Geoderma 305 (2017) 281–292

Contents lists available at ScienceDirect

Geoderma

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

A critical review on the research topic system of soil heavy metal pollution bioremediation based on dynamic co-words network measures

Rong Luo, Jia Li, Yiming Zhao, Xiuwen Fan, Peng Zhao, Lihe Chai*

School of Environmental Science and Engineering, Tianjin University, Tianjin 300354, China

ARTICLE INFO

Editor: A.B. McBratney

Keywords:
Research topic system (RTS)
Co-words network
Soil heavy metal pollution bioremediation (SHMPB)
Scale free network
Small world network

ABSTRACT

From the perspective of dynamic co-words network, for the first time this paper provide a unique flavored critical review on the evolution of the research topic system (RTS) of soil heavy metal pollution bioremediation (SHMPB). We built the dynamic co-words network-related measures of the RTS of SHMPB. Several main innovative findings were obtained. (i) We got the basic characteristics of the co-words network. The co-words network for the RTS had a relative wide and flat structure, which implies the SHMPB's RTS has spread in a broad range. This is differently from the usual radial structure of RTSs in many other scopes. (ii) We revealed the hotspots and development trends of the RTS: the contact degree between different hot fields gradually has increased as the years go by. We found that researchers tended to use plants to carry out SHMPB. The hotspots have shifted from hyperaccumulators to the plants like poplar. (iii) From the degree distribution of the dynamic co-words network, we revealed the evolution of the RTS had the scale free characteristics. The growth and preferential attachment mechanism of keywords were the driving force of the evolving RTS. (iv) By the calculations of the average path length and average clustering coefficients, we found that the evolution of the RTS had the small-world effect. (v) Basing on the obtained results, we gave some implications such as the confirmation of high efficient information transmission, the identification of potential destructive effect, the recognition of important nodes and the creation of new information transmission path within the RTS. Beyond the RTS of SHMPB, we would like to stress the importance of the proposed method, demonstrating its universal applicable merits for RTSs in wide scopes.

1. Introduction

1.1. Background

In recent years, with the rapid growth of industrial technology, heavy metals have been largely used in the world. Accordingly, more and more heavy metals have been released into soil, which has made heavy metal pollution become the primary problem in all kinds of soil pollutions (Calisi et al., 2013). It is well known that heavy metal contamination not only directly reduces soil biological activity and decreases nutrient availability, but also poses a serious threat to human health through entering into food chains and to environmental security by leaching into ground water (Wu et al., 2008). Facing this challenge, the research of the heavy metals contaminated soil remediation has been widely conducted, which mainly includes three aspects: physical, chemical and biological modes. Among them, soil heavy metal pollution bioremediation (SHMPB) has ever kept a research hotspot for due to its advantages of low cost, no secondary pollution and high effectiveness.

At the present stage, SHMPB can be divided into phytoremediation, animal remediation and microbial remediation (Hema et al., 2014). In phytoremediation technology, for example, it was found that hyperaccumulator *Pterisvittata* can extract arsenic, lead, zinc and other heavy metal pollutants from contaminated soil (Brunetti et al., 2012). In animal remediation technology, through experiments, researchers discovered that earthworms have a strong ability of enrichment to lead. With the increase of lead concentration, the amount of lead in earthworms also increases, which further demonstrates the feasibility of introducing earthworms in the animal remediation for heavy metals contaminated soil. In microbial remediation technology, researchers selected the isolated strains from the soil in which the concentration was 10 mmol/L in Cr^{6+}, Zn^{2+}, Pb^{2+}. They found the strain can reduce selenate and selenite into the colloidal Se, transform Pb^{2+} into Pb, and make the colloidal Se and colloidal Pb lost toxicity and be stable in structure (Barton and David, 1992).

Anyway, beyond the publications we can here mention, lots of

* Corresponding author.
E-mail address: lhchai@tju.edu.cn (L. Chai).

http://dx.doi.org/10.1016/j.geoderma.2017.06.019
Received 29 September 2016; Received in revised form 21 November 2016; Accepted 16 June 2017
Available online 26 June 2017
0016-7061/

investigations on SHMPB can be found in literature. However, most researches were mainly on the specific topical analyses of plant, animal and microbial remediation theories or technology respectively. The work of exploring the overall hotspots and development trends of the research topic system (RTS) of SHMPB by consulting a large amount of literature (e.g. bibliometrics) is rather rare. Thus, up to today, we cannot well understand the overall development characteristics of SHMPB's RTS.

Bibliometrics (e.g., literature metrology) is a branch of information science. Because of the arrival of the era of big data, it has been more and more widely applied in the analyses of the overall characteristics of RTSs (Behrens and Luksch, 2011). Bibliometrics can not only reflect the current research status of a RTS by the number counting, but also reveal the structural changes of the RTS behind the numbers by mapping (Borner et al., 2003). Although bibliometrics has been widely used in many fields, we noted that most applications were still mainly concerned with the quantity statistics of literature's information (Takeda et al., 2009). However, it is difficult to uncover the evolutionary dynamic mechanisms of a RTS just by the mere statistics of literature's information.

Recently, considering the keywords in publications can well represent the knowledge contents, knowledge elements, hot spots and focus of a RTS, co-words network of keywords has become a good bibliometric method. The analysis of the co-words network seems hopeful to reveal the evolutionary dynamic mechanisms of a RTS and is accordingly advantageous to forecast the development trends of the RTS, which has prompted its wide theoretical and applicable research in many fields (Choi et al., 2011; Ronda-Pupo and Guerras-Martin, 2012). However, since co-words network, which belongs to a typical complex network, is dynamic, its dynamic analyses have remained hard to be conducted and the associated implications for a RTS have thus kept elusive to be deduced for a long time. Fortunately, the last two upsurges of complex network theories originated from two streams may change this situation. One upsurge is the small-world networks discovered by Watts and Strogatz who described the transition from completely regular network to completely random network (Watts and Strogatz, 1998). They proposed that the small-world networks not only have similar clustering characteristics of regular networks, but also have similar short average path lengths of random networks. The other upsurge is the scale free networks elucidated by Barábasi and Albert who pointed out that the connection degree distributions of some real-world complex networks have power law forms (Barabási and Albert, 1999). This kind of complex networks is thus also called the scale free networks. The dynamic mechanism and laws uncovered by complex network theories have brought the new understanding on the evolution of networks, which has thus prompted many investigations of networks such as Internet (Faloutsos et al., 1999), co-author networks, social networks and many others (Newman, 2001). It is a logical conclusion that by taking the dynamic co-words network as a complex network, we may conduct its dynamic analyses and deduce the associated implications for an evolving RTS under the instructive complex network theories.

1.2. Review purpose

Background review tells us, it is important and possible to use the dynamic co-words network in bibliometrics to review the evolution of a RTS of a subject, and SHMPB is an urgent case of subject that waits to be analyzed.

Thus, with SHMPB's RTS as a case, this paper will for the first time build the dynamic co-words network-related measures to rigorously review the evolution of a RTS, expecting to find the characteristics and the key factors for the development of a RTS. In specific, we will realize several purposes. a) By analyzing the dynamic co-words network, we provide the general characteristics of the network and the RTS. b) We recognize and analyze the hotspot trends in the field of SHMPB. c) Through the analytical calculations of co-words network from complex network theories, we will elucidate the scale free and small world effect for the evolution of SHMPB's RTS in depth. d) We further propose the engineering application implications of the theoretical results. Beyond the specific case of SHMPB, we hope the analytical method of RTS from the dynamic co-words network related measures could find more universal applications in many other RTSs.

2. Review methods

2.1. The dynamic co-words network analysis

Co-words analysis has ever been a fundamental aspect of bibliometrics. Factor analysis, cluster analysis and MDS (Multi-Dimensional Scaling) were once the core foundations of the traditional co-words analysis (Cho, 2014). Now, in this paper, aiming at the analysis of the dynamics and content of an evolving RTS, we have to analyze varied frequencies of keywords and set up the dynamic co-words network. The interrelationships and the frequencies of the keywords, the pertinence of relative contents, the distance between keywords/nodes and the like in the dynamic co-words network have great importance for the evolving RTS. Thus, only from the dynamic co-words network can we confirm and predict the relation between the development and content of the targeted RTS.

The dynamic co-words network can be established and analyzed by Gephi basing on the database (Mathieu et al., 2009). According to the purpose of constructing dynamic network, we create the dynamic matrix by Ucinet (Yang et al., 2014) and visualize it by Netdraw (Takashi, 2010). Using Bibexcel to deal with all the related keywords, we can get a .net file. We can then open this file by Gephi to get a dynamic co-words network that includes lots of varied nodes and edges.

