Journal of Environmental Informatics 38(2) 116-130 (2021)

Journal of Environmental Informatics

www.iseis.org/jei

Comprehensive Study of Evolution of Global Environmental Quality Research Using Informetric Co-Word Network

Z. La[1] and L. H. Chai[1 *]

[1] *School of Environmental Science and Engineering, Tianjin University, Tianjin 300354, China*

Received 21 October 2016; revised 02 May 2018; accepted 12 November 2019; published online 24 March 2021

ABSTRACT. Global environmental quality is a rapidly developing and complicated subject in need of a new method for evolutionary analyses. This paper proposes an informetric dynamic co-word network to study the evolution of global environmental quality research (GEQR), making a significant contribution to the available literature. First, it was observed that GEQR has been vigorously developing and hotspots have emerged through a self-organized adjustment process, which other methods lacked. Second, small-world and scale-free effects, which are mechanisms of an evolving knowledge system (KS), were identified in GEQR's KS. Third, the dynamic co-word network yielded topological patterns and robust clustering in GEQR. An assessment strategy map/table of GEQR has been accordingly proposed as a mined innovative function of co-word networks, which provides new holistic understanding of global environmental quality. Thus, our studies on GEQR evolution using this informetric dynamic co-word network method may help researchers or managers learn GEQR development mechanisms, allowing for further topical selection and policy management of GEQR. Beyond GEQR, the informetric dynamic co-word network method may also provide a new universal method for evolutionary analysis of many other environmental problems.

Keywords: Global environmental quality research (GEQR), informetric, dynamic co-word network, complex network, environmental quality assessment

1. Introduction

Global environmental quality has important impacts on human life. In recent years, although enormous progresses on global environmental quality research (GEQR) has progressed enormously in recent years, most available research is specific to environmental, ecological, technical, economic, and systematic analyses of individual topics. Regarding singular aspects of environmental quality assessment, there have been investigations on soil (Vrscaj et al., 2008), climate change conditions (Penenko and Tsvetova, 2014), action of metals and pesticides (Singh et al., 2004), and uses of aquatic plants (Ferrat et al., 2003). For multi-factor environmental quality assessment that includes wide factors, it was found that the studies have been categorized by country (Larsson and Hanberger, 2016), city (Robati et al., 2015), ocean (Gao et al., 2014), indoor environments (Kolokotsa and Santamouris, 2015), and so on. These studies have used remote sensing, Geographic Information Systems (GIS), Analytical Hierarchy Process (AHP) and Support Vector Machine (SVM) models to evaluate environmental quality (Karimipour et al., 2005; Xiong et al., 2007; Wang et al., 2015).

In addition to specific research, some citation analyses and bibliometric research on global environmental assessment were performed to provide an overall view of GEQR (Khan and Ho, 2012; Li and Zhao, 2015). However, because using GEQR as a knowledge system (KS) for microscopic topical evolution (note: study here mainly refers to KS founding on the topical contents) requires rigorous analysis and meets many difficulties, academia is lack of the associated search, which causes us to be short of important policy making on global environmental quality. The main reason is intricacy of KS research, which relates to a large scale of published data as well as new information methods or techniques. Fortunately, with current improvements to data storage capacity, enhanced computer processing capacity, and establishment of many large-scale databases, an opportunity is ripe to investigate whether an appropriate information method can be developed to facilitate evolution of a KS.

An appropriate information method must first be identified. As early as the 1980s, Brookes (1980) aimed to combine human cognitive processes to concisely describe trends in concept development and proposed cognitive mapping to reveal the structure of KS. Cognitive mapping is "the relationship between concepts that creates the KS structure" (Brookes, 1980). Generally, concepts can be reflected by authors, nations, citations, and keywords in scholarly publications. Keywords summarize the detailed topics in a given publication and are better concepts for KS than other macroscopic parameters such as author names, citations, or nation of origin. It is more appropriate to use the co-occurrence of keywords (i.e., co-word) when constructing a

* Corresponding author. Tel.: 86-22-87402072; Fax: 86-22-87402076.
E-mail address: lhchai@tju.edu.cn (L. H. Chai).

SSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.202100449

KS to define the relationships between concepts. In this way, Callon et al. (1983) proposed a well-known informetric co-word analysis. The advantages of co-word analysis have also informed a large number of methodological applications in the field of software engineering (Coulter et al., 1998), information retrieval (Ding et al., 2001), higher education and science (Dehdarirad et al., 2014). In co-word analysis, the general process includes keyword extraction, high-frequency word selection, co-word matrix construction, and knowledge map drawings, and the most frequently used techniques are multi-dimensional scaling, clustering analysis, and strategy plotting.

Co-word networks, also called complex networks, usually relate to co-occurrences of keywords at large scales, and are complicated and dynamic (Newman, 2003). Complex network theories have been applied to analyze evolution within a co-word network. For example, by comparing and discussing co-citation and word networks, Yi and Choi (2012) identified structural features, evolutionary origin and meanings of co-word networks. Some researchers analyzed growth patterns of scientific knowledge from the evolutionary perspective of the co-word network (Wang, 2009; Luo et al., 2017; Zhang et al., 2017). Nevertheless, these only a few preliminary studies on the co-word network from the evolutionary perspective of complex networks. Scientists are still far from defining the evolutionary mechanisms of a co-word network with dynamic features, and the underlying details of the KS are still unknown.

Like other co-word networks, GEQR is likely a dynamic co-word network and a typical complex network (Newman, 2003; Wang, 2009; Yi and Choi, 2012; Luo et al., 2017; Zhang et al., 2017). With the aim of pioneering beyond available co-word research, this paper aims to develop an informetric dynamic co-word network to analyze the evolution of GEQR's KS, mainly revealing the effects of self-organization, small-world and scale-free. Self-organization refers to a system spontaneously arranges its components or elements in a purposeful manner under appropriate conditions but without the help of an external force. Small-world effects mean the link phenomenon within a network with short node path lengths, i.e. any two nodes tend to be linked through a small chain of nodes of nodes. Scale-scale effects indicate that the fraction of network nodes with degree follows a power law, i.e., a self-similar pattern free of scale. The informetric dynamic co-word network method has the following advantages and innovations above available co-word network analyses. (1) By screening the frequencies of dynamic keywords and links at different stages, the dynamic co-word network is constructed to rigorously identify a self-organized process over time, a functionality that is lacking in existing methods. (2) Although small-world and scale-free effects have been widely studied with other systems, these effects are deeply analyzed in the dynamic co-word network, to reveal underlying meanings to the KS of GEQR. Analysis of small-world and scale-free effects, which are often overlooked in existing methods, is therefore an important aspect of our research. (3) Through robust clustering of GEQR, the KS emerges from the dynamic co-word network, allowing for a novel assessment strategy map/table to be proposed, which is a function not found in existing methods.

The layout of this paper is as follows. The method is introduced in section 2, and Section 3 provides results and discussions that are mainly divided into three parts. The first part gives the basic evolutionary characterization, hotspot identification, and especially self-organization of GEQR by the dynamic visualization of the co-word network. The second part analyzes more underlying evolutionary mechanisms of GEQR by calculating and discussing the measures of the dynamic co-word network such as small-world and scale-free effects. The third part provides an assessment strategy map/table of GEQR based on the clustering of the dynamic co-word network. In last section, the main conclusions are given.

2. Method: Informetric Dynamic Co-Word Network

This study established a co-word network by the co-occurrence of keywords in the same publications, in which the network nodes (N) are keywords and the link sides (M) between the nodes are the two keyword appearances in the same publication. The size of the nodes (N) indicate the keyword appearance frequency, and the side thicknesses (M) are the link frequencies between two keywords, as shown in Figure 1. Both nodes and links are dynamic in an evolving KS, resulting in a dynamic co-word network. Our method framework on the dynamic co-word network is illustrated in Figure 2.

