Physica A 465 (2017) 49–61

Contents lists available at ScienceDirect

Physica A

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

The rapid bi-level exploration on the evolution of regional solar energy development

Qing Guan, Haizhong An *, Huajiao Li, Xiaoqing Hao

School of Humanities and Economic Management, China University of Geosciences, Beijing 100083, China
Key Laboratory of Carrying Capacity Assessment for Resource and Environment, Ministry of Land and Resources, Beijing 100083, China
Open Lab of Talents Evaluation, Ministry of Land and Resources, Beijing 100083, China

HIGHLIGHTS

- Bi-level networks are classified.
- Economic region are used to explain the results.
- Practical information is explored as the market feedback.
- Combined approaches are used to quickly acquire information.
- Visualized evolution of local solar energy development in China is provided.

ARTICLE INFO

Article history:
Received 12 November 2015
Received in revised form 24 June 2016
Available online 12 August 2016

Keywords:
Solar energy
Regional development
Text mining
Complex network

ABSTRACT

As one of the renewable energy, solar energy is experiencing increased but exploratory development worldwide. The positive or negative influences of regional characteristics, like economy, production capacity and allowance policies, make them have uneven solar energy development. In this paper, we aim at quickly exploring the features of provincial solar energy development, and their concerns about solar energy. We take China as a typical case, and combine text mining and two-actor networks. We find that the classification of levels based on certain nodes and the amount of degree avoids missing meaningful information that may be ignored by global level results. Moreover, eastern provinces are hot focus for the media, western countries are key to bridge the networks and special administrative region has local development features; third, most focus points are more about the application than the improvement of material. The exploration of news provides practical information to adjust researches and development strategies of solar energy. Moreover, the bi-level exploration, which can also be expanded to multi-level, is helpful for governments or researchers to grasp more targeted and precise knowledge.

1. Introduction

The global fossil energy crisis and the requirements of pollution remission faced by most countries have promoted the development of renewable energy [1]. Thus, in recent years, solar energy has received significant attention from not only governments but also the public [2], and has got vigorous development prospect in the production and the application.

* Corresponding author at: School of Humanities and Economic Management, China University of Geosciences, Beijing 100083, China. Fax: +86 1082321783.
E-mail address: ahz369@163.com (H. An).

http://dx.doi.org/10.1016/j.physa.2016.08.007
0378-4371/

However, because of limit precedents of renewable energy, the development of solar energy is still in the exploratory stage. In fact, although lots of academic studies and research institution has made significant contributions to the material improvement of solar cells [3–5], distributed electricity generation [6–8], the ecological relation between solar energy and urban development [9–13], practical applications are still lagging compared with them, having to deal with conflicts with reality. As a result, acquiring the information on the development in solar energy utilization is instructive for both the realistic application and further studies. However, the utilization is affected by many exterior factors, like geology, economy or production capacity. It means that, for a certain country, different regions with their own characteristics have their own concerns on solar energy utilization. Especially for China, whose regional differences are apparent, its solar energy development is unevenly distributed across all provinces. As a result, as a typical case, the quick exploration of practical information on Chinese regional solar energy development is essential, which can be references for the governments to make or adjust their policies.

The evolution of regional solar energy development has become the focus point for researchers. C.L. Kwan [14] explored the distribution and influence elements of PV in the United States; A.M. Adil [15] provides general analysis on the renewable energy systems evolution from a socio-technical perspective; O. Yetemen [16] selects the ecohydrologic perspective to analyze their landscape evolution. These studies provide abundant information with technical, authentic or professional measurement, but few take the market as a whole to explore the application condition. As a result, our research tends to know the provincial solar energy development characteristic with a rapid exploration method.

To quickly grasp the general situation of a certain field [17,18], text mining has been widely used in previous studies. In contrast to academic articles which describe technological developments, news represents the timely communication of practical information [19]. In particular, news titles are the most direct and precise reflection of entire news articles, thereby providing an easy way for readers to quickly obtain the important points of the article. This type of text can reflect not only the real application and development of solar energy but also concerns from the public.

However, the information in the text is abundant but complex, thereby contributing to its unstructured characteristics. To solve this problem, the complex network, which has been widely improved for broader application [20,21] or used on energy issues [22,23], is useful for analyzing the features of each object and the relations among them [24,25]. In particular, two-actor networks can effectively explore information between two sets of objects with different attributes [26]. For example, they can be used to study the relations between energy companies and their shareholders [26] or between ethnic violence and other factors such as environmental resources or social–political terms [23]. These studies, which apply two-actor networks to the text mining have demonstrated that this method is useful for quickly obtaining required information [27]. These papers focus on the feature and evolution of the entire pattern [28]. However, local information represents a fresh perspective for exploring concentrated features of required objects. Thus, our work combines two-actor networks with text mining to study regional solar energy development in China from the perspective of both global and local evolution features as the two different levels.

Our work aims at quickly extracting practical and targeted information for energy issue, especially the renewable energy market. We use news titles as data source, and two-actor networks between the set of regional nouns and non-regional nouns have been constructed as the models. First, we make general analysis on the global features of bi-level networks, including the pattern scale and controllability. Then, we explore top 10 hot provinces that are mentioned by most news articles many times to explore provinces that receive substantial attention from the media and the public. Top 10 key provinces are also explored as the bridge for the network construction. Moreover, closeness among provinces is measured, and the local information from the relation between two levels is given to differentiate key development aspects for each province.

The remainder of this paper is organized as follows. The next section introduces the data source and the methodology. Section 3 analyzes provincial solar energy development from global and local information. Section 4 presents a discussion, and Section 5 concludes this study.

2. Data and methodology

2.1. Data

2.1.1. The data extraction

We obtained our data from the Chinese Energy Website (http://www.china5e.com/new-energy/solar-power/). Not similar with some common Chinese search engine, the Chinese Energy Website is a professional one with the collection of all energy-related news and information, especially focusing on the practical application of energy. As a result, the data collection from this website corresponds with our research goal with relative low amount of noisy data.

For the data extraction, the selected toolkit in this research is LocoySpider (http://www.locoy.com/), which is a commonly used one to collect online textual data. Because there is a specialized news board about solar energy development, we directly extract all news titles in this news board from 2010 to 2014. A total of 16 893 pieces of news titles provide 70 288 valid relations among words. Data for each year constitute a data set. Thus, we obtain five data sets and study on them individually.