In details, the construction and visualization progress of a dynamic co-words network are conducted by the steps below: 1) run the Force Atlas arithmetic; 2) count the network's Modular Value, and put the nodes in groups; 3) sort the nodes by their degrees and then adjust their sizes for following analyses; 4) adjust the tags of nodes in color, size and font; 5) get a new image. Here, we give a randomly generated network as Fig. 1 for illustrations, which is made by 1000 keywords. There are 8378 edges between these 1000 nodes.

This network as in Fig. 1 can well show the connections between the

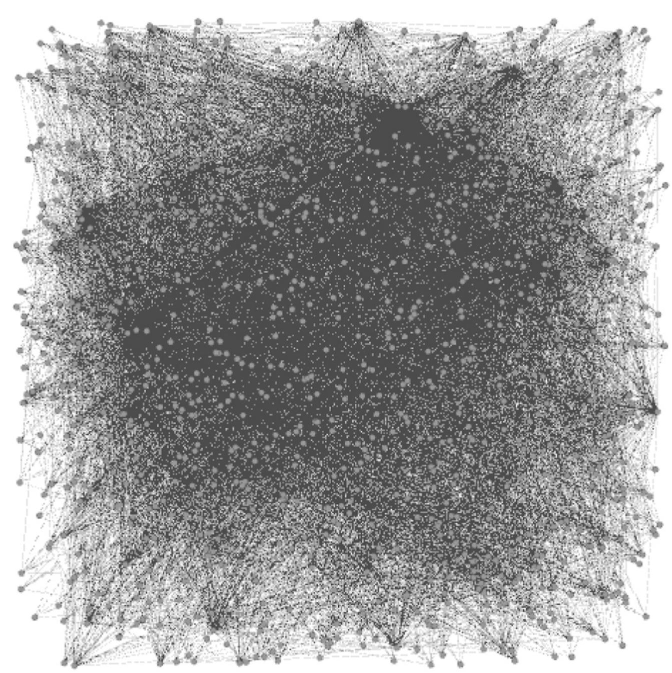

Fig. 1. A randomly generated network for illustrations.

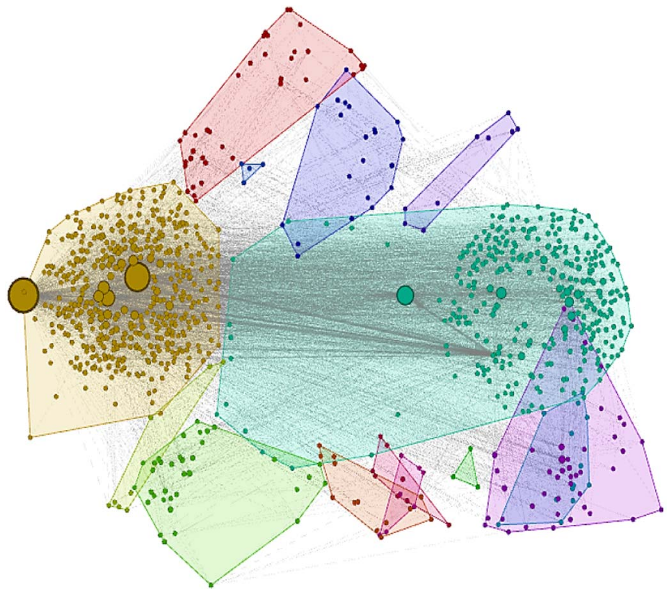

Fig. 2. The analyzable network changed from Fig. 1 by an operated visualization technique.

nodes, but cannot clearly show the degrees or the modularizations of the network. We can change Fig. 1 to an analyzable network as shown in Fig. 2 by an operated visualization technique. Fig. 2 also has 1000 nodes and 8378 edges. Different with Fig. 1, the analyzable network can clearly show us not only different connection degrees between the nodes by different thickness of the edges between these nodes, but different sizes of nodes dependent on their degrees. This can make the network more intuitive and be easier for us to analyze the nodes. In addition, in the analyzable network the nodes are set in groups by modularizations, which could benefit the clearer discovery of hotspots and development trends of RTS.

Through the way of establishing the pertinence of content in the targeted RTS, above dynamic co-words network can shift the content of keywords to the expression of nodes, and visualize the relations of nodes/keywords to the edges in the network. This can help us to quickly grasp the hotspots and trends of RTS of a subject like SHMPB. The future development of the RTS will be reasonably analyzed, predicted and adjusted.

2.2. The introduction of complex network-related measures

To further accomplish the aim of analyzing the dynamic interaction and inter-dependent relationship between inner elements in a RTS, we must have a deep consideration on the topological structure of the dynamic co-words network. Above all of that, we must pay more attention to the inner characters and their function changes in the entire network. In this connection, it would be helpful to discuss the elements and their dynamic developments of the network that could make difference to the forecast of the whole network (Hu et al., 2013).

Complex network theories can well tell us how to decide the type of complex network, and describe the topological structural changes of the entire network from the inner characters and their function changes, and reveal how the dynamic developments of elements in the network could affect the forecast of the whole network (Ronda-Pupo and Guerras-Martin, 2012). In details, under perspective of complex network theories, through the calculations of some network parameters, the type of complex network, which is the basis for further study, can be determined. By analyzing the degree distribution, clustering coefficient, path length and other network parameters, we can determine whether there is a mutual dependence between the hot fields and whether the network has scale-free and small-world characteristics. Furthermore, the practical value of measures in complex networks can also be found. For example, we can analyze the efficiency and effect of information transmission in the network. Through finding out the core nodes, we will analyze the impact of nodes' changes on the whole network, the stability of the network and the creation of new information transmission path within the RTS. In this way, the engineering and management applications implications of the theoretical results can be elaborated as well.

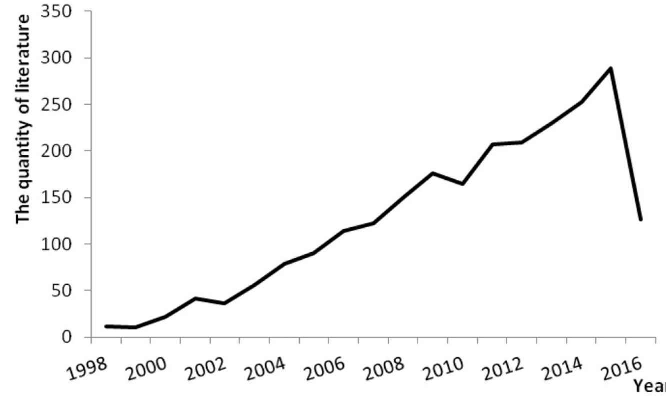

Fig. 3. The overall publication distribution for SHMPB (up to June 2006).

2.3. Data sources from an emerging RTS

We here give data sources of our selected case SHMPB and discuss the emergences characteristics of the interdisciplinary RTS. We select all literature on the SHMPB from the core collection database of Web of Science. The types of literature are article, review and proceedings papers. The retrieval keywords are bioremediation heavy metal contaminated soil and phytoremediation heavy metal contaminated soil with the retrieval time from 1997 to 2016. After the exclusion of some informal papers, a total of 2387 articles were obtained as our data sources, and the time distribution of the literature is shown in Fig. 3.

To analyze where and how the interdisciplinary SHMPB's RTS is emerged, we can discuss in what way the literature of SHMPB appears in different categories. By the statistics of categories, we find that among the publications of SHMPB, the most popular categories are environmental science & ecology, agriculture, plant science, chemistry, geology, engineering. For the distribution of all categories, we establish the relationship between the amount of literature (frequency *F*) and the 149 categories (the rank *r*) by using the generalized Zipf formula

$$F = C(r + m)^{-q}. \quad (1)$$

The results are shown in Fig. 4. Eq. (1) is a kind of hyperbolic distribution, skew distribution or power law distribution. The generalized Zipf formula holds that a small part of low ranks categories share high appearing frequencies, whose appearance implies a self-organized

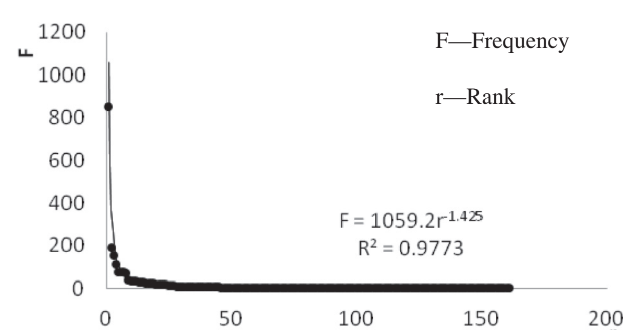

Fig. 4. The generalized Zipf's relationship between the publication output (frequency *F*) and the rank *r* of 149 categories.

dynamic process. As shown in Fig. 4, the occurrence frequency of categories in the field of SHMPB can meet the power law distribution, which shows that the interdisciplinary RTS demonstrates a self-organized emerging process. Among all categories, environmental science & ecology, agriculture, plant science, chemistry, geology and engineering appear with higher frequencies. It can be explained that these top productive categories act as the mainstreams to prompt the emergence of the interdisciplinary RTS. In other words, in the emergence of SHMPB's RTS, the knowledge of these categories is more preferential to be used. Surely, the vigorous use of the knowledge from these top productive categories is the key of promoting the development of SHMPB's RTS.