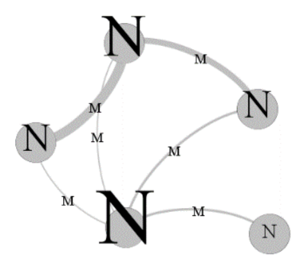

Figure 1. A schematic construction of co-word network.

Step 1. For environmental quality, TS = "environment* quality" is selected as the search term to retrieve data from the 1998 ~ 2015 Scientific Citation Index (SCI), Scientific Citation Index Extension (SCI-E), Social Science Citation Index (SSCI) and Proceedings Citation Index-Science (CPCI-S) in the WoS database. A total of 6,819 publications have been retrieved, of which 5,482 articles have keywords. For a publication, in addition to keywords provided by authors, the WoS editorial system may also provide some keywords for a publication; this is called "keyword plus". These additional keywords are added to improve the ability of keywords to reflect the contents of the publiccations. 5,064 of the articles have additional keywords. After pre-treating keywords with +s, +ing, +ed, and abbreviations to identify all forms, a total of 22,904 keywords were identified. After selecting the typical data, Bibexcel software was used to re-check the data. Data reduc ions are usually needed for better display and more accurate data analysis when working at such a large scale.

Step 2. In the Gephi software, information is classified by year in the table of publication information. High-frequency key-words can be chosen for each year and specific keywords within a given time frame can be extracted to provide dynamic keyword frequencies and links from different stages. By entering this information to Gephi, an informetric dynamic co-word network that tracks changes over time can finally be realized. Analyzing gradual development from original small groups to large groups and gradual emergence of major hot-spots during the evolutionary process allows for analysis of the self-organized effects of the GEQR topical system. The innovative method presented here can thus provide dynamic self-organized evolutionary results (i.e., Figure 3 and Table 1 in Section 3.1) that existing methods lack.

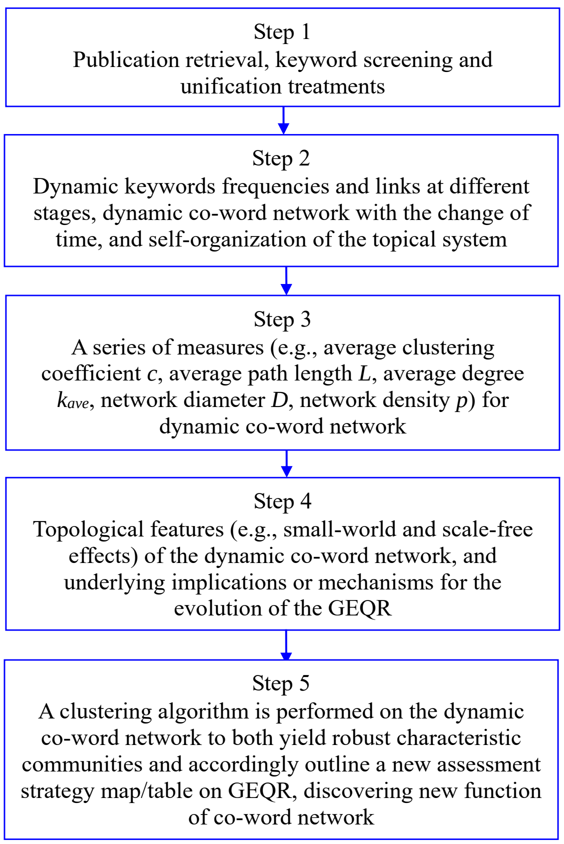

Figure 2. An illustration of method framework.

Step 3. The dynamic co-word perspective for KS is based on the network structure and its evolution, rather than on a single keyword. The dynamic co-word network must next be quantitatively described to understand the quantitative implications for KS. This paper addresses the constructed informetric dynamic co-word network that evolves through interactions between coupled keywords; a series of concepts (i.e., clustering coefficient, path length, average degree, network diameter, and network density) are introduced and calculated as follows (Newman, 2003). The clustering coefficient of a co-word network denotes shared connections by topic and aggregate features of a KS. The larger the clustering coefficient, the stronger the connections between various research areas in the KS. The clustering coefficient c_i of node i is expressed by:

$$c_i = \frac{2E_i}{k_i(k_i - 1)} \tag{1}$$

Here, k_i is the degree of node i and E_i is the number of its neighbors. The average clustering coefficient c of the entire network of all nodes is:

$$c = \frac{1}{N}\sum_{i=1}^{N} c_i \tag{2}$$

N is the number of all nodes. The path length between nodes in a co-word network represents the semantic distance between topics. Generally, shorter distance indicates stronger semantic correlation between topics and thus increased opportunities for keyword co-occurrence. The average path length L is defined as the average path length d_{ij} between any keyword pairs in the co-word network:

$$L = \frac{2}{N(N+1)}\sum_{i \geq j} d_{ij} \tag{3}$$

In a co-word keyword, higher-degree keywords represent more important topics than others in the KS. The average degree, k_{ave}, is defined as the average of k_i (number of adjacent edges or degree of keyword i) in the co-word network:

$$k_{ave} = \frac{1}{N}\sum_{i=1}^{N} k_i \tag{4}$$

Additionally, the co-word network diameter, D, is the maximum path length between any two keywords (i.e., $D = \max_{1 \leq i, j \leq N} d_{ij}$) and the co-word network density, p, is the ratio of the actual number of edges to the possible maximum number of edges (i.e., $p = 2M / N(N - 1)$). Here M is the actual edge's number in the co-word network. The density of the co-word network density reflects keyword co-occurrence opportunities or KS maturity.

Step 4. Path length, clustering coefficient and degree distribution are used to identify the topological features of the co-word network (i.e., small-world and scale-free effects). Specifically, the small-world effect can determined by average path length L and average clustering coefficient c of a co-word network. In a same-scaled random network with the same number of nodes and edges as the co-word network can be considered, the average path length L_g and average clustering coefficient c_g can be calculated. If $L \leq L_g$ and $c > c_g$ are satisfied, a co-word network with parameters L and c is considered to have a small-world effect (Watts and Strogatz, 1998). Addiionally, the node degree distribution $P(k)$ of the entire co-word network is analyzed. If the distribution approximately follows power law functions, i.e.:

$$P(k) \sim k^{-r} \quad (r \text{ is the scale exponent}) \tag{5}$$

Such a power function is scale invariant and the co-word network has scale-free effects (Caldarelli, 2007). In this step,