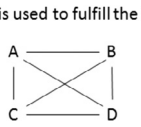

Fig. 1. The relations among the nouns in each title.

	t_1	t_2	⋯	t_m
x_1	a_{11}	a_{12}	⋯	a_{1m}
x_2	a_{21}	a_{22}	⋯	a_{2m}
⋮	⋮	⋮	⋯	⋮
x_n	a_{n1}	a_{n2}	⋯	a_{nm}

(a) Affiliation matrix: T–X.

	t_1	t_2	⋯	t_m
y_1	b_{11}	b_{12}	⋯	b_{1m}
y_2	b_{21}	b_{22}	⋯	b_{2m}
⋮	⋮	⋮	⋯	⋮
y_k	b_{k1}	b_{k2}	⋯	b_{km}

(b) Affiliation matrix: T–Y.

Fig. 2. Two affiliation relations between titles and region nouns or non-region nouns.

2.1.2. The data processing

Based on the extracted data, we use the toolkit called NLPIR (the original name is ICTCLAS) to make word segmentation. NLPIR has been acknowledged by the researchers as the most authoritative and effective Chinese word segmentation software, which contains the dictionary with almost 260 thousand Chinese words and have distinguished word segmentation ability. We input all news data into it, and then get the output data, which includes the segmented words and their corresponding part-of-speech for each news title.

We have cleaned some noisy data, and have corrected some wrong segmentation results by modifying the word dictionary. Furthermore, our study attempts to explore the evolution of regional solar energy development in China. However, certain region nouns are cities that appear in the same title as their corresponding province. This leads to a redundant calculation for the same province, which causes confusion in the subsequent analysis. To solve this problem, we de-duplicate them. As a result, we obtain one set of province nouns and another set of non-region nouns.

2.2. Methodology

In our study, we construct networks of two sets of actors from 2010 to 2014. One set involves region nouns, and the other set includes all non-region nouns that are related to the development or industry chain of solar energy. After using ICTCLAS to code the type for each word in each title, which includes nouns, verbs, adjectives or adverbs, we extract all nouns, clean them and remove incorrect nouns. Then, we consider the nouns from one title having relations with each other. Fig. 1 shows the relations among the nouns A, B, C and D, which are coded as nouns in one sentence.

In the following subsections, we classify the nouns into region nouns, such as Beijing, Shanghai and Xuzhou, and non-region nouns such as photovoltaic, roof and grid-connected. In addition, we explain the relational networks, which are expressed mathematically.

2.2.1. Affiliation relations

The region nouns and non-region nouns both belong to their news titles, thereby contributing to affiliation relations. Let $T = \{t_1, t_2, \ldots, t_m\}$ represents the set of all news titles obtained from the website, where $t_i \in T$ is each title. Additionally, $X = \{x_1, x_2, \ldots, x_n\}$ represents the set of all region nouns that are Chinese districts, and $Y = \{y_1, y_2, \ldots, y_k\}$ represents the set of all non-region nouns that belong to a title but that are not region nouns. As a result, we obtain two types of affiliation relations (T and X, T and Y).

Fig. 2 shows the two affiliation relations, which are defined as [19]

$$a_{ij} = \begin{cases} 0, & x_i \quad \text{does not belong to } t_i \\ 1, & x_i \quad \text{belongs to } t_j \end{cases} \tag{1}$$

$$b_{ij} = \begin{cases} 0, & y_i \quad \text{does not belong to } t_i \\ 1, & y_i \quad \text{belongs to } t_j \end{cases} \tag{2}$$

where a_{ij} describes the affiliation between titles and region nouns and b_{ij} indicates the affiliation between titles and non-region nouns.

	y_1	y_2	...	y_k
x_1	c_{11}	c_{12}	...	c_{1k}
x_2	c_{21}	c_{22}	...	c_{2k}
⋮	⋮	⋮	...	⋮
x_n	c_{n1}	c_{n2}	...	c_{nk}

Fig. 3. Co-occurrence relations between region nouns and non-region nouns.

2.2.2. Two sets of actors

According to Batagelj [29] and Li et al. [30], when there are two nouns that have the same relationship with a title, they are equivalent. This type of relation between region nouns and non-region nouns is always called co-occurrence, which is shown in Fig. 3.

We can define the co-occurrence relations as

$$c_{ij} = \sum_{l=1}^{m} a_{il} b_{jl} \tag{3}$$

where each co-occurrence relation can be calculated using the two defined affiliation relations.

2.2.3. Network modeling and topological indicators

Taking region and non-region nouns as nodes and their co-occurrence relations as edges, we model a network that describes regional solar energy development. Because we use two sets of nouns with different types, the network can be called two-actor network. There is no direction, but the weight of each edge is defined by the number of times the pair of words appears in the same title. Basically, province nouns and related nouns are the two actors, and the number of times that they co-occur in the same news title is the weight of the edge between them. We construct five networks from 2010 to 2014.

Compared with the network with only one kind of nodes, two-actor networks help to classify objects, which would provide two-dimensional perspective with more abundant information. We measure provincial solar energy development based on their direct related topics and the indirect ones, because of the connections among topics.

Topological indicators are used to explore the statistic feature of solar energy development for each province and each related news topic. We select a number of classic indicators to measure the topological characteristics from different perspectives.

(1) Indicators for global pattern analysis.

Average degree (AD), Average weighted degree (AWD), network diameter (ND) and average path length (APL) are selected to measure the evolution of global pattern. These indicators provide different information.

AD is used to describe the general number of news topics that are related to each province. If the network has a high AD value, in that year, most provinces are active in various solar energy development fields. AD is defined as [31]

$$S^{AD} = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{4}$$

where N is the total number of nodes in a network, and k_i is the node v_i's number of neighbor nodes.

AWD is similar with AD, but considers the concurrence times of each pair of news topics. As a result, for example, the province with high AWD appears frequently in that year, and is related with many solar energy development news topics. AWD's definition is as follows [31]:

$$S^{AWD} = \frac{1}{N} \sum_{i=1}^{N} S_i^{WD} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \in N_i} w_{ij} \tag{5}$$

where N_i represents the set of v_i's neighbor nodes, w_{ij} is the weight of the link between v_i and v_j.

ND measures the scale of the network, which is always used to estimate the information transmission speed [32]. In our work, the network with relative high ND approves that solar energy news includes abundant but relative independent sub-topics.

$$S_{ND} = \max_{1 \le i,j \le N} d_{ij} \tag{6}$$

where d_{ij} is the minimum path length from v_i to v_j.

APL helps to describe the relevance for each pair of nouns. Smaller APL means that these two nouns are easier to be reported and talked together with the media. It is defined as follows [32]:

$$S^{APL} = \frac{2}{N(N-1)} \times \sum_{i>j} d_{ij}. \tag{7}$$

(2) Indicators for local pattern analysis.