3. Results and discussions

3.1. Basic evolutionary characteristics of the RTS from the dynamic co-words network

3.1.1. The selection of high-frequency keywords

By the counting of all the keywords in the 2387 papers that are searched by keywords in the data base of Web of Science, we get 4131 original keywords.

To make the keywords as simple as and to avoid the repetition or messes brought up by nonstandard expressions, this paper uses some methods to pre-treat all of the 4131 keywords: (1) combine the keywords that share same meanings but use different descriptions, such as Zinc and Zn, phytoremediation and bioremediation; (2) combine the keywords that totally mean the same things but being singular or plural, such as soils and soil; (3) delete the keywords that have little connections to the topics of a paper, such as place names.

After many times of revision and inspection, we finally choose 2037 keywords. The top 10 high-frequency keywords are shown in Table 1. Using the software, we sort all 2037 keywords by the times they appear in papers, and finally select the first 1000 keywords as high-frequency keywords for our object of study.

3.1.2. The illustration of the matrix for the relationship between edges

Here by picking out the top 50 high-frequency keywords of SHMPB, we establish the keyword co-occurrence networks. We use matrix to describe the phenomenon that a keyword appear in different publications. We can thus judge the correlations between both keywords by the way of analyzing the edges in the network. It is an intuitive method to make survey of the interaction extent and possibility between keywords/nodes. Fig. 5 shows a part of the matrix, which is one of the representative parts of the whole matrix. In order to display the general situation of the matrix of the top 50 keywords, we further make a simple matrix table as Fig. 6.

From Fig. 6, we can make a primary conclusion that from the whole view, the keywords have not strong correlations. But in Fig. 5, we cannot deny that there is a relatively close contact between some keywords. Obviously, the close contact is only limited in some ranges of the entire network. Thus, in the whole view on the RTS of SHMPB, only some research topics are relatively hot while lots of other themes have not high centralities.

Table 1
The top 10 high-frequency keywords.

Order number	High-frequency keywords	Frequency
1	Bioremediation	1105
2	Heavy metal	747
3	Phytoextraction	310
4	Cd	268
5	Pb	168
6	EDTA	126
7	Zn	124
8	Polluted soil	103
9	Phytostabilization	89
10	Soil	85

3.1.3. Basic characteristics of the dynamic co-words network

Basing on the study of the matrix, we take use of Ucinet and Netdraw to visualize the co-words network shown by the matrix. Fig. 7 is the co-words network that devotes convenience to describe the relationships between keywords. Fig. 8 is the network visualization by a part of keywords. From the both, we conclude that: The research topics of the SHMPB's RTS generally present a homogenized tendency, and only some of them link with each other very tightly. In general, most topics stay in average levels and the contents have a broad spread in wide interdisciplinary scientific fields, which highlights the diverse development of the SHMPB's RTS. In addition, from the topological characters of the network, it has an obvious feature of complex network, different from a random network. We can also find, comparing to the usual radial structure shown in the networks in many other scopes, the co-words network has a relative wide and flat structure, which implies the research on SHMPB has spread in a broad range.

3.2. The hotspots and development trends of the RTS

We now discuss the development trends of the RTS from the dynamic co-words network. To get dynamic co-words network at different stages, we need extract and select keywords within a range of time. Here we tend to analyze the keywords within each 3 years. In the table of all information about 2387 papers, we classify the information by years in which the papers appear, and make new tables in term of each 3 years' information. Then we can choose high-frequency keywords. Table 2 and Fig. 9 respectively give the hot topic words and evolutionary co-words network at different stages.

By Table 2 and Fig. 9, we can summarize the change characteristics of the keywords and the evolution of the hot fields of the RTS in each stage as follows:

(1) 1997–1999: The boundary between hot fields is clear and the connection is little. It illustrates the correlation degree between hot fields is not high. In addition, four fields that have no prominent topic words exist, which indicates that this is the initial stage of SHMPB research.
(2) 2000–2002: Although the number of hot fields reduces, the area of every hot field expands and different fields begin to have certain overlapping. It indicates that the research on SHMPB is gradually developing. Through the analysis on the hot topic keywords in this stage, we find the research mainly focuses on the heavy metal elements especially Cd and Cr, which demonstrates Cd and Cr, the earliest research objects, have higher contents in the soil and attract the first attention of researchers.
(3) 2003–2005: The notable feature in this stage is the connections between hot fields are strengthened and the connection degree of different key words increases obviously. By the study of topic words, we find the heavy metal hyperaccumulators become the hotspot of research. In addition, there is a further research on heavy metal elements, which means besides Cd, other elements like Pb have gradually been incorporated into the scope of the study. All of these can illustrate that the research on SHMPB has entered a new stage.
(4) 2006–2009: The overlapping degree of different hot fields is further deepened. It indicates the connections between these hot fields become strengthened, interdependence is enhanced and mutual catalytic relations are more obvious. Besides, heavy metal element Cd is still a hot topic and attracts further study, which reflects that the removing of this element in soil is challenging. Another characteristic in this stage is that EDTA begins to be used in the remediation of heavy metals contaminated soil and has drawn

$$\left\{\begin{array}{cccccc}
\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\
\cdots\cdots & 2 & 4 & 3 & 1 & \cdots\cdots \\
\cdots\cdots & 7 & 26 & 14 & 3 & \cdots\cdots \\
\cdots\cdots & 1 & 13 & 66 & 3 & \cdots\cdots \\
\cdots\cdots & 3 & 43 & 169 & 17 & \cdots\cdots \\
\cdots\cdots & 2 & 10 & 10 & 1 & \cdots\cdots \\
\cdots\cdots & 1 & 9 & 34 & 0 & \cdots\cdots \\
\cdots\cdots & 0 & 0 & 9 & 1 & \cdots\cdots \\
\cdots\cdots & 0 & 3 & 4 & 0 & \cdots\cdots \\
\cdots\cdots & 0 & 17 & 43 & 4 & \cdots\cdots \\
\cdots\cdots & 1 & 5 & 4 & 0 & \cdots\cdots \\
\cdots\cdots & 0 & 2 & 9 & 0 & \cdots\cdots \\
\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots
\end{array}\right\}$$

Fig. 5. A part of the matrix of the top 50 high-frequency keywords.

attention of scholars.

(5) 2010–2012: In this stage, not only the connections between original hot fields become strengthened, but also some new hot fields appear. Combining the analysis of topic words, we find the new study of using phytochelatin to remediate the heavy metals contaminated soil appears. At the same time, we also find the plant of phytoremediation has gradually shifted from hyperaccumulator to other plants like poplar, reflecting control technology of heavy metals contaminated soil develops continually.

(6) 2013–2016: The main characteristic of this stage is that there are more similar keywords and fewer new keywords between hot fields. This phenomenon may be explained that the study on SHMPB has gradually entered the mature stage, which mainly focuses on the development of existing technologies.

Through above analyses, we can preliminarily predict that researchers begin to tend to use plants to carry out the remediation of heavy metals contaminated soil. The hotspots of the study mainly concentrate in choosing plants that can absorb heavy metals in the soil with higher efficiency, and the trend of SHMPB research has gradually shifted from hyperaccumulators to the plants like poplar with fast growth, deep root, large biomass and high transpiration rate.

In addition, through analysis, we find that there is not distinct regularity in the change of the number of hot fields with time. The possible reason is that, with the development of the SHMPB, some hot fields disappear and some new ones appear, leading to a fluctuation in keywords number. In the partitions and overlapping degree, we find before 2013 there are some independent hot fields that have clear boundaries. However, the contact degree between different hot fields has gradually increased as the years go by. This phenomenon can be explained that with the development of SHMPB, new technology and research idea continuously emerge and mutually permeate, which results in the relationship between hot fields becomes increasingly close.