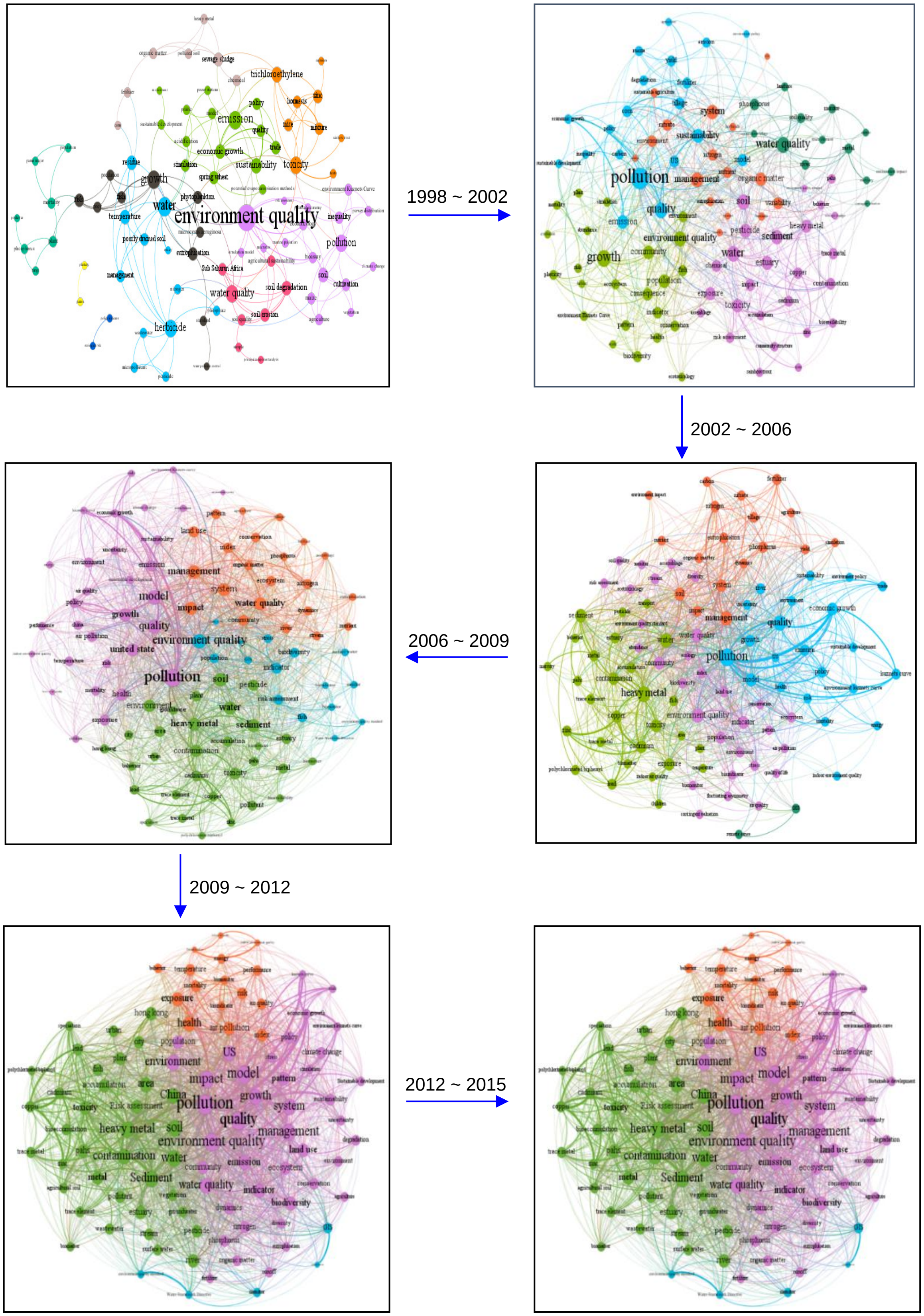

Figure 3. Self-organized evolution of dynamic co-word network by top 100 keywords.

Table 1. The Comparative Analyses of the Top 100 Keywords in Different Periods

1998 keywords	1998: 1998 ~ 2002	1998 ~ 2002: 1998 ~ 2006	1998 ~ 2006: 1998 ~ 2009	1998 ~ 2009: 1998 ~ 2012	1998 ~ 2012: 1998 ~ 2015	1998 ~ 2015 keywords
Environment quality	0	0	0	0	0	Environment quality
Water quality	0	0	0	0	0	Pollution
Toxicity	0	0	0	0	0	Heavy metal
Growth	0	0	0	0	0	Model
Eater	0	0	0	0	0	Sediment
Economic growth	0	0	0	0	0	Management
Sustainable development	0	0	0	0	0	Water quality
Emission	0	0	0	0	0	Quality
Pollution	0	0	0	0	0	Soil
Soil degradation	0	0	0	0	0	Impact
Mortality	0	0	0	0	0	Water
Soil	0	0	0	0	0	China
Sample	1	0	0	0	0	System
Sustainability	1	0	0	0	0	Contamination
Organic matter	0	0	0	0	0	Health
Herbicide	0	0	0	0	0	Indoor environment quality
Management	0	0	0	0	0	Growth
Environment Kuznets curve	0	0	0	0	0	PAHs
Quality	1	0	0	0	0	Indicator
Soil quality	0	0	0	0	0	Economic growth
Spring wheat	0	0	0	0	0	Policy
Pattern	0	0	0	0	0	Performance
Pesticide	0	0	0	0	0	Environment
Phosphate	0	0	0	0	0	Air pollution
Phosphorus	1	0	0	0	0	Emission
Corn	0	0	0	0	0	Exposure
Hormesis	1	0	0	0	0	Metal
Temperature	1	0	0	0	0	Toxicity
Sub-Saharan Africa	0	0	0	0	0	Sustainability
Chemical ecology	1	0	0	0	0	GIS
Heavy metal	1	0	0	0	0	Risk assessment
Chemical	1	0	0	0	0	Area
Parus major	0	0	0	0	0	Environment
Standard	0	0	0	0	0	United States
Cadmium	1	0	1	0	0	Land use
Tests	1	0	0	0	0	Cadmium
Climate change	1	0	0	0	0	Biodiversity
Ecological risk	1	0	0	0	0	Climate change
Estuary	0	0	0	0	0	Community
Risk assessment	1	0	0	0	0	Environment Kuznets curve
Principal component analysis	0	0	0	0	0	Estuary
Eutrophication	0	0	0	0	0	City
Rusle	1	0	0	0	0	Air quality
Environment quality standard	1	0	0	0	0	Sustainable development
Rainbow trout	0	0	0	0	0	River
Environment	0	0	0	0	0	Nitrogen
Risk	0	1	0	0	0	Thermal comfort
Ecotoxicology	1	0	0	0	0	Conservation
Residue	1	1	0	0	0	Energy
Cultivation	1	1	0	0	0	Trace metal
Plant	1	0	0	0	0	Fish

Continues

1998 keywords	1998: 1998 ~ 2002	1998 ~ 2002: 1998 ~ 2006	1998 ~ 2006: 1998 ~ 2009	1998 ~ 2009: 1998 ~ 2012	1998 ~ 2012: 1998 ~ 2015	1998 ~ 2015 keywords
Simulation model	0	0	0	0	0	Copper
Fungi	1	1	0	0	0	Wastewater
Soil erosion	1	0	0	0	0	Biomarker
Phytoplankton	0	0	0	0	0	Population
Sewage sludge	1	0	0	0	0	Eutrophication
Fertilizer	1	0	1	0	0	Temperature
Predation	1	1	0	0	0	Surface water
Simulation	1	1	0	0	0	Lead
Policy	1	0	0	0	0	Accumulation
Fish	1	1	0	0	0	Pesticide
Bioassay	1	0	0	0	0	Organic matter
Wastewater	0	0	0	0	0	Plant
Urban sustainability	1	0	0	0	0	Trace element
Acidification	1	0	0	0	0	Environment quality standard
Rrade	0	0	1	0	0	Kuznets curve
Rrichloroethylene	0	0	1	0	0	Indoor air quality
Agriculture	1	0	0	0	0	Index
Agricultural sustainability	1	1	1	0	0	Diversity
Inequality	0	1	0	0	0	Biomonitor
Water pollution control	1	1	0	0	0	Pattern
Model	0	0	1	0	0	Agriculture
Nitrogen	1	0	0	0	0	Behavior
Mixture	0	0	0	0	0	Phosphorus
Maize	1	1	0	0	0	Water framework directive
Vegetation	1	1	0	0	0	Risk
Zinc	0	0	0	0	0	Ecosystem
Prairie	1	0	1	0	0	Dynamics
Meuse	0	0	0	0	0	Agricultural soil
Power stations	0	1	0	0	0	Monitor
Precursor	1	0	0	0	0	Zinc
Po Delta	0	1	0	0	0	Groundwater
Metal bioavailability in soil	1	1	1	1	1	Perception
Plasticity	1	0	0	0	0	Polychlorinated biphenyl
Prey	0	1	0	1	1	Satisfaction
Primate population	0	1	0	0	0	Uncertainty
Microtox	1	0	0	0	0	Degradation
Power distribution	1	1	0	0	0	Bioindicator
Microcystis aeruginosa	1	1	0	0	0	Stream
Mice	0	0	1	0	0	Urban
Micropollutant	0	1	0	0	0	Simulation
Political economy	1	0	1	0	0	Hong Kong
Polluted soil	1	1	0	0	1	Comfort
Policy indicator	1	0	0	0	0	Speciation
Potential evapotranspiration methods	0	1	1	1	0	Pollutant
Powdery mildew	0	0	1	1	0	Bioaccumulation
Postlarvae	1	1	0	1	0	Fertilizer
Poorly drained soil	0	1	0	1	0	Stress
Population	1	0	0	1	0	Remote sense
Marine pollution	1	0	1	0	1	Productivity

Note: 0: same; 1: different

Table 2. The Dynamic Co-word Network Measures Calculated in Different Periods

Year	*N*	*M*	*C*	*L*	k_{ave}	*D*	*p*
1998	613	2073	0.454	2.18	6.763	7	0.011
1998 ~ 2002	3151	14327	0.437	3.717	9.094	11	0.003
1998 ~ 2006	7191	41193	0.427	3.778	11.457	11	0.002
1998 ~ 2009	11662	74533	0.421	3.741	12.782	12	0.001
1998 ~ 2012	16813	116333	0.419	3.702	13.838	13	0.001
1998 ~ 2015	22904	177960	0.446	3.93	15.54	12	0.001

N: node; *M*: edge; ***c***: average clustering coefficient; *L*: average path length; k_{ave}: average degree; *D*: network diameter; *p*: network density.