We analyze local pattern characteristics for provinces based on direct and indirect information. According to direct information, we use weighted degree (WD) to describe the degree of attention for provinces. A high or low value of the WD represents two types of information. One type is whether the province has relations with more topics, and another type is whether the province is frequently reported by the media. The node v_i's WD is defined as [31]

$$S_i^{WD} = \sum_{j \in N_i} w_{ij}. \tag{8}$$

To describe provinces' importance based on indirect information, we adopt three indicators: clustering coefficient (CC), betweenness centrality (BC) and closeness centrality (CLC). In our research, CC represents the possibility that the linked topics also co-occurred in same news titles, BC is used to describe the times that the province appear in the indirect topic links, and CLC describes each province's closeness with any other topics.

CC's definition is [33]

$$S_i^{CC} = 2M_i / [k_i (k_i - 1)] \tag{9}$$

where M_i represents the number of links among v_i's k_i neighbor nodes.

BC's definition is [34]

$$S_i^{BC} = 2 \sum_{j,l,j \neq l \neq i} \left[N_{jl}(i) / N_{jl}\right] / [(N-2)(N-1)] \tag{10}$$

where N_{jl} represents the number of minimum paths between v_j and v_l, and $N_{jl}(i)$ is the number of nodes that the minimum paths between v_j and v_l go through.

CLC's definition is [35]

$$S_i^{CLC} = (N-1) \Big/ \left[\sum_{j=1, j \neq i}^{N} d_{ij}\right]. \tag{11}$$

(3) Indicator for measuring regional solar energy development communities.

Social networks are always divided into communities to explore nodes with close relations than others. In our studies, based on the algorithm developed by Vincent et al. [36], we divided the network in each year into communities. As a result, provinces in the same community have closer relations from the perspective of news topics.

This algorithm estimates the network with an indicator called modularity, which is used to measure the density of links inside communities compared with the links among communities. Its high value approves that the partition is good. It is defined as

$$Q = \frac{1}{2m} \sum_{i,j} \left[w_{ij} - \frac{A_i A_j}{2m}\right] \delta\left(c_i, c_j\right) \tag{12}$$

where w_{ij} is the weight of the link between v_i and v_j, $A_i = S_i^{WD}$, c_i is the community that v_i is assigned, $\delta\left(c_i, c_j\right)$ is 1 if $c_i = c_j$ and 0 otherwise, and $m = \frac{1}{2}\sum_{i,j} w_{ij}$.

The community detection includes two phases that are repeated iteratively. First, each node v_i itself is considered as an independent community, and then the algorithm evaluates its modularity ΔQ when v_i is placed into its neighbor community. After the consideration of v_i's every neighbor community, if the gain is positive, v_i would join in the community with highest ΔQ. The first phase would be repeated for all nodes until no further improvement can be achieved. The definition of ΔQ is as follows:

$$\Delta Q = \left[\frac{\sum C_{in} + A_{i,in}}{2m} - \left(\frac{\sum tot + A_i}{2m}\right)^2\right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum tot}{2m}\right)^2 - \left(\frac{A_i}{2m}\right)^2\right] \tag{13}$$

where $\sum C_{in}$ is the sum of all the links' weight inside the community C, $\sum tot$ is the sum of the links' weight that are incident to all nodes in the community C, $A_{i,in}$ is the sum of the links' weight that from v_i to all the nodes in C.

In the second phase, a new network is built up with communities found in the first phase as nodes, and the links' weight between nodes in the corresponding two communities is the new network's link weight. Edges between nodes of the same community lead to self-loop in the new network. The completion of the second phase would lead to the reapplication of the first phase on the resulting network. The iteration would stop is there are no more changes.

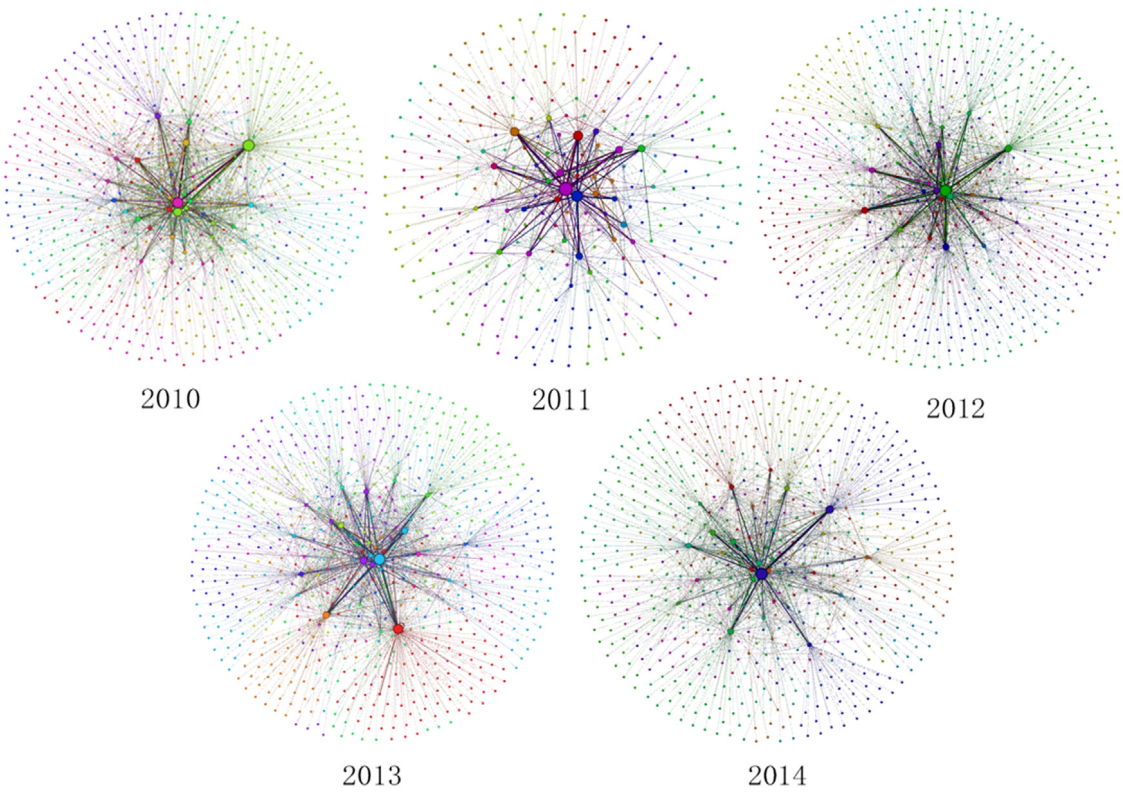

Fig. 4. The two-actor networks from 2010 to 2014.