3.3. *The scale-free characteristics of co-words network*

The quantitative studies of the dynamic co-words network based on complex network related measures can yield the more mechanistic understanding on the evolution of RTS. The scale free models of degree distribution obviously are the one of most important aspects of complex network. The scale free models have two remarkable characters as growth and preferential attachment, which are also the necessary conditions to judge whether the degree distribution can be regarded as a robust measure for a network. Here we calculate the degree distribution of the co-words network to analyze its scale free characteristics.

We correlate the degree distribution of the co-words network of keywords by a Zipf's power-law relationship as

$$F \sim I^{-\lambda} \text{ or } D \sim I^{-\lambda}. \tag{2}$$

Here F is appearance frequency of node, I is ID of node degree, and D is node degree. Or in another expression as

High-frequency word	Accumulation	Bioaccumulation	Brassica juncea	Citric acid	EDTA	hyperaccumulator	maize
Accumulation	0	0	0	1	2	2	0
Bioaccumulation	0	0	0	0	0	1	0
Brassica juncea	0	0	0	0	0	1	0
Citric acid	0	0	0	0	4	0	2
EDTA	0	0	1	6	0	2	0
hyperaccumulator	1	0	0	0	0	0	0
maize	0	0	0	0	1	0	0
Pb	1	0	0	1	3	0	1
phytotoxicity	0	0	0	0	0	0	0
Salix	0	0	0	0	1	0	0
soil remediation	0	0	0	2	1	1	0
Trace elements	1	0	0	0	0	0	0
zinc	1	2	0	0	1	3	0
	Pb	phytotoxicity	Salix	soil remediation	Trace elements	zinc	
Accumulation	0	0	1	0	0	2	
Bioaccumulation	0	0	0	1	0	0	
Brassica juncea	0	1	0	0	0	1	
Citric acid	1	0	0	0	0	0	
EDTA	7	1	0	2	0	3	
hyperaccumulator	1	0	0	2	0	4	
maize	1	0	0	1	0	1	
Pb	0	0	0	1	0	0	
phytotoxicity	0	0	0	0	0	0	
Salix	2	0	0	0	0	4	
soil remediation	0	0	0	0	0	0	
Trace elements	0	0	0	1	0	0	
zinc	0	0	0	2	0	0	

Fig. 6. A simple matrix table for the general representation of keywords correlations.

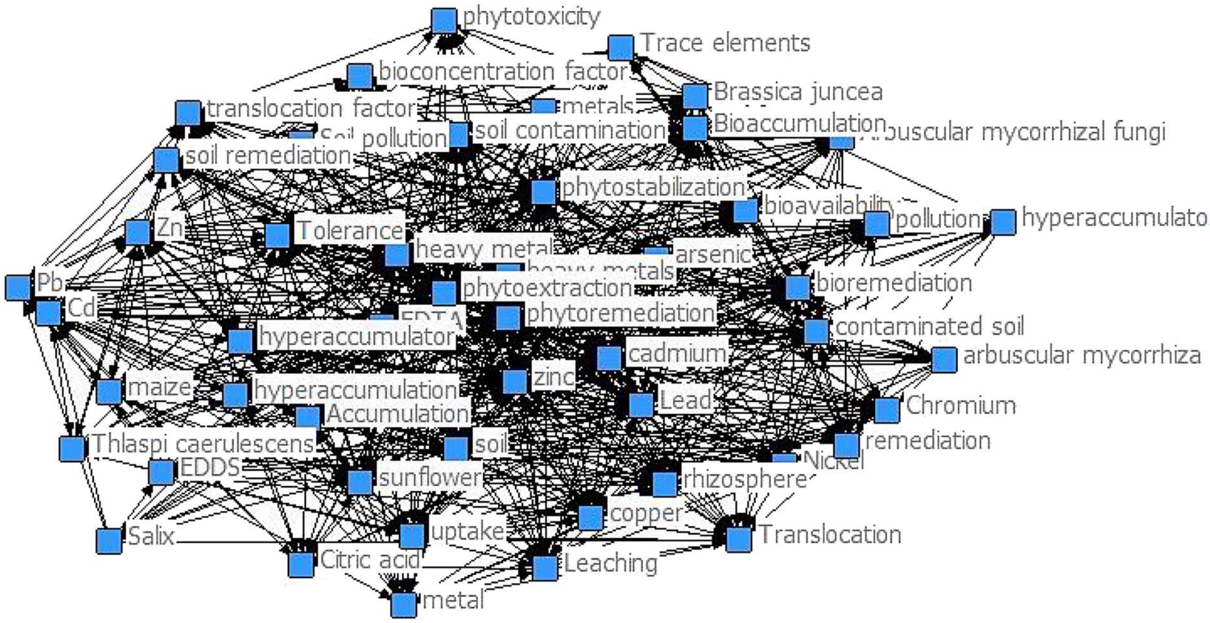

Fig. 7. The co-words network by the entire matrix of keywords visualized by Netdraw.

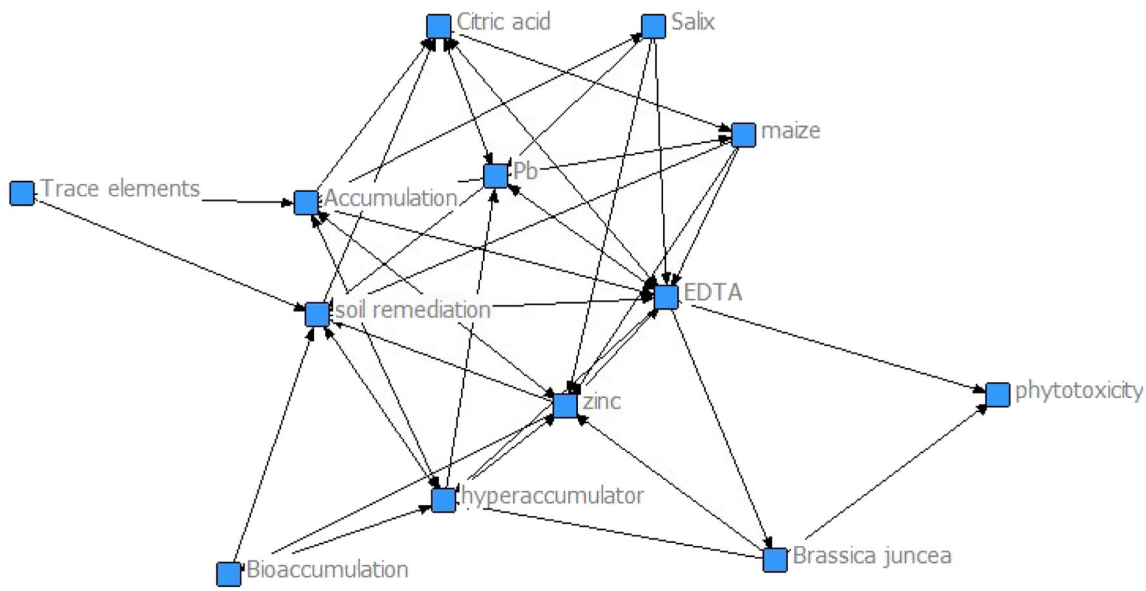

Fig. 8. The co-words network visualization by a part of keywords.

Table 2
Hot topic words at different stages.

Stage	Hot topic words
1997–1999	Heavy metal, chlorophyll fluorescence, soil remediation
2000–2002	Bioremediation, Cd, Cr
2003–2005	Speciation bioremediation, phytoextraction, Cd, hyperaccumulator
2006–2009	Bioremediation, Cd, phytoextraction, polluted soil
2010–2012	Bioremediation, Phytochelatins, Phytoextraction, phytostabilisation, poplar
2013–2016	Bioremediation, phytoextraction

$$P \sim D^{-\gamma}. \tag{3}$$

Here P is node number with corresponding degrees. The empirically correlated degree distributions are shown in Figs. 10–12.

By presenting the general frequency and degrees distributions as shown in Figs. 10–11, it is clear that, among lots of keywords, the evident high-frequency is exactly limited in the several nodes, i.e., the specific topics (Barabási and Oltvai, 2004). In fact, the linear relation with a not large slope (0.879 or 0.794) indicates that the most keywords have not strong connections with others. As shown in Fig. 12, the scale γ is another important parameter for the characteristics of the network. In general, the higher the scale γ is, the stronger influence on the comprehensive network the related keywords have. By the appearance frequency and the degree distributions of the nodes shown in Figs. 10–12, their comparability implies that, for some important nodes in a scale free network, their frequencies and their influences have a significant high-pertinence.