Table 3. The Six Clusters for Strategy Table of GEQR

	Keywords	Perspectives	Cases
Cluster 1	Management; Water quality; Impact; System; Indicator; Community; Land use; River; Biodiversity; Dynamics; Organic matter; Diversity; Nitrogen; Pattern; Forest.	Management	Assessment objects: River Important reference: Water quality; Indicator; Impact; System; Management; Land use; Community; Dynamics; Pattern; Conservation; Biodiversity; Bay; Basin etc.
Cluster 2	Pollution; Quality; Model; Growth; Environment; Air pollution; Emission; United States; Policy; Sustainability; Climate change; Risk; City; Degradation; Sustainable development.	Pollution	Assessment objects: Air pollution Important reference: Environment; Growth; Quality; Model; Emission; City; Policy; United States; Risk; Climate change; Sustainability; Environment Kuznets curve etc.
Cluster 3	Health; Exposure; Performance; Temperature; Indoor environment quality; Air quality; Perception; Indoor air quality; Energy; Behavior; Productivity; Climate; Children; Simulation; Thermal comfort.	Health	Assessment objects: Indoor environment quality Important reference: Exposure; Health; Temperature; Air quality; Performance; Energy; Children; Satisfaction; Behavior; Perception etc.; Indoor air quality; Particulate matter etc.
Cluster 4	Sediment; Water; PAHs; Pesticide; Fish; Monitor; estuary; Biomarker; Surface water; Oxidative stress; Polychlorinated biphenyl; Contaminant; Estuary; Sea; Monitor.	Water	Assessment objects: Water Important reference: Sediment; Metal; Pesticide; Fish; Estuary; Sea; PAHs; Biomarker; Contaminant; Pollutant; Groundwater; Sample; Environment quality standard; Polychlorinated biphenyl etc.
Cluster 5	Heavy metal; Soil; Contamination; Cadmium; Toxicity; Risk assessment; Wastewater; Trace metal; Copper; China; Area; Plant; Lake; Transport; Availability.	Heavy metal	Assessment objects: Soil Important reference: Heavy metal; Contamination; Toxicity; Copper; Waste water; Risk assessment; Zinc; Area; Plant; Mercury; Lead; Removal; China; Accumulation; Lake etc.
Cluster 6	Population; Evolution; Egg size; Selection; Fluctuating asymmetry; Body size; Natural selection; Consequence; Fitness; Life history; Reproductive success; Birds; Escherichia coil; Survival; Trade off.	Evolution	Assessment objects: Species Important reference: Population; Evolution; Reproductive success; Egg size; Life history traits; Trade off; Fitness; Survival; Sexual selection; Parasite; Brood size etc.

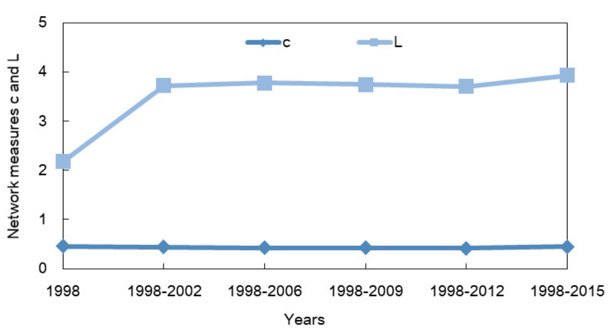

Figure 4. The change of average clustering coefficient and average path length.

the innovation of the present method is apparent in the implications for dynamic co-word network evolution and GEQR's KS is revealed through small-world and scale-free mechanisms (as discussed in Section 3.2), which are usually neglected by existing methods.

Step 5. The topological pattern of the dynamic co-word network, which emerges from self-organized interactions between factors/keywords, has robust characteristic communities. Keywords are closely linked within communities, while links between communities are weak. The GEQR cluster is dependent on a set of factors/keywords with close connections and is neither determined by a single factor/keyword nor defined by a group of unrelated factors/keywords. Thus, the topological patterns generated by a set of factors/keywords with close connections show close clustering on GEQR. When a clustering algorithm is performed on the dynamic co-word network, a new assessment strategy map/table is created in GEQR. Derivation of this assessment strategy map/table (Table 3, Section 3.3) is a mined innovative function of the co-word network that is not found in existing methods. Understanding the structural features of the KS of GEQR from a microscopic knowledge unit (keyword) allows for a better description of the KS. Thus, a more comprehensive assessment strategy map/table for GEQR can be identified.

The underlying evolutionary mechanisms of the KS are quantitatively revealed by the informetric dynamic co-word network through the five steps discussed above, allowing for the development of a useful informetric dynamic co-word network method for knowledge discovery, identification predicttions and assessment strategy maps/tables of GEQR.

3. Results and Discussions

3.1. The Self-Organized Mechanism of the Dynamic Co-Word Network

Dynamic topical evolution reflects development trends of a KS, an important area of study. Because of data source and computation restrictions, the 1998-2015 timeframe was selected. The sample is theoretically complete and can be extended before 1998 and after 2015. Since a KS tends to change signifycantly over several years and the study resource is limited, we selected an interval of four years to determine the average KS development. Figure 3 shows the visualized configuretions of the dynamic co-word network for 1998, 1998 ~ 2002, 1998 ~ 2006, 1998 ~ 2009, 1998 ~ 2012 and 1998 ~ 2015, derived by selecting the top 100 nodes. It is apparent in Figure 3 that as the number of individual nodes (keywords) increases, the connections between nodes in the dynamic co-word network gradually increase as well. Thus, the dynamic evolution of the GEQR involves gradual conversion of many small groups of topics into several large collective topics. From topologies identified by different colors, this gradual development from small to large groups is visually represented by the different-colored topologies. Certain keywords (i.e., pollution, heavy metal, model, sediment, management, water quality, soil, impact, environment, contamination, and the likes) emerge as prominent topics in the dynamic co-word network, implying that with the addition of new keywords and stronger internal interactions, stable major hotspots such as pollution, heavy metal, model, sediment, management, and water quality have emerged by a self-organized process. As such, GEQR hotspots and the self-organized mechanism of their emergence are recognized by the dynamic co-word network.

Table 1 compares keyword changes by time period based on the identified top 100 keywords for each period. Between 1998 and 1998 ~ 2002, 50 of the top 100 keywords changed, while 24 of the top 100 keywords changed from 1998 ~ 2002 to 1998 ~ 2006. Sequential comparative analyses between 1998 ~ 2009 and 1998 ~ 2015 showed that only 4 of the top 100 keywords had changed. These hotspots have dominated the self-organized dynamic process of developing the GEQR's KS stably. Analyses of the reduction in detailed keyword changes among the top 100 keywords are consistent with the results shown in Figure 3, which clearly shows the dynamic formations of hot trends in GEQR by self-organized adjustments. This is the first quantitative evidence that self-organized dynamic mechanisms in GEQR have formed a mature KS with current hotspots. By demonstrating how the ordered trends of GEQR (i.e., pollution, heavy metal, model, sediment, management, water quality, etc.) emerge from initial disorder, our results are instructive for scientists with different research interests.