3. Results and analysis

In this paper, we construct five networks of two sets of actors, which are shown in Fig. 4. One set includes region nouns, and the other includes non-region nouns. The relations among these two sets or in the same set are the times that nouns co-occurred in same news titles.

Although the features are different for each year, Fig. 4 shows that there are some common characteristics among these five networks. First, the relations at the core position are dense and complicated. Second, certain points connect to a number of points that only have this one connection. This means that, in different regions, the common points of focus are similar. However, the particularities of each province are such that they each have their own characteristics in terms of solar energy development. For example, in 2014, only Beijing was related to reports on inverters, and only Shanghai was related to reports on allowances.

To study the features and the evolution of regional solar energy development, we analyze the networks from three aspects: the global patterns of five networks, the local topological characteristics of provinces, the evolution and local feature of hot solar energy related news topics, and the importance of certain co-occurrence relations between the two sets.

3.1. The global pattern of the bi-level networks

The global topological characteristics of the five networks of two sets of actors are shown in Fig. 5, which approves the evolution of news titles about integral solar energy development in China.

Although China is a leader of global photovoltaic export, Fig. 4 shows that the abundance and scale of news topics from the public media is not increasing, but exhibits fluctuations. Generally, the fluctuation degree of the nodes is larger than that of the edges, which means that the fluctuation of a small number of nodes can lead to larger fluctuations of edge numbers.

Additionally, it is apparent that all global indicators have apparent fluctuation in 2011. In this year, news topics and their relations decrease significantly with the decrease of nodes from 611 in 2010 to 316 in 2011 and edges from 1395 in 2010 to 625 in 2011, which sequentially leads to the obvious change of network with higher ND and APL but lower AD and AWD. High ND and APL approve the controllability of the network is hard in 2011, while the low AD and AWD is mostly related to the decrease of nodes and links. Moreover, the same ND and huge gap in number of nodes and links in 2011 and 2012 approves that the recovery of network controllability is slower than the recovery of the network scale.

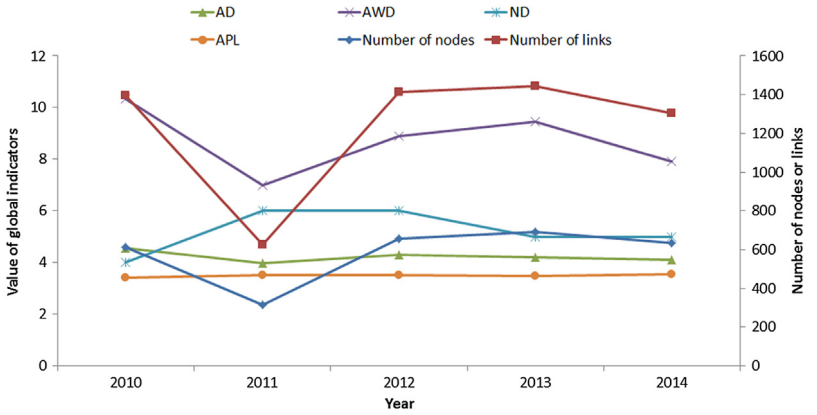

Fig. 5. The scale of the two-actor network from 2010 to 2014.

3.2. Hot provinces

Hot provinces are those that are mentioned in the news titles with highest times. These provinces get most focus from the media, which approves that they are the most active provinces of developing solar energy. Moreover, WD considers not only connection universality but also the frequency of these connections for each node. As a result, a province with a higher WD is likely to be a hot one that has become the focus at that period of time. We show the evolution of regional solar energy development during these five years on the map of China in Fig. 6. The WD values of all province nouns are classified equally into eight degrees, which are shown in the legend. This facilitates comparisons among these five years.

Generally, eastern provinces received the most attention in 2010, but then all topics received substantially less attention in 2011. However, in 2012, western provinces received substantially more attention compared with other years, for example, Sinkiang is almost as popular as Jiangsu. In 2013, the number of topics in the western parts was smaller in number, but it was greater in eastern parts. However, in 2014, all topics received substantially less attention again. This demonstrates that the media focused on solar energy development in western provinces but that this concern was short lived and decreased substantially in recent years. By contrast, the attention paid by the media to eastern provinces continuously increased during these five years, except in 2011. It seems that the media lost interest in solar energy development in 2011 and again in 2014.

In detail, based on the above analysis, we conclude that the most popular province with high development of solar energy is Jiangsu because of its high productivity. In addition, western provinces' solar energy development receive little attention from the media in recent years. Generally, solar energy development in the cities located in the eastern part of China is receiving more attention from the media.

3.3. Provincial importance

We explore provincial importance to measure their contribution in the topic correlation. We select three indicators to measure province's importance from different perspectives, and respectively, top 10 provinces with highest value are selected.

To explore the law between provincial importance and provincial economic development, we classify the results into different regions. According to the new situation of Chinese economic and social quick development, provinces are divided into four economic regions: eastern region, northeastern region, central region and western region. Different regions have their own development characteristics. Western region focuses on the further development, which includes Sinkiang, Gansu, Qinghai and so on. Northeastern region focus on the revitalization, which includes Liaoning, Jilin and so on. Central region, including Jiangxi, Anhui, Shanxi, Hubei and so on, is rising, and the eastern region, which includes Jiangsu, Zhejiang, Beijing and so on, takes the lead in the economic development. Apart from these four economic regions, special administrative region include Hong Kong, Macao and Taiwan.

CLC is used to approve that the word has closer relation with most other topics, which means that this word is the most active one. Top 10 words with highest value of CLC are shown in Table 1. It shows that most active provinces are concentrated on eastern region with relative better economic development, but also on western region which is not the focus from the media in recent years. It means that although western region provinces are not the hottest ones with most reports from the media, but it is still active, and plays an essential role in the solar energy development in China.

BC is used to describe whether a province is the bridge among groups of topics. Table 2 shows the top 10 provinces with highest BC value. It approves that more western provinces play the role of connecting different groups of topics. Moreover, some provinces in central region have better rankings compared with the Table 1. Especially for Anhui, its main role is the connection hub, which means that Anhui tends to be the solar energy development province with relative independent policies for different regions.