Growth and preferential attachment mechanism are the preconditions for the further analysis on the power-law distribution in a scale free network (Chai and Wen, 2004; Chai and Li, 2007; Todorova and Vogt, 2011). In Figs. 10–12, from the power law function, it is obvious that only some small parts of nodes have the higher degrees. The phenomenon is consistent with the typical characters of scale free network model in two aspects:

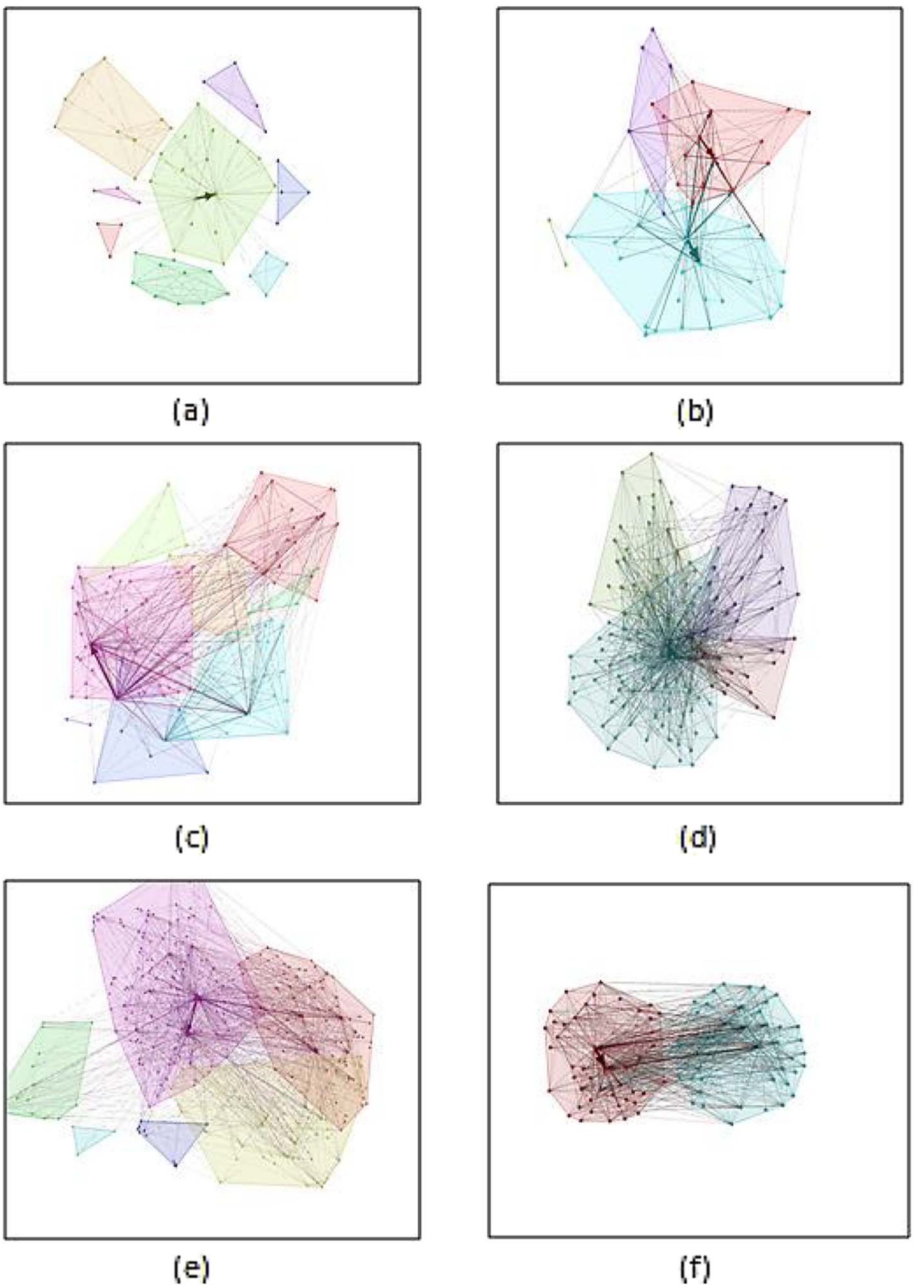

Fig. 9. The patterns of the co-words network in each three years.
(a. 1997–1999, b. 2000–2002, c. 2003–2005, d. 2006–2009, e. 2010–2012, f. 2013–2016).

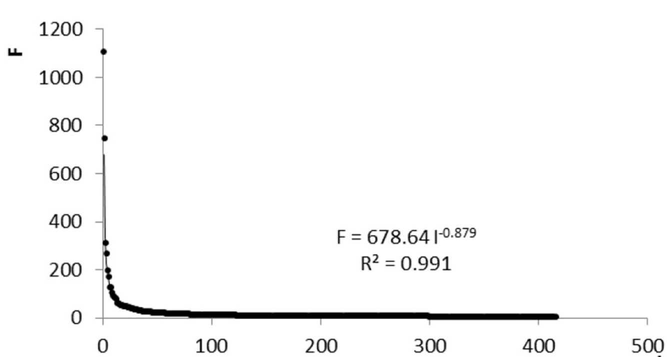

Fig. 10. The correlation between appearance frequency and ID of nodes (1997–2016).

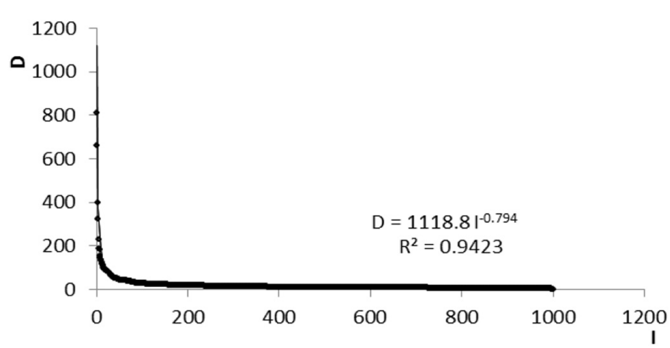

Fig. 11. The correlation between degrees of nodes and ID of nodes (1997–2016).

(1) Growth. Usually most RTSs have the growing tendencies to increase their elemental abundance of knowledge. That is, the topic fields of most RTSs must become more and more extensive with the time goes. All these characteristic would verify that the keywords, as knowledge elements in all RTSs, their numbers usually are in a high growth level (Callon et al., 1983). For the RTS of SHMPB, like other RTSs, the growth of keywords/nodes prompts the development of

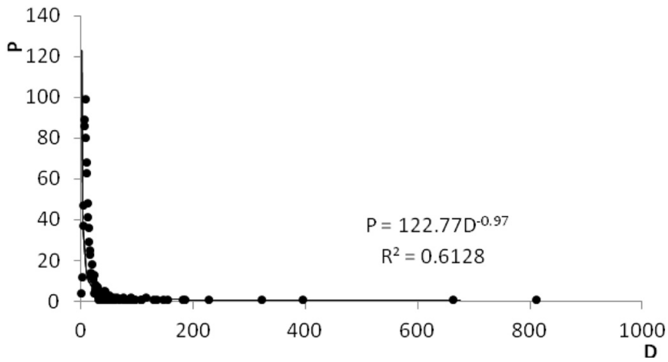

Fig. 12. The correlation between the node number (with corresponding degree) and degree (1997–2016).

the comprehensive co-words network.

(2) Preferential attachment mechanism. For the RTS of SHMPB, like other RTSs, these productive keywords/nodes surely always occupy the more possibility to establish the relation and attachment to other keywords/nodes. Through the developing process, they mostly have the privilege to be members of the new topics or directions. This can be regarded as a logical tendency for the development of any RTSs.

The two characters and associated results in Fig. 10-Fig. 12 provide the evidence that the co-words network of the SHMPB is a scale-free network. The evolution of the co-words network of SHMPB research corresponds to the scale free network model.