3.2. Small-World and Scale-Free Effects for GEQR Evolution

To identify additional underlying evolutionary mechanisms of a KS from the dynamic co-word network, it is necessary to build on the calculations for measuring the entire dynamic network. Here, we calculate the primary measurements related to dynamic characteristics, which include the average clustering coefficient c_i, average path length L, average degree k_{ave}, network diameter D, and network density p.

Measurements of the dynamic co-word network were conducted using Gephi, and the results are shown in Table 2. It is apparent from Table 2 that N nodes and M edges increase over time in the GEQR co-word network. New keywords are constantly added, and old keywords repeat as well. Additionally, growing edges include new relationships between old keywords, relationships between old and new keywords, and relationships among new keywords. The clustering coefficient c hovers around 0.02 with time. The average path length L, average degree k_{ave} and networks diameter D are generally increasing, causing density p to decrease and then stabilize. All measurements indicate that the entire network is developing and GEQR is in a growth stage. New terms and methods are constantly added, resulting in the diverse development of KS. Various research topics within the KS are closely related and environmental quality has been clearly defined over time in GEQR. The following sections discuss small-world effects and scale-free characteristics of the KS in terms of the calculated network measurements.

3.2.1. Small-World Effects

Figure 4 shows the average path length and clustering coefficient of GEQR's dynamic co-word network. As seen from Figure 4, GEQR's co-word network has an average path length of L = 3.7 and clustering coefficient c = 0.424. The average path length and clustering coefficient of the corresponding random network are L_g = 5.463 and c_g = 0.001. Since $L \leq L_g$ and $c > c_g$ are satisfied, the GEQR co-word network has small-world effects.

The evolutionary paths of specific topics throughout the network demonstrate the small-world effect. In the case of geographic information systems (GIS), the keyword is extracted to identify its dynamic appearance frequencies and links, and the local dynamic co-word network can be obtained as shown in Figure 5 by retaining all original data except some necessary unifications (i.e., abbreviation, singular/plural forms, dash, upper-/lower-case). In 1998, the keyword "GIS" was not within the top 100 keywords, while by 2002, GIS was among the top 100 keywords and was therefore connected to the keyword "system". By 2006, GIS had an improved ranking and an increasing number of nodes connected to it. In 2009, 2012 and 2015, as the keyword "GIS" increased in frequency, so did the node connections. The development path of this specific example illustrates how the node "GIS" can build links between surrounding nodes, shortening the distances within the co-word network. This small-world effect shows how, since its appearance in the 1960s, GIS has become one of the most popular hotspots for rapid development in environmental quality and other research. The current GIS has firmly occupied the hot field of GEQR. The evolution of GIS illustrates the small-world mechanism in the KS, which can be applied to other cases in the dynamic co-word network. The development paths of many specific topics with small-world mechanisms has prompted the entire KS of GEQR to evolve towards an interdisciplinary field focusing on hot trends such as pollution, heavy metals, modeling, sediment, management, and water quality that relate to the lithosphere, hydrosphere, atmosphere, biosphere, and anthroposphere.

The small-world mechanistic results have practical uses in showing how GEQR can be associated with other knowledge to multiply connections within the network. Information is transmitted at high speeds this type of small-world knowledge network; by adjusting a small number of connections, the performance of the whole network can be changed drastically. Thus, by changing a few node links, one can likely significantly adjust and improve the overall KS structure in the existing GEQR network. The small-world effect may also be a universal characteristic applicable to other KSs beyond GEQR, and therefore it is important that we improve information transmission paths, find new links and prompt innovative development of KSs (Luo et al., 2017; Zhang et al., 2017).

3.2.2. Scale-Free Characteristics

Figure 6 shows the calculated degree distributions of keywords for 1998, 1998 ~ 2002, 1998 ~ 2006, 1998 ~ 2009, 1998 ~ 2012 and 1998 ~ 2015 in the dynamic co-word network of GEQR, which are r = 1.428, 1.583, 1.522, 1.483, 1.438 and 1.369, respectively. For all degree distributions, fitting performances and robust changes of r exponents around an approximate 1.5 value indicate a KS with scale-free effects. In the dynamic co-word network for GEQR's KS, most studies are linked to only a small number of the hottest topics. In other words, a few research hotspots with large degrees comprise the prominent research centers, linking numerous KS studies and marking the formation of core themes, ideas, and schools within the KS. The scale-free effects indicate that the evolution of GEQR's KS is driven by nodal growth and preferential connection.

To intuitively illustrate the scale-free mechanism from the dynamic co-word network, it is useful to analyze the evolution

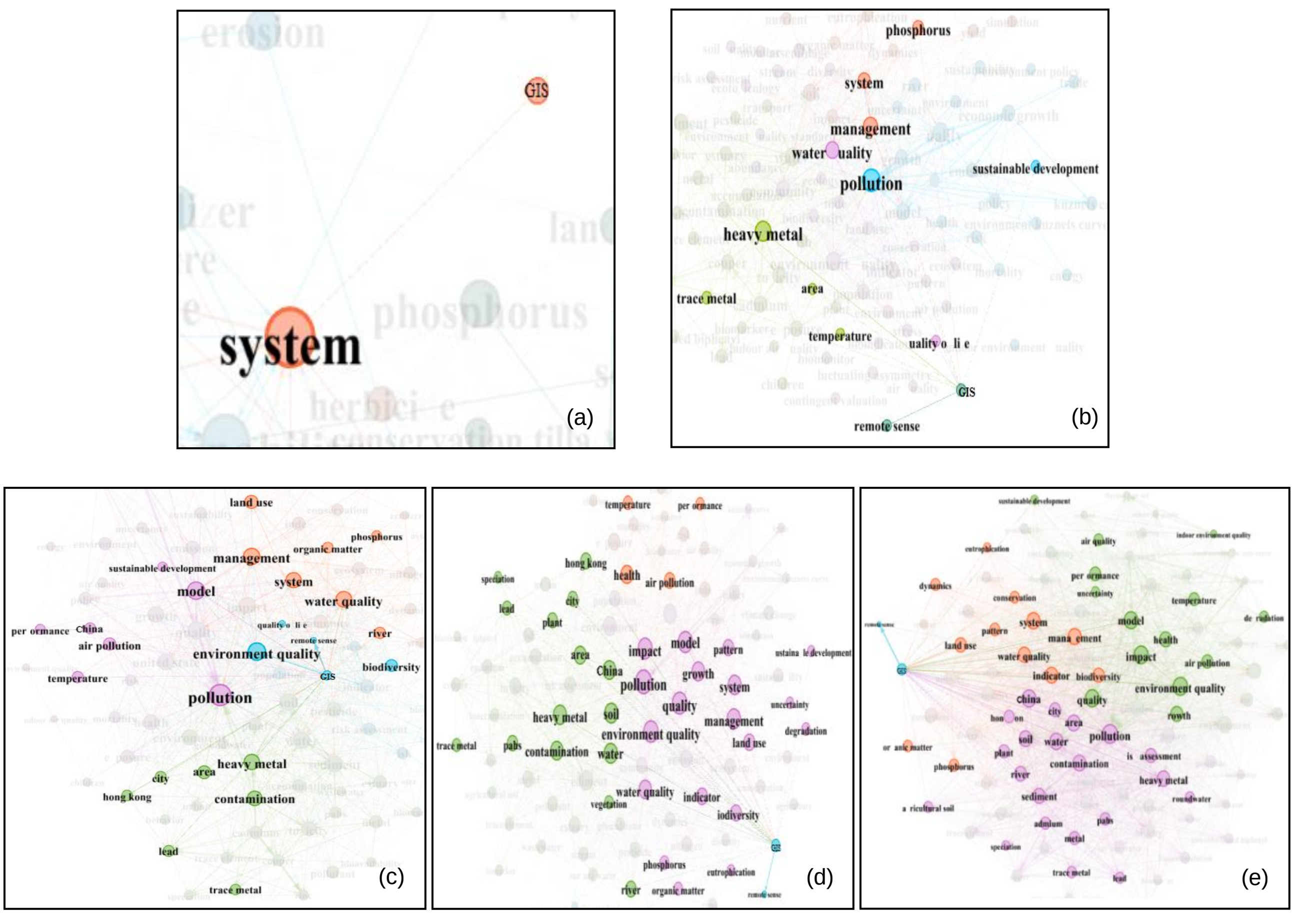

Figure 5. The dynamic evolutionary path of GIS in the top 100 hotspots: (a) 1998 ~ 2002, (b) 1998 ~ 2006, (c) 1998 ~ 2009, (d) 1998 ~ 2012, and (e) 1998 ~ 2015.