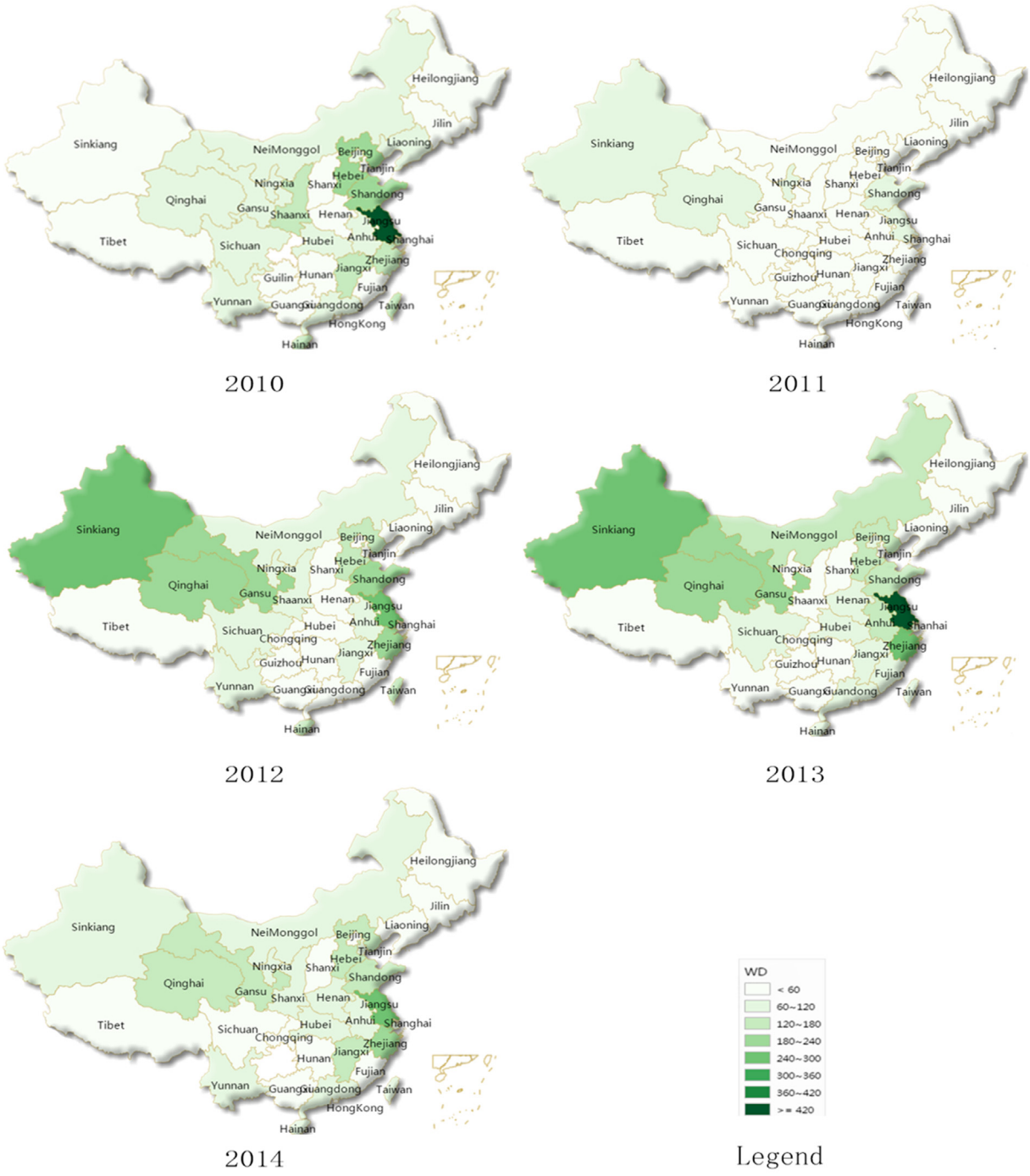

Fig. 6. Regional evolution of focus points from solar energy news.

In our research, the province with high CC value is not always the most important or hot one, but has its own solar energy development characteristics. As a result, we can see from Table 3, which shows top provinces with high CC value, that some special administrative region, like Hong Kong, appears at the top ranking. Under the difference of regime and developing policies of renewable energy, Hong Kong's related topics are different from others.

3.4. The closeness among provinces

There are cooperation, similarities or indirect relations among provincial solar energy development, which promote the closeness among part of provinces. The exploration on it is helpful for understanding provincial solar energy development

Table 1
Top 10 provinces with highest CLC value.

	2010		2011		2012		2013		2014	
	Province	R	Province	R	Province	R	Province	R	Province	R
1	Jiangsu	E	Jiangsu	E	Jiangsu	E	Jiangsu	E	Jiangsu	E
2	Zhejiang	E	Sinkiang	W	Zhejiang	E	Sinkiang	W	Gansu	W
3	Taiwan	S	Shandong	E	Sinkiang	W	Zhejiang	E	Zhejiang	E
4	Shandong	E	Qinghai	W	Shandong	E	Ningxia	W	Shanghai	E
5	Jiangxi	C	Tibet	W	Guangdong	E	Qinghai	W	Qinghai	W
6	Hebei	E	Guangdong	E	Qinghai	W	Shandong	E	Shandong	E
7	Beijing	E	Hebei	E	Jiangxi	C	Shanghai	E	Hebei	E
8	Shanghai	E	Zhejiang	E	Gansu	W	Nei Mongol	N	Sinkiang	W
9	Liaoning	N	Jiangxi	C	Beijing	E	Gansu	W	Jiangxi	C
10	Hong Kong	S	Ningxia	W	Anhui	C	Anhui	C	Anhui	C

Notes: R: Region; E: Eastern; S: Special Administrative; W: Western; C: Central.

Table 2
Top 10 provinces with highest BC values.

	2010		2011		2012		2013		2014	
	Province	R	Province	R	Province	R	Province	R	Province	R
1	Jiangsu	E	Jiangsu	E	Jiangsu	E	Jiangsu	E	Jiangsu	E
2	Hebei	E	Qinghai	W	Zhejiang	E	Zhejiang	E	Zhejiang	E
3	Shanghai	E	Shandong	E	Shandong	E	Shandong	E	Gansu	W
4	Shandong	E	Zhejiang	E	Sinkiang	W	Anhui	C	Shandong	E
5	Jiangxi	C	Hebei	E	Hebei	E	Qinghai	W	Anhui	C
6	Taiwan	S	Ningxia	W	Qinghai	W	Sinkiang	W	Qinghai	W
7	Zhejiang	E	Guangdong	E	Taiwan	S	Gansu	W	Taiwan	S
8	Shannxi	W	Sichuan	W	Gansu	W	Taiwan	S	Jiangxi	C
9	Beijing	E	Jiangxi	C	Shannxi	W	Jiangxi	C	Hebei	E
10	Guangdong	E	Sinkiang	W	Sichuan	W	Nei Mongol	N	Shanghai	E

Notes: R: Region; E: Eastern; S: Special Administrative; W: Western; C: Central.