In details, in the evolving co-words network of SHMPB, "bioremediation" and "heavy metal" are evidently the highest-frequency keywords whose degrees are the largest. In term of lots of work in this field, we can find some results as follows. Bioremediation is generally divided into two aspects, namely, phytoremediation and microbial remediation. Among them, phytoremediation includes degradation, extraction, volatility, root passivation and fixation, and microbial remediation mechanism mainly includes the bioaccumulation, biological sorption and bioconversion. From the analyses of keywords, it is easily to tell that there are many studies in the field of phytoremediation, which are the hotspots of major topics. Nevertheless, in the study process of phytoremediation, there are also limitations, such as oneness accumulation of heavy metal in hyperaccumulators, the slowness of plant growth, the low efficiency of accumulation, the weakness of adaptability, etc. Therefore, in the evolution process of the SHMPB's RTS, the keywords on common species occur, and they are also gradually becoming major topics. For example, poplar and pine are used for resolving heavy metals pollution (Vaajasaari et al., 2002), which are embodied by the evolution of keywords in the SHMPB. This also reflects that in scientific research, all the fields continuously make innovations. They absorb knowledge. At the same time, they also further create new areas and seek new directions, bringing out more and more treating methods, and continuously exploring more efficient and economic processing methods of soil heavy metal pollution. This is the self-organized evolution of co-words network, and also is the development and progress in the SHMPB's RTS, highlighting the mechanisms for the generation of scale free distribution of node degrees.

3.4. The small world characteristics of co-words network

Small-world is another characteristic of complex network. To analyze the small world characteristics of co-words network, we introduce two measures.

One is a measure of the global properties of the network: the average path length. It is defined as the number of edges in the shortest path between two nodes averaged over all pairs of nodes (Zhang et al., 2006b). If $d(i, j)$ denotes the shortest distance between a pair of nodes i and j, the total distance $\sigma(N)$ is expressed by

$$\sigma(N) = \sum_{1 \le i < j \le N} d_{ij}. \tag{4}$$

The average path length $L(N)$ is then defined as

$$L(N) = \frac{2\sigma(N)}{N(N-1)}. \tag{5}$$

The other is a local measure: the average clustering coefficient. It describes the cliquishness properties of a generic node. Clustering coefficient C_i of a generic node is the number of edges that exist in the clique of a generic node (e.g., the number of neighbors) divided by the most probable number of edges in the clique. C_i is used as a description of local feature of network, measuring whether there are subsystems which are relative stable. Namely, clustering coefficient C_i can be regarded as a parameter that is used to describe neighboring nodes' relation degree (Zhang et al., 2006a). Generally, we assume a node i in the network has k_i edges connecting it to other nodes, these k_i nodes are what we called the node's neighbors. Apparently, among these k_i nodes there are at most $k_i(k_i - 1)/2$ edges, and the ratio of the number of actual edges E_i among these k_i nodes to all of the possible edges number $k_i(k_i - 1)/2$ is what we defined as node's clustering coefficient (C_i)

$$C_i = \frac{2E_i}{k_i(k_i-1)}. \tag{6}$$

The average clustering coefficient C of the whole network is then the average of every node's clustering coefficient

$$C = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{7}$$

Obviously $0 \le C \le 1$. Only when all the nodes have no connections to each other, $C = 0$, namely, there are not edges at all in the network. Only when all the nodes have connections to every other nodes, $C = 1$. When N is very large, $C = O(N^{-1})$, when $N \to \infty$, $C = O(1)$. Many large networks have obvious clustering effects, although their clustering coefficient is far more < 1 but far larger than $O(N^{-1})$, which means these actual networks are not totally random but in some ways they have the non-random feature similar to the those in the social world "birds of a feather flock together". Of course, C and L (defined as the simple arithmetic mean of d_{ij}) can be applied only to connected graphs and cannot be used for cliques subgraphs, in which most of the cases are disconnected (Marchiori and Latora, 2000).

Though there is not an acceptable definition of small-world phenomenon, it is generally considered that small-world feature is mainly measured by two parameters, average clustering coefficient and average path length. If a network has average clustering coefficient far larger than that of its corresponding random network, but has average path length similar to (usually slightly smaller than) that of the corresponding random network, then we can say that this network has a small-world phenomenon (Comellas et al., 2000). Thus, there is a specific judgment of small-world effect.

For comparisons, we select keywords in different periods. According to the number of nodes and edges, the probability of connection can be calculated and a comparable random network in every three years can then be produced. In details, in the Gephi, by selecting "file- generation-random graph" and inputting the quantity of nodes and the probability of connection, which are required for each network, the comparable random network that has the same number of nodes and edges as the studied co-words network can be generated as shown in Fig. 13. Table 3 gives the basic parameters of the generated network at different stages.

We compare the statistical parameters of the two networks in Table 4. It can be seen that the studied co-words network has far more larger average clustering coefficient than the comparable random network while two networks have similar average path lengths, which can verify the co-words network has an obvious small-world effect. The

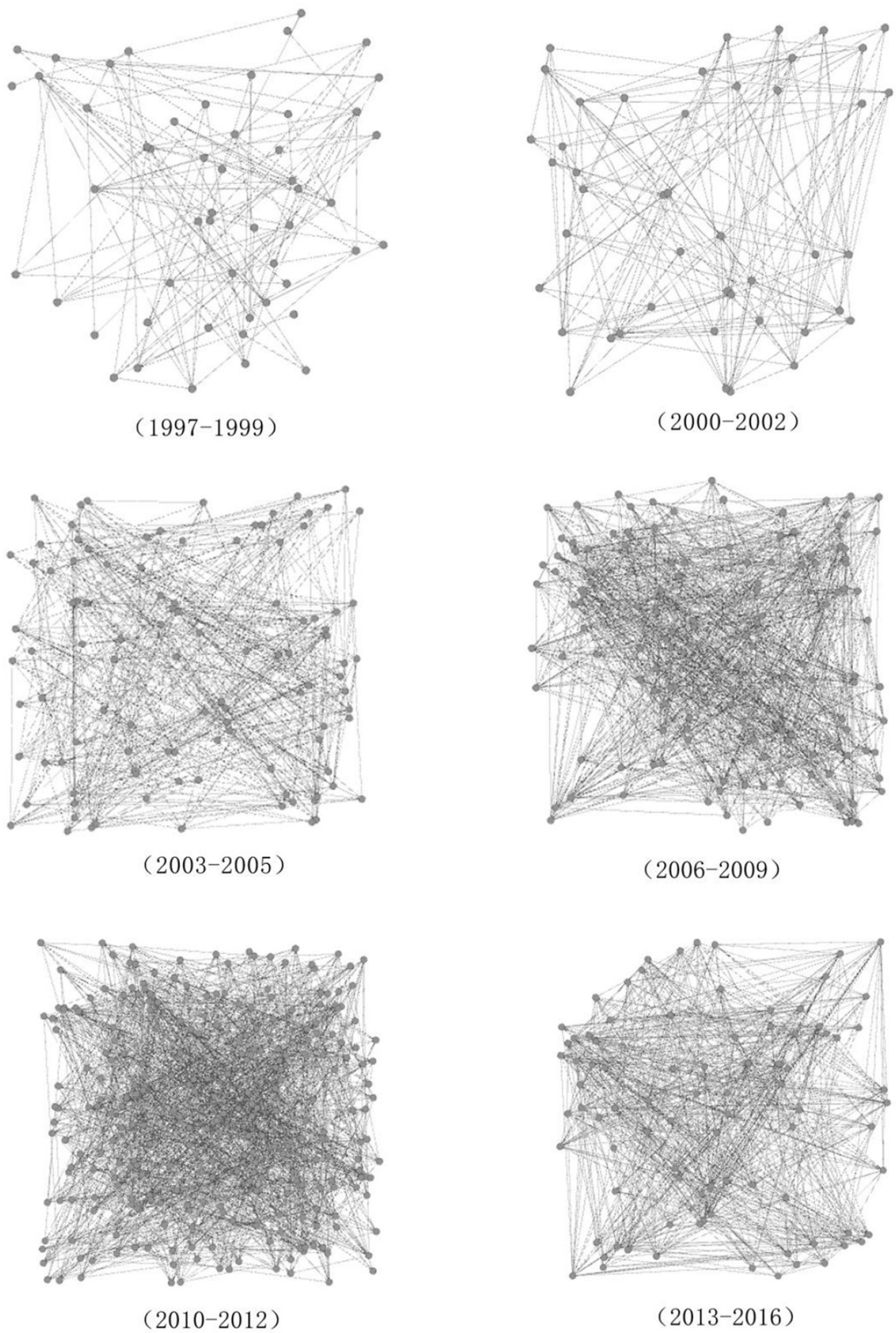

Fig. 13. The comparable random network in every three years.