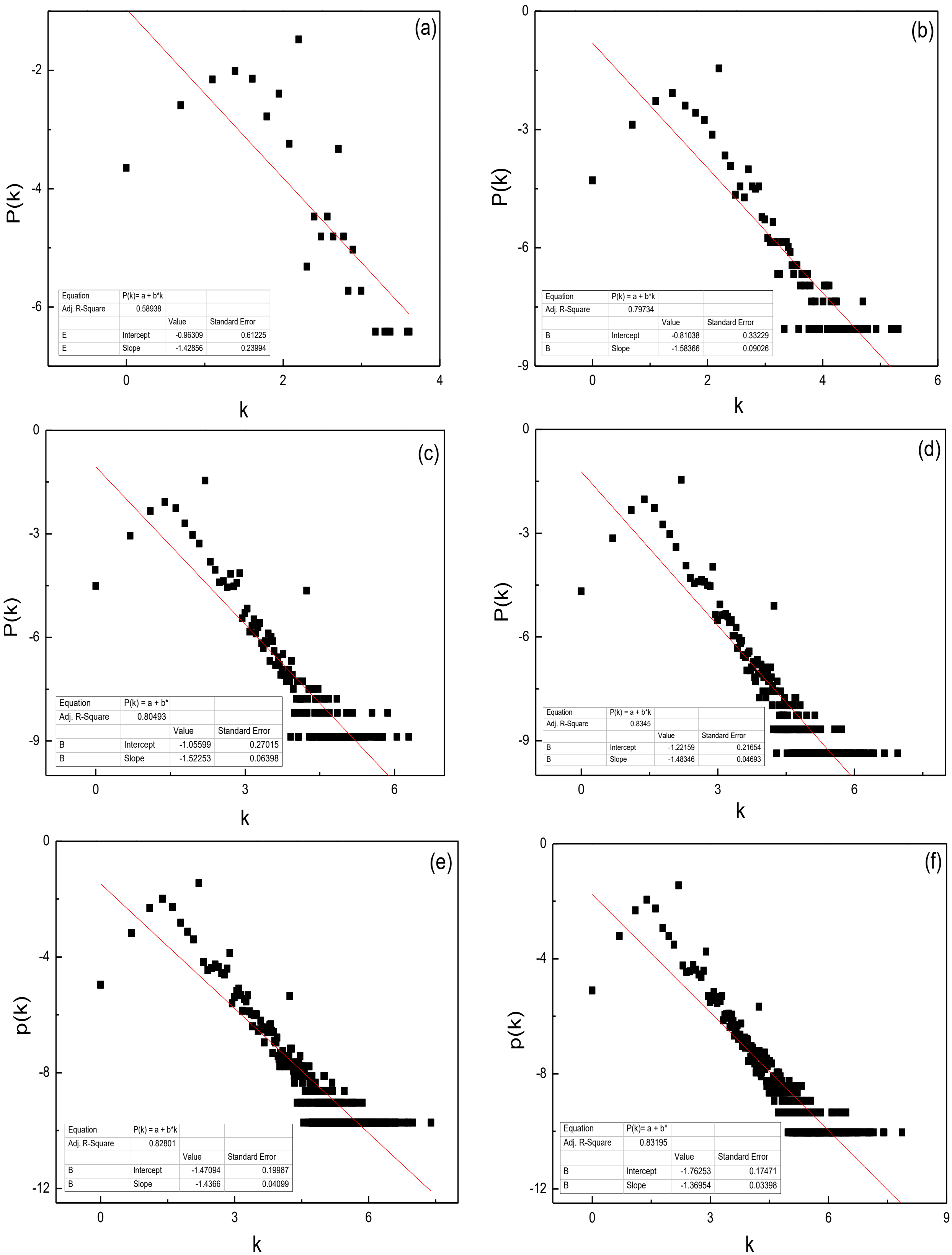

Figure 6. The degree distribution for the dynamic co-word network of GEQR: (a) 1998, (b) 1998 ~ 2002, (c) 1998 ~ 2006, (d) 1998 ~ 2009, (e) 1998 ~ 2012, and (f) 1998 ~ 2015.

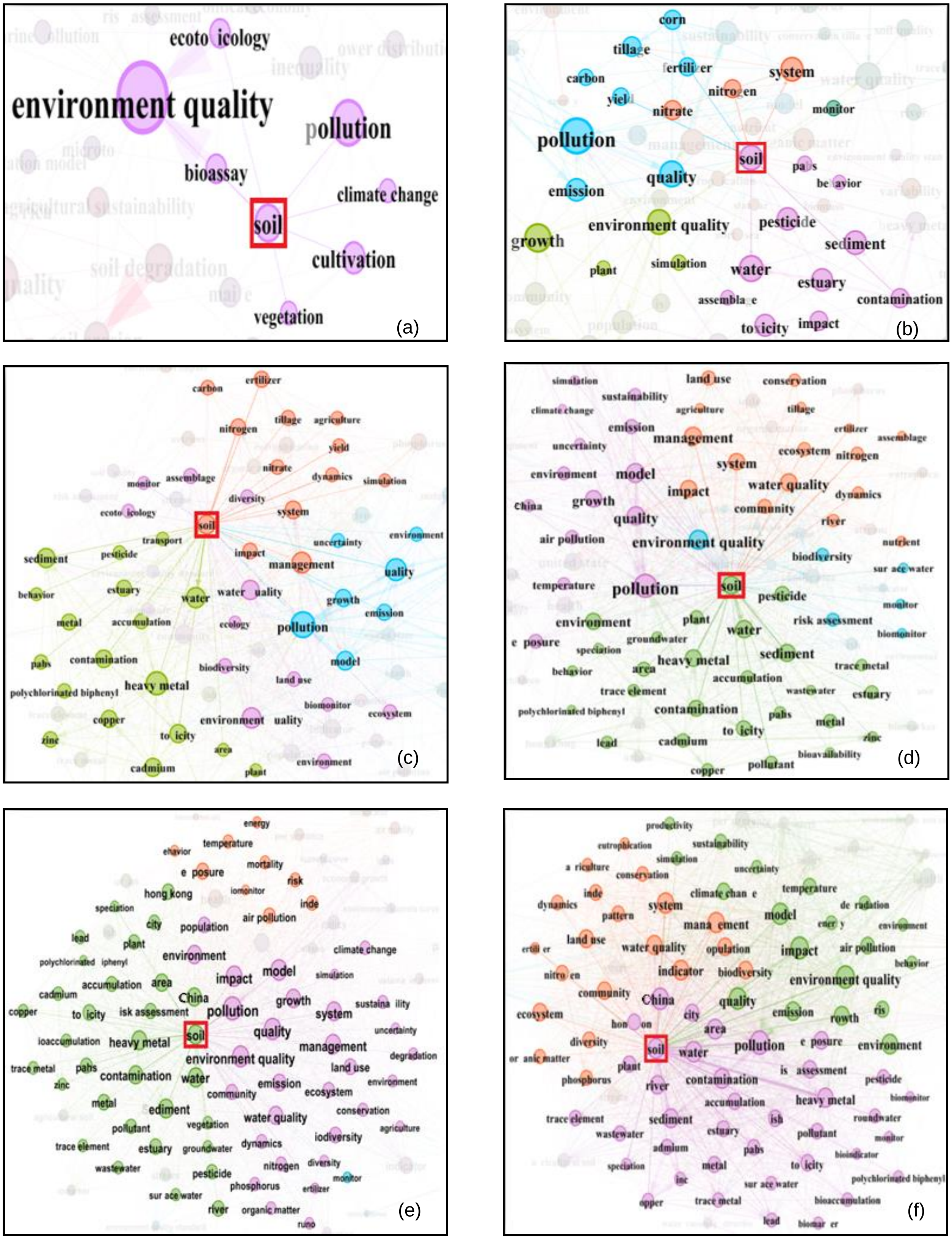

Figure 7. The gradual evolution of soil research interfacing with the lithosphere, hydrosphere, atmosphere, biosphere and anthroposphere: (a) 1998, (b) 1998 ~ 2002, (c) 1998 ~ 2006, (d) 1998 ~ 2009, (e) 1998 ~ 2012, and (f) 1998 ~ 2015.