Table 3
Top 10 provinces with highest CC values.

	2010		2011		2012		2013		2014	
	Province	R	Province	R	Province	R	Province	R	Province	R
1	Hong Kong	S	Jilin	N	Jilin	N	Ningxia	W	Hong Kong	S
2	Zhejiang	E	Sinkiang	W	Shanxi	C	Sinkiang	W	Sichuan	W
3	Taiwan	S	Tibet	W	Shanghai	E	Shanghai	E	Shanghai	E
4	Tianjin	E	Shanghai	E	Guangdong	E	Guangdong	E	Nei Mongol	N
5	Shanxi	C	Taiwan	S	Beijing	E	Gansu	W	Beijing	E
6	Gansu	W	Guangdong	E	Nei Mongol	N	Qinghai	W	Yunnan	W
7	Anhui	C	Jiangsu	E	Hunan	C	Nei Mongol	N	Gansu	W
8	Liaoning	N	Shandong	E	Yunnan	W	Hubei	C	Sinkiang	W
9	Beijing	E	Qinghai	W	Jiangxi	C	Shandong	E	Qinghai	W
10	Jiangsu	E	Zhejiang	E	Zhejiang	E	Zhejiang	E	Jiangsu	E

Notes: R: Region; E: Eastern; S: Special Administrative; W: Western; C: Central.

in groups, which is also good for making related policies for certain regions. In 2010–2014, provinces belong to different communities with each other, which are based on their changes of closeness. The closeness has been quantitatively described based on the measurement of communities in the networks. We also divide provinces into four economic regions to explore potential rules, and it is shown in Table 4. For each community in a year, we count the number of provinces which belongs to different regions.

Although the distribution of regions is complex, it can be seen that many communities tend to consists of provinces from same region. For example, in 2013, the provinces in the 7th community are all eastern ones, and in 2012, provinces in the first community are most western ones. Moreover, it is apparent that the region composition in communities has changed. Especially, in 2014, most communities are composed by one eastern province and one central province or one western province. That is to say, some solar energy topics happen in them are similar, and their development are consistent.

From the perspective of closeness and economic level, we explored potential rules of solar energy development for provinces. We find that provinces with similar economic level tend to have close development relations. But in 2014, the correlation or cooperation tends to happen between one eastern province and one central or western province.

Table 4
Number of provinces in the community in certain year.

Community	Region	2010	2011	2012	2013	2014
1	Eastern	1	1	1	3	
	Northeastern		1		1	
	Central		1		1	
	Western		2	4	2	2
2	Eastern	1	1	1		1
	Northeastern					
	Central	1				1
	Western	2			3	
3	Eastern	3	2		2	2
	Northeastern	2			1	
	Central	1		1	2	
	Western	1				3
4	Eastern		1	2		
	Northeastern					2
	Central	1	2		1	1
	Western	1	3		1	1
5	Eastern	1	1	2		1
	Northeastern					
	Central		1			1
	Western	2			1	
6	Eastern	1	1		1	1
	Northeastern					
	Central		1			
	Western		1	1		1
7	Eastern	1			3	1
	Northeastern					
	Central			1		1
	Western	1	2	1		
8	Eastern		2			1
	Northeastern				1	
	Central			1		1
	Western	1			1	
9	Eastern	1	1	1		1
	Northeastern		2			
	Central	3	1	1	2	
	Western		2			1
10	Eastern	1		2		
	Northeastern	1		2		
	Central			1		
	Western	2		2	1	2
11	Eastern			1	1	2
	Northeastern			1		1
	Central			1		1
	Western			2	1	

3.5. Common hot nouns for hot provinces

Based on the information from bi-level networks, we have made more abundant analysis on provincial solar energy development, and we have obtained general results for all provinces. In this section, to get more local and targeted results for provinces, we expand the analysis into the relations between two levels. We reserve top 10 provinces for each year. Simultaneously, to determine which nouns are common topics for hot provinces, we keep related nouns that have connections with at least five provinces. Fig. 7 shows the five networks where red nodes represent provinces, the other colors are their related nouns, and the size of each noun is the WD.

Although the results for each year fluctuate, common nouns include ‘EG’, ‘Energy’, ‘Enterprise’, ‘GC’, ‘Industry’, ‘PS’ and ‘Project’. In detail, first, by comparing the results for these five years, we find that there are common topics among these five years. Although certain nouns, such as ‘Roof’ and ‘Electricity’, may obtain high rankings in a given year, they are not common topics for hot provinces over all five years. Second, ‘System’ and ‘Market’ are nouns with low WD value in five bi-level networks, which means that they have not received significant attention by the media in any year. However, they are the common focus for most hot provinces, which is not same with the results from the perspective of global network pattern. Thus, these nouns should also be focused on by researchers.