Table 3
The basic parameters of comparable random network at different stages.

Stages	The number of nodes	The number of sides	Probability
1997–1999	53	136	0.09869
2000–2002	45	189	0.19091
2003–2005	108	495	0.08567
2006–2009	137	1000	0.10735
2010–2012	285	1528	0.03777
2013–2016	89	737	0.18822

small-world effect means the contact between the nodes is very close and the information between different the nodes can spread very quickly along the co-words network of SHMPB's RTS.

Table 4
The comparison of the average path length and the average clustering coefficient between co-words network and comparable random network.

Network	Nodes	Total edges	L	C
Co-words network	1000	8378	3.193	0.299
Comparable random network	1000	8378	3.448	0.008

3.5. The implications in engineering applications

Under the perspective of complex network related measures, the results of the dynamic co-words network for the evolution of RTS bring a new flavor, which should be embodied not only in academy but in management engineering. We now discuss the implications of the results in engineering applications.

3.5.1. The importance of high efficient information transmission within RTS

For small-world network, the clustering effect between each field is stronger than one can even perceive. If a network is a random network, the average path length would be longer and the average clustering coefficient would be smaller, which makes the information transmission in the network not be smooth or interruption phenomena often occur. Therefore, information cannot be transmitted in time or be fed back from one field to other related fields.

As for the co-words network of SHMPB, its small-world effect means the spread of information in various fields of the RTS is rapid and the hot areas can efficiently promote the development of new emerging areas. In addition, the communication barriers between different fields of the RTS are low and each field has a strong ability to absorb the knowledge from other fields. By quickly absorbing the knowledge into the field, new knowledge is created and the knowledge diffusion cycle works. In this way, with the high efficient information transmission, mutual links and cooperation between different fields are established, which promotes the rapid diffusion of the research results and innovated knowledge between researchers. At the same time, due to the complexity and multi-dimension of the small-world network and the non-exclusive nature of knowledge diffusion, the diffused information along the co-words network is not easy to disappear or decrease. All these characteristics are important for SHMPB's RTS (of course, other RTSs have similar implications). Here we just give a typical example, during the evolution of SHMPB's RTS, poplar trees in one clique, which take over hyperaccumulator plants in another completely different clique, have gradually become the new method to solve the heavy metal pollution with the change of time. This is a case of sufficiently utilizing small world effect. For any RTSs, making good use of small world phenomena, we can effectively make knowledge discovery, innovation and prediction in the evolving RTSs.

3.5.2. The potential destructive effect within RTS

Co-words network is largely different from random network. The comparisons of the co-words network with the same scaled random network are as in Table 5. We can see that, though the difference of the average degrees is small, they have largely different biggest degrees. The studied co-words network's degrees of key nodes are much higher than the random network, which indicates that the co-words network has better connectivity, and the connections between the key nodes and other nodes are closer. This also means that with the gradual fading out of some nodes and/or entering a new and immature research field of some nodes, some particular nodes that have strong connections with the other nodes have prevailed in the co-words network.

In other words, the co-words network holds uneven distribution of degrees. It is clear that when shocks come (e.g., the non-natural decline or disappearance of some special key areas with large degrees), the transmission of a key part of information would be interrupted and the network would be greatly affected. Due to the rapidity and continuity of information transmission, when a key field in the network is influenced by shocks, the impact would quickly spread to other areas. This is similar to destructive shocks in economical networks, for example, the disillusion of the real estate bubble in the United States in 1926 triggered indirectly and rapidly the world crisis in the 1930s. In 2007, the subprime crisis in the United States became deeper in just one year and spread around the globe, causing the world financial “tsunami”.

Therefore, it is necessary to recognize the core nodes from the co-words network of a RTS. By regulating and controlling their developments, we can improve the robustness of the network, enhance its anti-interference ability and adjust the spread speed of interference in the whole network. By evaluating the importance of the nodes to identify “core nodes”, on one hand, the protection of these “core nodes” can improve the reliability of the whole network; on the other hand, the attack of “core nodes” can achieve the purpose of destroying the entire network. In most cases, the most meaningful thing is to adjust the connection between key nodes to stabilize or protect the co-words network to realize the sustainable development of a RTS. Here, focusing on SHMPB's RTS, we give a typical example. During the evolution of SHMPB's RTS, poplar tree instead of hyper accumulator plants acts as the new method of solving the heavy metal pollution create new important links between key nodes, which has largely stabilized the co-words network to prompt the sustainable development of the RTS. In most RTSs, we should make good use of this kind of effects to stabilize the co-words networks and achieve the better development of the RTSs.

Table 5
The comparisons of node degrees for co-words network and the same scaled random network.

Network	Number of the nodes	Number of edges	Maximum degree	Minimum degree	Average degree
Co-words network	1000	8378	812	1	8.751
The same scaled random network	1000	8378	35	5	8.364

3.5.3. The recognition of important nodes within RTS

The connection degree of the nodes (e.g., the number of node connection edges) in a network can be regarded as a measure of node's importance. It is thought that the more the connection edges of nodes have, the bigger their degrees are, and the more important the nodes are. This method of evaluating the important of nodes is very simple. According to the theory of social network, for RTSs, the keywords in publications are fundamental knowledge elements. The connections between the keywords/nodes are important aspects of an evolving RTS. The number of node connection edges, i.e., the connection degree of the nodes, can primarily identify the importance of the research topics of a RTS. If the node degree is high, it illustrates that there are many direct connections between it and many other nodes, then we can regard the node as a “core node” that has plenty of information can be used or transported and plays an important role in the process of the information transmission in the whole network. In this connection, we can hold that the node degree is the most simple and reasonable measure of node's importance. Of course, one may argue that there is certain one-sidedness when using simple node degree to determine “core nodes”, for some significant “core nodes” do not necessarily has larger connection degrees. At the same time, some high-frequency keywords that have few connections may also disturb the judgment of “core nodes”. Thus, some scholars brought up some other methods to help to determine the “core node” (Kim et al., 2016). The betweenness degree of a node is another measure. It is, indeed, a measure of the importance of the node in the network, and in the recent years has been used intensively for network analysis (Giorgio et al., 2013). However, in most cases, considering that betweenness degree is more complicate in operation and lacks some intuitions, here we would to select the node's degrees as the measure of keywords' importance. The top 10 keywords with high degrees in SHMPB's RTS are shown in Table 6.

In our studied case of co-words network as shown in Table 6, the top 10 high frequencies keywords are also the top 10 high degrees

Table 6
The top 10 keywords with high degrees.

Order	Keyword	Degree	Order	Keyword	Degree
1	Bioremediation	812	6	Zn	185
2	Heavy Metal	662	7	EDTA	181
3	Phytoextraction	396	8	Polluted soil	155
4	Cd	322	9	Phytostabilization	147
5	Pb	228	10	Soil	134

keywords, which also show that these nodes can well act as the “core nodes” in the network. In addition, according to the Table 6 (and Table 5), we can see that though the difference of the average degree between the co-words network and the same scaled random network is small, the difference of the biggest degree is large. The co-words network has better connectivity. In a word, the measure of node degree can well divide the co-words network from the same scaled random network. This again shows that the node degree is reasonable but simple method for identifying important nodes.

3.5.4. The creation of the new information transmission path within RTS

In a complex network, the clustering characteristics of nodes mean some weak links between two clusters are often very vital for the global topology (Zhang et al., 2006a). In the network, one can usually reduce the average path length by adding more edges between two cliques to create the new links. Of course, due to some economical and technical restrictions, the added edge between nodes (or cliques) must be controlled by some acceptable conditions. The new links are actually the new information transmission path. By adding new information transmission paths, we can reduce the path length between the two nodes (or cliques), strengthen the connection between two fields, facilitate information transmission along the network and ultimately promote the development of a RTS. Available results show that it is easily possible to construct networks with a higher clustering index and a lower diameter by deterministic techniques (Comellas and Sampels, 2002). The key is to determine which path between nodes can be added to be most effective and meaningful in a real complex network.

