *Renewable Energy*, *66*, 696–706.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: Strengths and weaknesses. *The FASEB Journal*, *22*(2), 338–342.
- Friedman, D. D., Landes, W. M., & Posner, R. A. (1991). Some economics of trade secret law. *The Journal of Economic Perspectives*, *5*(1), 61–72.
- Garfield, E. (1979). *Citation indexing*. New York: Wiley.

- Guan, J. C., & Liu, N. (2015). Invention profiles and uneven growth in the field of emerging nano-energy. *Energy Policy*, 76, 146–157.
- Guan, J. C., & Yan, Y. (2016). Technological proximity and recombinative innovation in the alternative energy field. *Research Policy*. doi:10.1016/j.respol.2016.05.002.
- Guan, J. C., Zhang, J. J., & Yan, Y. (2015). The impact of multilevel networks on innovation. *Research Policy*, 44(3), 545–559.
- HaCohen-Kerner, Y. (2003). Automatic extraction of keywords from abstracts. In V. Palade, R. J. Howlett, & L. Jain (Eds.), *Knowledge-based intelligent information and engineering systems: Proceedings of the 7th international conference, KES 2003, Oxford, UK, September 2003, Part 1* (pp. 843–849). Berlin, Heidelberg: Springer.
- Hagstrom, W. O. (1965). *The scientific community*. New York: Basic books.
- Hammarfelt, B. (2010). Interdisciplinarity and the intellectual base of literature studies: Citation analysis of highly cited monographs. *Scientometrics*, 86(3), 705–725.
- Karsai, M., Perra, N., & Vespignani, A. (2014). Time varying networks and the weakness of strong ties. *Scientific Reports*. doi:10.1038/srep04001.
- Kazi, P., Patwardhan, M., & Joglekar, P. (2016). Towards a new perspective on context based citation index of research articles. *Scientometrics*, 107(1), 103–121.
- Koschützki, D., & Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regulation and Systems Biology*, 2, 193.
- Lee, B., & Jeong, Y.-I. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis. *Scientometrics*, 77(1), 3–19.
- Letchford, A., Preis, T., & Moat, H. S. (2016). The advantage of simple paper abstracts. *Journal of Informetrics*, 10(1), 1–8.
- Lewis, N. S., & Nocera, D. G. (2006). Powering the planet: Chemical challenges in solar energy utilization. *Proceedings of the National Academy of Sciences*, 103(43), 15729–15735.
- Lewis, G., Thornicroft, G., Szmukler, G., & Tansella, M. (2007). Fair assessment of the merits of psychiatric research. *The British Journal of Psychiatry*, 190(4), 314–318.
- 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.
- Li, X., Chen, H., Huang, Z., & Roco, M. C. (2007). Patent citation network in nanotechnology (1976–2004). *Journal of Nanoparticle Research*, 9(3), 337–352.
- Li, E. Y., Liao, C. H., & Yen, H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy*, 42(9), 1515–1530.
- MacRoberts, M. H., & MacRoberts, B. R. (2010). Problems of citation analysis: A study of uncited and seldom-cited influences. *Journal of the American Society for Information Science and Technology*, 61(1), 1–12.
- McClellan, J. E. (2003). Specialist control: The publications committee of the Académie Royale des Sciences (Paris) 1700–1793. *Transactions of the American Philosophical Society*, 93(3), i-134.
- Merton, R. K. (1979). *The sociology of science: An episodic memoir*. Carbondale: Southern Illinois University Press.
- Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis*, 79(4), 606–623.
- Muñoz-Leiva, F., Viedma-del-Jesús, M. I., Sánchez-Fernández, J., & López-Herrera, A. G. (2012). An application of co-word analysis and bibliometric maps for detecting the most highlighting themes in the consumer behaviour research from a longitudinal perspective. *Quality & Quantity*, 46(4), 1077–1095.
- Narin, F. (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Washington, DC: Computer Horizons.
- O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5), 673–690.
- Oppenheim, C., & Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, 29(5), 225–231.
- Persson, O. (1994). The intellectual base and research fronts of “jasis” 1986–1990. *Journal of the American Society for Information Science*, 45(1), 31–38.
- Price, D. J. (1970). *Citation measures of hard science, soft science, technology, and nonscience. Communication among scientists and engineers*. Lexington: D.C. Heath and Company.
- Real, R., & Vargas, J. M. (1996). The probabilistic basis of Jaccard's index of similarity. *Systematic Biology*, 45(3), 380–385.

- Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2), 131–134.
- Sanz-Casado, E., Garcia-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.
- Schubert, A., & Glänzel, W. (1983). Statistical reliability of comparisons based on the citation impact of scientific publications. *Scientometrics*, 5(1), 59–73.
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal*, 314(7079), 498–502.
- Simard, C., & West, J. (2006). *Knowledge networks and the geographic locus of innovation. Open innovation: Researching a new paradigm*. Oxford: Oxford University Press.
- Small, H. (1981). The relationship of information science to the social sciences: A co-citation analysis. *Information Processing and Management*, 17(1), 39–50.
- Tahamtan, I., Afshar, A. S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, 107(3), 1195–1225.
- Takeda, Y., & Kajikawa, Y. (2008). Optics: A bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*, 78(3), 543–558.
- Tijssen, R., Visser, M., & van Leeuwen, T. (2002). Benchmarking international scientific excellence: Are highly cited research papers an appropriate frame of reference? *Scientometrics*, 54(3), 381–397.
- Timilsina, G. R., Kurdgelashvili, L., & Narbel, P. A. (2012). Solar energy: Markets, economics and policies. *Renewable and Sustainable Energy Reviews*, 16(1), 449–465.
- Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42(1), 35–67.
- Wang, J. (2016). Knowledge creation in collaboration networks: Effects of tie configuration. *Research Policy*, 45(1), 68–80.
- Wang, C., Rodan, S., Fruin, M., & Xu, X. (2014). Knowledge networks, collaboration networks, and exploratory innovation. *Academy of Management Journal*, 57(2), 484–514.
- Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, 113(2), 331–360.
- Yamashita, Y., & Yoshinaga, D. (2014). Influence of researchers' international mobilities on publication: A comparison of highly cited and uncited papers. *Scientometrics*, 101(2), 1475–1489.
- Zhang, J. J., Yan, Y., & Guan, J. C. (2015). Scientific relatedness in solar energy: A comparative study between the USA and China. *Scientometrics*, 102(2), 1595–1613.