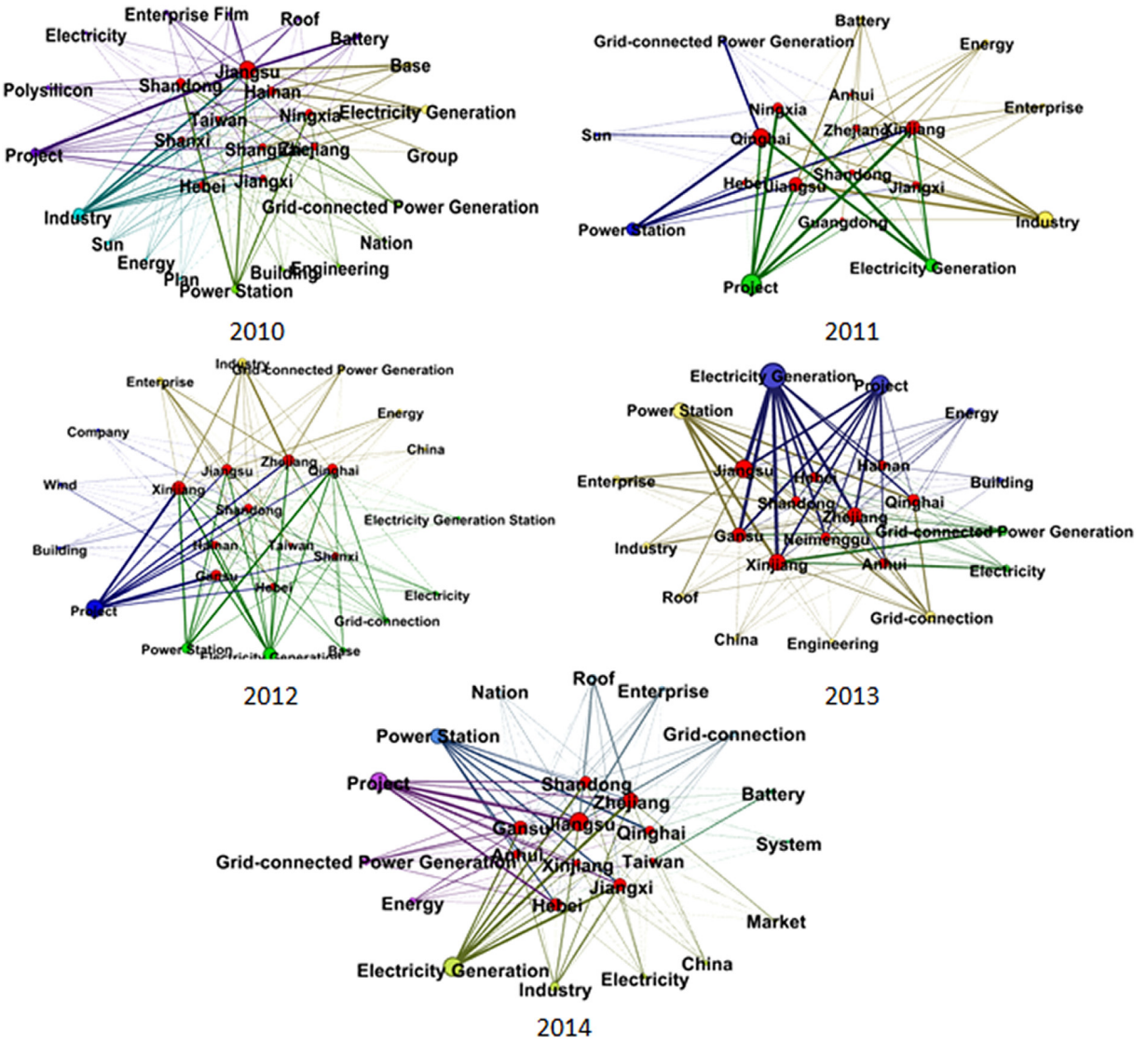

Fig. 7. The network of common hot words and top 10 provinces. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

(1) The application and contribution on energy related issues.

Little previous researches combine the text mining and complex network to study energy related issues. However, it is an effective way to gather market information as soon as possible and objective. Text mining from the news titles provides energy market with practical information instead of just the improvement of technologies. That is to say, apart from improving theoretical basis, it is also necessary to accumulate the feedback from the practice. As a result, applying text mining and complex network on energy related issues is the main contribution of our work. We use news titles as our data source to provide most fresh practical information. And bi-level networks have cleared our research object and have improved information extraction speed.

(2) Provincial solar energy development is uneven.

Our results have revealed Chinese provincial solar energy development characteristics. The whole development pattern is unstable with fluctuate related number of provinces or topics. Although eastern provinces are the hot focus for the media, western provinces also play an important role for bridging the whole topic network. Moreover, it is interesting that the closeness among provinces is related to their economic development levels, but in 2014, the closeness tend to happen between a eastern province and a central or western province with cooperation or similar development way. Furthermore,

the information between two levels approves the development focus for hot provinces, like industry, power station, grid-connection and so on.

As a result, our method is able to reveal abundant information for provincial development through the information from another set of actor, like the news topics in our research. The analysis can be expended from one level to bi-level, which include more target results.

(3) The information exploration from bi-level network with two set of actors.

Two sets of actors help us to distinguish the information. If we see networks with two sets of actors as a whole, the information is abundant but redundant, which make the analysis inaccurate. However, the application on energy issue approves that exploring information from a certain set of actors can provide more accurate and accurate information. For example, in our research, provincial solar energy development is known from the set of related news topics. Our data includes all the related topics to provide a general and macroscopic perspective to understand the market. Further, we can divide actors into more sets, like the words of materials, finance, trade and so on, to make more target analysis.

5. Conclusion

In our study, we applied the text mining of news titles and the bi-level network of two actors on an energy issue. Especially for renewable energy market, the exploration of practical information is necessary for governments, researchers or enterprise leaders to adjust their development strategies. News titles help to provide fresh practical information quickly and the bi-level networks is good for exploring targeted knowledge. Their combination on energy issue helps us find that the whole pattern is fluctuant with changes of provincial solar energy development roles. Based on the information from the set of topics, the closeness among provinces reveals the changes of their development strategies and route. Moreover, the local characteristics of the relation between two levels help to avoid missing information that embedded in global information. For example, industry is still the development focus for relative solar energy developed provinces, but it is neglected when considering the global information. As a result, the application of news titles text mining and the two actors bi-level networks on energy issue is useful and feasible to promote the understanding of energy market.

In our future studies, abundant data from authoritative websites must be obtained to allow us to analyze detailed information in future studies. Additionally, we will improve our analysis by considering the complexity of Chinese terms.

Acknowledgments

This research is supported by grants from the National Natural Science Foundation of China (Grant No. 71173199) and the open project of Key Laboratory of Carrying Capacity Assessment for Resource and the Environment, Ministry of Land and Resources (No. CCA2016.01). The authors would like to express their gratitude to Xiangyun Gao, Xiaoliang Jia, Shupei Huang, Xiaojia Liu and Lijun Wang who helped a lot during their work.