of some specific nodes in the overall network, using soil in GEQR as an example. Soil is an important part of the Earth's surface and natural geographical environment. Before the 16th century, human understanding was based on intuition and some experience with agricultural production. After the 19th century, scientists began to consider soil an important element of global ecosystems (Feng and Chai, 2008; Luo et al., 2017). In this example, a table was created to show all publiccations containing the keyword "soil". The publications are next classified by year, resulting in a new table showing publications that included the keyword "soil" by year. By extracting the keyword "soil" to ascertain its dynamic appearance frequencies and dynamic links to other keywords at different stages, the local dynamic co-word network is obtained and the results are shown in Figure 7. In this way, the evolutionary path of the single keyword "soil" throughout the network can finally be obtained. As Figure 7 shows, the keyword "soil" had links with only a few keywords among the top 100 in 1998. By 2002, its ranking had gradually increased and with more links to the top 100 keywords and the nodes connected to "soil" had significantly increased. In 2006, 2009, 2012 and 2015, the connections to soil increased constantly, implying that its related scopes had increased as well. From the evolutionary path, it is found that soil research changed from the original focus on minerals/organic matter/organisms, to plant growth, water storage/supply/purification, modification of Earth's atmosphere and = organism habitat, among others. The degree to which soil research has gradually developed into a wide inter-disciplinary hot topic interfacing with the lithosphere, hydrosphere, atmosphere, biosphere and anthroposphere is apparent. On the development path of a specific topic, the node grows larger as it develops more connections to the top 100 keywords. Specifically, soil had only a few links at its origin, yet the number of node connections increased further as it became connected to more nodes. This explains the scale-free effect driven by nodal growth and preferential connection. The research nodal number of the network refers to the constantly increasing network size. By the preferential connection mechanism, the continuously produced new research nodes are more likely to connect to nodes that already have a high degree of connection. This mechanism controls the development path of a specific topic in the dynamic co-word network. The above-mentioned nodal growth and preferential connection mechanisms of the KS, however, were only analyzed using the simple "GIS" case with a relatively large time step (years). In future analysis, a more rigorous process should be used to track connection and growth with a shorter time step and more computational power.

Like the soil example, other cases can be similarly analyzed in the dynamic co-word network. The development paths of many specific topics with scale-free mechanisms prompt the evolution of the entire GEQR KS towards an emerging discipline that includes pollution, heavy metals, modeling, sediment management, and water quality as the hot trends. Scale-free nature is an important mechanism that may have universalities for many KSs. In fact, many KSs may have the characteristics of keyword growth and preferential attachment that can generate scale-free degree distributions (Luo et al., 2017; Zhang et al., 2017). The scale-free nature and its formation mechanism have practical uses for knowledge discovery and the self-organized regulation of many specific topics in a KS. By detecting the development path of a specific topic in the scale-free co-word network, the development of a KS for expected purposes can be more intentionally predicted and adjusted.

3.3. A New Strategy Map/Table on GEQR Based on Dynamic Co-Word Networks

3.3.1. Clustering on GEQR

Topological patterns emerge via a set of highly correlated keywords in the dynamic co-word network. The criteria of classifying clusters are based on community structure detection algorithms. The algorithms detect when nodes in a community are densely connected while connections between communities, i.e., the links between the nodes of one community and the nodes of another community, are sparse. By using quickly unfolding algorithms and partitioning the co-word network, all clusters of topological patterns are divided into different colors, as shown in Figure 8.

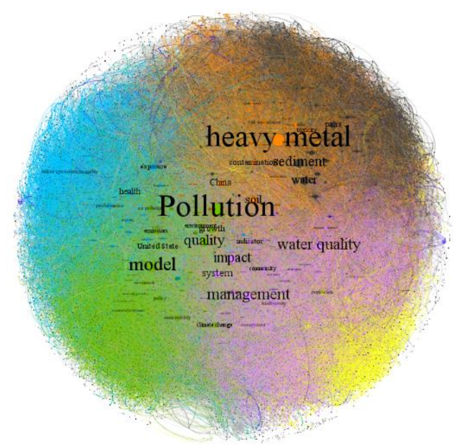

Figure 8. The topological patterns emerged by a set of highly correlated keywords in the dynamic co-word network.

Figure 8 classifies topological patterns by different colors, with the result that all dominant topics in GEQR diffuse to topical centers such as pollution, heavy metal, management and nitrogen. In Figure 8, the size and depth of nodes in the network represent the appearance number of keywords. The greater and deeper nodes represent more appearance times, whereas thinner edges indicate connection tightness between the different nodes. The thicker the edges, the closer the connections between nodes. Conversely, lighter color implies weaker connections. From the topological patterns, it is observed that GEQR comprises natural as well as social environmental aspects. Natural environmental aspects may be physical, chemical or biological. Social environmental aspects can be divided into social-economic, psychological, educational, cultural aspects of the environment, and so on. Indeed, as a measurement of human suitability for survival, reproduction, and social or economic development, environmental quality reflects human-specific requirements. It