As for the RTS of SHMPB, the creation of the new links is important and its method is also clear. For example, seen from the co-words network, the keywords about microorganism are limited. This field still belongs to the progressive development stage, but the microorganism can respond quickly to the environmental changes and adapt themselves to environmental conditions. Changes in the population or activity of microorganisms are easier to be tested than the physical and chemical properties of soil, which means the changes in microorganism can be considered as an early sign of improvement in soil quality or an early warning of soil degradation (Pankhurst et al., 1995). Therefore, microorganism can break some limitations of phytoremediation and develop hopefully into a new discipline. However, at the same time, problems such as microbial species antagonism effects are also a significant limitation of this field. If so, we can create the new links between the two cliques, that is, we can try to combine the phytoremediation and microbial remediation. By using the symbiosis or synergy between plants and microbes, we can largely improve the bioremediation effectiveness of soil heavy metal pollution. Of course, this is just a case in SHMPB's RTS. In many other RTSs, using the small world characteristics to create the new links (i.e., new information transmission path) between the nodes (or cliques) is important to discover new hotspots and realize the innovative development of the RTSs.

4. Concluding remarks

This paper for the first time provides a review on the evolution of research topic system (RTS) of soil heavy metal pollution bioremediation (SHMPB) based on a unique flavored avenue (i.e., the dynamic co-words network). We here summarize several main innovative findings:

(1) We analyzed the basic characteristics of the dynamic co-words network. Different from the usual radial structure of RTSs in many other scopes, the dynamic co-words network for the RTS had a relative wide and flat structure, which implies the SHMPB's RTS has spread in a broad range. (2) We elaborated the hotspots and development trends of the RTS: the contact degree between different hot fields gradually increases as the years go by. We found that researchers tended to use plants to carry out SHMPB. The hotspots have shifted from hyperaccumulators to the plants like poplar. (3) By calculating the degree distribution, we revealed the evolution of the RTS corresponds to the scale free network. The growth and preferential attachment mechanism of keywords prompts the development of the RTS. (4) We calculated the average path length and average clustering coefficients to find that the evolution of the RTS has an obvious small-world effect. (5) With the obtained results, we elaborated some engineering application implications such as the confirmation of high efficient information transmission, the identification of potential destructive effect, the recognition of important nodes and the creation of new information transmission path within the RTS. In this way, we would like to show the paper not only is instructive in academy, but has important engineering applicable merits.

Currently, the whole world has come to realize the environmental problems have impacted the human being terribly. Above all, since the heavy metals soil pollution is so close to our life, our food, our water, its research should urgently develop, not only from the theories or technique related to specific topics, but also from the comprehensive investigation method. Our bibliometric investigative reviews on the evolution of SHMPB's RTS over time basing on the dynamic co-words network related measures can help researchers and government to know the development mechanisms of the global SHMPB's RTS. This can offer a foundation for the further regulation and optimization of the global SHMPB's RTS. This is the merit of the studied case.

Of course, beyond the case of SHMPB's RTS, our method and findings may be universal and have important meanings in many other areas. By introducing varied edges to describe the interactions between keywords and constructing a series of dynamic co-words network-related measures, the evolution and engineering implications of any other RTSs can be similarly revealed or deduced. Thus, though the present paper just gives a simple case study of SHMPB's RTS, the method and findings can be universally applied to many other RTSs. As a model owning a mathematical-physical foundation, we hope the present framework could readily be extended and applied to many other specific RTSs in other scopes in the future.

References

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.

Barabási, A.L., Oltvai, Z.N., 2004. Network biology: understanding the cell's function organization. Nat. Rev. Genet. 5 (2), 101–113.

Barton, L.L., David, A.S., 1992. Transport and Remediation of Subsurface Contaminants. American Chemical Society, Washington DC, pp. 99–107.

Behrens, H., Luksch, P., 2011. Mathematics 1868–2008: a bibliometric analysis. Scientometrics 86 (1), 179–194.

Borner, K., Chen, C., Boyack, K.W., 2003. Visualizing knowledge domains. Ann. Rev. Inf. Sci. Technol. 37 (1), 179–255.

Brunetti, G., Farrag, K., Soler-Rovira, P., et al., 2012. The effect of compost and *Bacillus licheniformis*, on the phytoextraction of Cr, Cu, Pb and Zn by three brassicaceae species from contaminated soils in the Apulia region, Southern Italy. Geoderma 170 (1), 322–330.

Calisi, A., Zaccarelli, N., Lionetto, M.G., 2013. Integrated biomarker analysis in the earthworm Lumbricus terrestris: application to the monitoring of soil heavy metal pollution. Chemosphere 90 (11), 2637–2644.

Callon, M., Courtial, J.P., Turner, W., et al., 1983. From translation to problematic networks: an introduction to co-word analysis. Soc. Sci. Inf. 22 (2), 191–235.

Chai, L.H., Li, H.B., 2007. A new theoretical analysis on organizing principles of water supply networks. J. Water Supply 56 (4), 233–244.

Chai, L.H., Wen, D.S., 2004. Hierarchical self-organization of complex systems. Chem. Res. Chin. Univ. 20 (4), 440–445.

Cho, J., 2014. Intellectual structure of the institutional repository field: a co-word analysis. J. Inf. Sci. 40 (3), 386–397.

Choi, J.H., Yi, S.Y., Lee, K.C., 2011. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. Inf. Manag. 48 (8), 371–381.

Comellas, F., Sampels, M., 2002. Deterministic small-world networks. Physica A 309 (1–2), 231–235.

Comellas, F., Ozon, J., Peter, J.G., 2000. Deterministic small-world communication networks. Inf. Process. Lett. 76 (1–2), 83–90.

Faloutsos, M., Faloutsos, P., Faloutsos, C., 1999. On power-law relationships of the internet topology. Comput. Commun. Rev. 29 (4), 251–260.

Giorgio, A., Donatella, F., Luigi, L., 2013. The (betweenness) centrality of critical nodes and network cores. In: International Wireless Communications and Mobile Computing Conference, pp. 90–95.

Hema, T.G., Getha, K., Tan, G.Y., et al., 2014. Actinobacteria isolates from tin tailings and forest soil for bioremediation of heavy metals. J. Trop. For. Sci. 26 (1), 153–162.

Hu, L.D., Jin, N.D., Gao, Z.K., 2013. Characterization of horizontal gas-liquid two-phase

flow using Markov model-based complex network. Int. J. Mod. Phys. C 24 (24), 957–967.
Kim, C.M., Kim, Y.H., Han, Y.H., et al., 2016. Efficient estimation of betweenness centrality in wireless networks. Mob. Net. Appl. 21 (3), 469–481.
Marchiori, M., Latora, V., 2000. Harmony in the small-world. Physica A 285 (3–4), 539–546.
Mathieu, B., Sebastien, H., Mathieu, J., 2009. Gephi: An Open Source Software for Exploring and Manipulating Network. In: International AAAI Conference on Weblogs and Social Media (ICWSM 2009) 3rd, San Jose, California, USA.
Newman, M.E.J., 2001. The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. U. S. A. 98 (2), 404–409.
Pankhurst, C.E., Hawke, B.G., McDonald, H.J., et al., 1995. Evaluation of soil biological properties as potential bioindicators of soil health. Aust. J. Exp. Agric. 35 (7), 1015–1028.
Ronda-Pupo, G.A., Guerras-Martin, L.A., 2012. Dynamics of the evolution of the strategy concept 1962–2008: a co-word analysis. Strateg. Manag. J. 33 (2), 162–188.
Takashi, Y., 2010. A Draw Plug-in for a Wild Software. In: 10th Annual International Symposium on Applications and the Internet (SAINT 2010) 10th, Seoul, South Korea.
Takeda, Y., Mae, S., Kajikawa, Y., et al., 2009. Nanobiotechnology as an emerging research domain from nanotechnology: a bibliometric approach. Scientometrics 80 (1), 23–38.
Todorova, L., Vogt, B., 2011. Power law distribution in high frequency financial data: an econometric analysis. Physica A 390 (23–24), 4433–4444.
Vaajasaari, K., Joutti, A., Schultz, S., et al., 2002. Comparisons of terrestrial and aquatic bioassays for oil-contaminated soil toxicity. J. Soils Sediments 2 (4), 194–202.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-world network. Nature 393 (6684), 440–442.
Wu, C., Wu, J., Luo, Y., et al., 2008. Statistical and geoestatistical characterization of heavy metal concentrations in a contaminated area taking into account soil map units. Geoderma 144 (1–2), 171–179.
Yang, H., Li, Z.F., Wei, W., 2014. Instance analysis of social network based Ucinet tool. Inf. Technol. J. 13 (8), 1532–1539.
Zhang, Z.Z., Rong, L.L., Comellas, F., 2006a. A high-dimensional random apollonian networks. Physica A 364 (C), 610–618.
Zhang, Z.Z., Rong, L.L., Guo, C.H., 2006b. A deterministic small-world network created by edge iterations. Physica A 363 (2), 567–572.