References

[1] B. Algieri, A. Aquino, M. Succurro, Going "green": trade specialisation dynamics in the solar photovoltaic sector, Energy Policy 39 (2011) 7275–7283.
[2] D. Yang, J. Liu, J. Yang, N. Ding, Life-cycle assessment of China's multi-crystalline silicon photovoltaic modules considering international trade, J. Cleaner Prod. 94 (2015) 35–45.
[3] G. Li, V. Shrotriya, J.S. Huang, Y. Yao, T. Moriarty, K. Emery, Y. Yang, High-efficiency solution processable polymer photovoltaic cells by self-organization of polymer blends, Nature Mater. 4 (2005) 864–868.
[4] M. Law, L.E. Greene, J.C. Johnson, R. Saykally, P.D. Yang, Nanowire dye-sensitized solar cells, Nature Mater. 4 (2005) 455–459.
[5] S. Juneja, S. Sudhakar, J. Gope, K. Lodhi, M. Sharma, S. Kumar, Highly conductive boron doped micro/nanocrystalline silicon thin films deposited by VHF-PECVD for solar cell applications, J. Alloys Compd. 643 (2015) 94–99.
[6] R. Kannan, K.C. Leong, R. Osman, H.K. Ho, C.P. Tso, Life cycle assessment study of solar PV systems: An example of a 2.7 kW(p) distributed solar PV system in Singapore, Sol. Energy 80 (2006) 555–563.
[7] C.A. Hill, M.C. Such, D.M. Chen, J. Gonzalez, W.M. Grady, Battery energy storage for enabling integration of distributed solar power generation, IEEE Trans. Smart Grid 3 (2012) 850–857.
[8] W. Grossmann, I. Grossmann, K.W. Steininger, Solar electricity supply isolines of generation capacity and storage, Proc. Natl. Acad. Sci. USA 112 (2015) 3663–3668.
[9] Y. Xiao, H.Y. Xu, H.M. Xie, Z.H. Yang, G.M. Zeng, Comparison of the treatment for isopropyl alcohol wastewater from silicon solar cell industry using SBR and SBBR, Int. J. Environ. Sci. Technol. 12 (2015) 2381–2388.
[10] T.W. Hamann, R.A. Jensen, A.B.F. Martinson, H. Van Ryswyk, J.T. Hupp, Advancing beyond current generation dye-sensitized solar cells, Energy Environ. Sci. 1 (2008) 66–78.
[11] G. Frisari, M. Stadelmann, De-risking concentrated solar power in emerging markets: The role of policies and international finance institutions, Energy Policy 82 (2015) 12–22.
[12] J.A. Khan, M. Mustafa, T. Hayat, A. Alsaedi, Three-dimensional flow of nanofluid over a non-linearly stretching sheet: An application to solar energy, Int. J. Heat Mass Transfer 86 (2015) 158–164.
[13] B. Alorda, R. Pujol-Nadal, G. Rodriguez-Navas, A. Moia-Pol, V. Martinez-Moll, Collaborative distributed sun-tracking control system for building integration with minimal plant area and maximum energy-conversion efficiency, Int. J. Electr. Power Energy Syst. 70 (2015) 52–60.
[14] A.M. Adil, Y. Ko, Socio-technical evolution of decentralized energy systems: A critical review and implications for urban planning and policy, Renewable Sustainable Energy Rev. 57 (2016) 1025–1037.
[15] C.L. Kwan, Influence of local environmental, social, economic and political variables on the spatial distribution of residential solar PV arrays across the United States, Energy Policy 47 (2012) 332–344.
[16] O. Yetemen, E. Istanbulluoglu, J.H. Flores-Cervantes, E.R. Vivoni, R.L. Bras, Ecohydrologic role of solar radiation on landscape evolution, Water Resour. Res. 51 (2015) 1127–1157.
[17] W.J. Lee, S.Y. Sohn, Patent analysis to identify shale gas development in China and the United States, Energy Policy 74 (2014) 111–115.

[18] N.L. Papanikolaou, G.A. Pavlopoulos, T. Theodosiou, I. Iliopoulos, Protein-protein interaction predictions using text mining methods, Methods 74 (2015) 47–53.
[19] H.J. Li, W. Fang, H.Z. An, X. Huang, Words analysis of online Chinese news headlines about trending events: A complex network perspective, PLoS One 10 (2015).
[20] Z.K. Gao, Y.X. Yang, L.S. Zhai, M.S. Ding, N.D. Jin, Characterizing slug to churn flow transition by using multivariate pseudo Wigner distribution and multivariate multiscale entropy, Chem. Eng. J. 291 (2016) 74–81.
[21] Z.K. Gao, N.D. Jin, A directed weighted complex network for characterizing chaotic dynamics from time series, Nonlinear Anal. RWA 13 (2012) 947–952.
[22] T. Van Holt, J.C. Johnson, J.D. Brinkley, K.M. Carley, J. Caspersen, Structure of ethnic violence in Sudan: a semi-automated network analysis of online news (2003-2010), Comput. Math. Organ. Theory 18 (2012) 340–355.
[23] J. Diesner, K.M. Carley, L. Tambayong, Extracting socio-cultural networks of the Sudan from open-source, large-scale text data, Comput. Math. Organ. Theory 18 (2012) 328–339.
[24] Z.K. Gao, P.C. Fang, M.S. Ding, N.D. Jin, Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow, Exp. Therm Fluid Sci. 60 (2015) 157–164.
[25] Z.K. Gao, Y.X. Yang, P.C. Fang, N.D. Jin, C.Y. Xia, L.D. Hu, Multi-frequency complex network from time series for uncovering oil-water flow structure, Sci. Rep. 5 (2015) 8222.
[26] H.J. Li, W. Fang, H.Z. An, L.L. Yan, The shareholding similarity of the shareholders of the worldwide listed energy companies based on a two-mode primitive network and a one-mode derivative holding-based network, Physica A 415 (2014) 525–532.
[27] A. Menendez-Manjon, K. Moldenhauer, P. Wagener, S. Barcikowski, Nano-energy research trends: bibliometrical analysis of nanotechnology research in the energy sector, J. Nanopart. Res. 13 (2011) 3911–3922.
[28] R. Irfan, C.K. King, D. Grages, S. Ewen, S.U. Khan, S.A. Madani, J. Kolodziej, L.Z. Wang, D. Chen, A. Rayes, N. Tziritas, C.Z. Xu, A.Y. Zomaya, A.S. Alzahrani, H.X. Li, A survey on text mining in social networks, Knowl. Eng. Rev. 30 (2015) 157–170.
[29] V. Batagelj, Social network analysis: Methods and applications, Psychometrika 63 (1998) 103–104.
[30] H.J. Li, H.Z. An, X.Y. Gao, J.C. Huang, Q. Xu, On the topological properties of the cross-shareholding networks of listed companies in China: Taking shareholders' cross-shareholding relationships into account, Physica A 406 (2014) 80–88.
[31] L.C. Freeman, Centrality in social networks: Conceptual clarification, Soc. Netw. 1 (1979) 215–239.
[32] D.J. Watts, S.H. Strogatz, Collective dynamics of "small-world" networks, Nature 339 (1998) 440–442.
[33] J.P. Onnela, J. Saramaki, J. Kertesz, K. Kaski, Intensity and coherence of motifs in weighted complex networks, Phys. Rev. E 71 (2005) 065103(R).
[34] K.I. Goh, E. Oh, B. Kahng, D. Kim, Betweenness centrality correlation in social networks, Phys. Rev. E 67 (2003) 017101.
[35] U. Brandes, A faster algorithm for betweenness centrality, J. Math. Sociol. 25 (2001) 163–177.
[36] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 10 (2008) 10008.