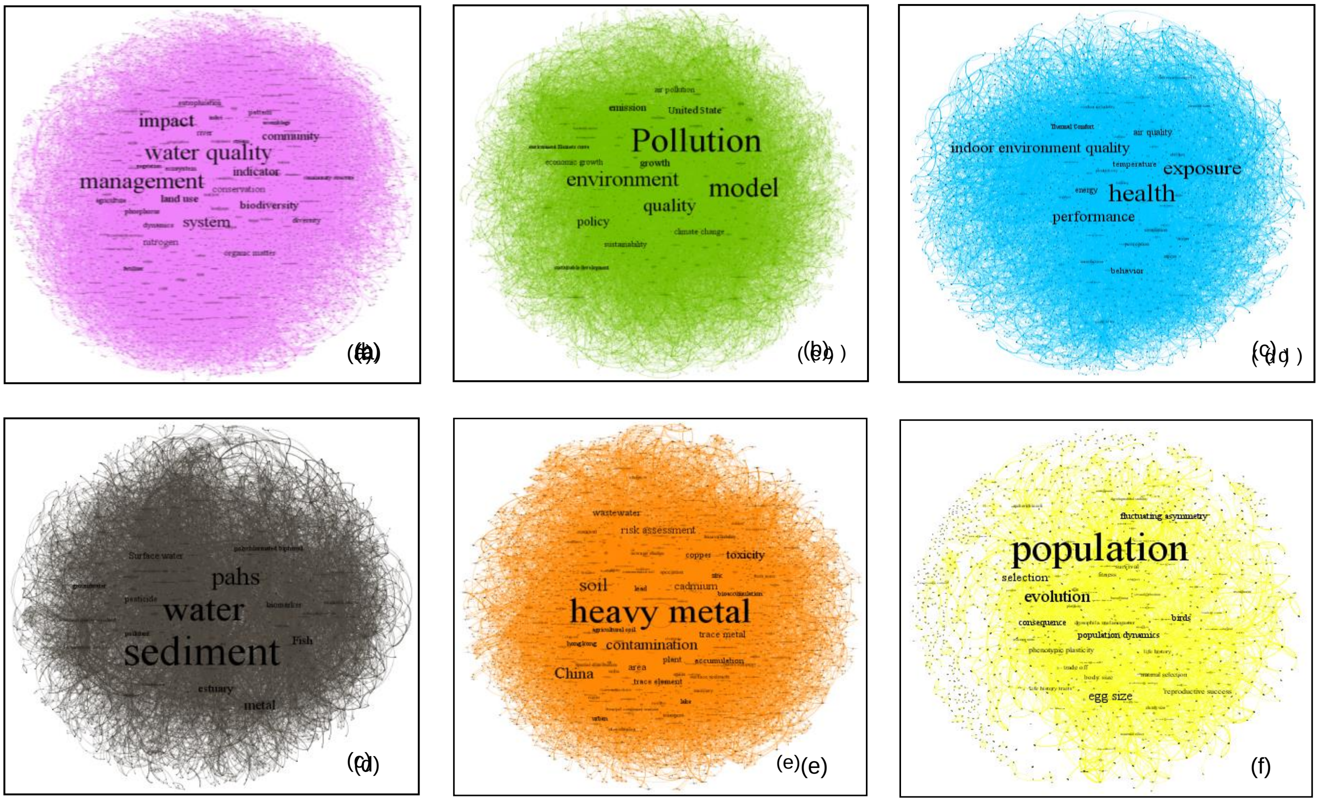

Figure 9. Co-word network's topological pattern categories of the six clusters for strategy map of GEQR: (a) Cluster 1, (b) Cluster 2, (c) Cluster 3, (d) Cluster 4, (e) Cluster 5, and (f) Cluster 6.

must consider various aspects of natural and social demands, including total environmental factors from the atmosphere, hydrosphere, biosphere, lithosphere and Anthroposphere. For humans to improve environmental quality, a specific assessment framework must be established to integrate each of these aspects for specific purposes. From a comprehensive perspective, the topological patterns that self-emerged from many related factors (i.e., keywords) of the dynamic co-word network are good descriptors of the clusters of GEQR.

3.3.2. An Assessment Strategy Map/Table for GEQR

Because of the large number of keywords and the complexity of factors, a precise cluster for GEQR has not yet been defined, however the topological patterns of dynamic co-word networks offer a new angle of investigation. This study uses the fast clustering algorithms in Gephi software to assign different colors for different topological patterns in GEQR, which are then divided into six perspectives as shown in Figure 9. The perspectives yield six clusters of GEQR relating to specific situations and purposes.

Table 3 provides a more intuitive understanding of the six clusters' strategy map of GEQR from the links between keywords shown in Figure 9. In Table 3, the top 15 important keywords (large nodes) are extracted from each cluster to establish the following six assessment systems: management, pollution, health, water, heavy metals and evolution. As shown in Table 3, specific assessment systems can be built into GEQR in accordance with the specific situations and purposes. For example, cluster 2 can be applied to assess environmental quality from an "air pollution" standpoint. The assessment system focuses on building a reference for keywords including environment, growth, quality, model, emission, city, policy, United States, risk, climate change, sustainability, and Kuznets curve. Thus, for the assessment object "pollution", cluster 2 can be used because all nodes connected with air pollution are references that can be invoked. To assess GEQR from the viewpoints of "indoor environment quality", cluster 3 can be applied. The purpose of the assessment system in this case is to build references for health, exposure, temperature, air quality, performance, energy, children, satisfaction, behavior, perception, indoor air quality, and particulate matter. For research involving keywords "China" and "sea", clusters 4 and 5 can be applied. This assessment system establishes references to heavy metal, pollution, contamination, PAHs, surface sediment, sediment, estuary, trace metal, and water. When studying "city", "population" and "management" themes, clusters 1, 2 and 6 can be used. In this case, the assessment system establishes references to soil, lead, metal, exposure, heavy metal, pollution, Kuznets curve, quality, indicator, system, policy, water, health, and biodiversity. In summary, cluster can be selected from Table 3 corresponding to any assessment object allowing all related factors to be conveniently extracted for construction of an assessment index system for that object. When the assessment object relates to numerous interdisciplinary factors, associated assessment index systems can be constructed with multiple corresponding clusters to extract more comprehendsive factors.

The topological patterns of the co-word networks (Figure 9) and the six clusters for the GEQR strategy table (Table 3) demonstrate the basic features of the new GEQR assessment system. The new assessment strategy map/table can facilitate further quantitative analyses by calculating network parameters. Figure 10 shows the calculated parameters of the co-word networks of six clusters, from which it is apparent that that edges *M*, diameter *D*, and average degree *k* have large vibrations, while other parameters do not change significantly. Cluster 1 has the most edges and the largest average connection degrees, which represent connections between the closest keywords, and mutual distances are shortest in this assessment system. The assessment strategy map/table proposed for GEQR is mechanistic and flexible, making it easy to operate and extract information from various interdisciplinary angles. Additionally, the self-organized, small-world and scale-free cluster characteristics have more applications for global environmental quality assessment in the future.

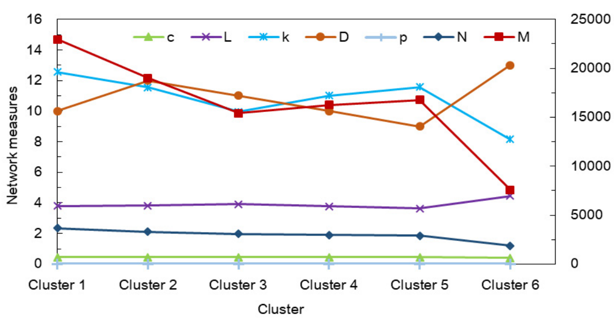

Figure 10. The parameter calculations for the co-word networks of the six clusters.

The practical use of the proposed assessment map/table is to offer a comprehensive macro-level view of GEQR. This may provide a new perspective for understanding global environmental quality of the total environment, which encompasses the atmosphere, hydrosphere, biosphere, lithosphere and anthroposphere. It also represents innovative development in the functioning of co-word networks.

Environmental related research focuses on the strong interactions between nature and the activities of all living beings in the world. Environmental problems are not only national affairs that require nations to handle environmental issues by state planning and comprehensive decisions, but also important parts of international politics, diplomacy and trade activities. Thus, compared with other subjects, environmental related research is unique in which the co-word network of the subject is tremendously large in scale and complicated in dynamics. The informetric dynamic co-word network method has the unique advantages, which can well reveal the self-organized dynamics, small-world and scale-free mechanisms during the large-scaled network evolution process. In addition, the assessment strategy map/table well provides new holistic understanding of the global environmental quality from the perspectives of a large informetric dynamic co-word network.

4. Conclusions

This paper develops a novel informetric dynamic co-word network method to analyze the evolution of GEQR's KS. The main contributions and conclusions can be summarized as follows:

The dynamic co-word network of GEQR shows how the self-organized and highly developed KS evolved from many small groups to a few large groups from 1998 to 2015. The entire network follows a growth trend that is closely related to the various directions of GEQR. Identified hotspots such as pollution, heavy metals, modeling, sediment, management, and water quality have emerged through a self-organized adjustment process that cannot be detected by existing methods.

Calculating the parameters of the dynamic co-word network quantitatively shows the underlying evolutionary dynamics of GEQR, including the oscillating clustering coefficient, growing average path length, degree and network diameter and decreasing or stabilizing density. GEQR's KS under the dynamic co-word network shows small-world and scale-free effects. Specifically, the dynamic co-word network confirms the importance of small-world and scale-free effects, which drive the evolution of the entire GEQR KS towards an inter-disciplinary structure with current hotspots, which existing methods usually ignore.

The topological patterns that emerged from the dynamic co-word network show robust clustering in GEQR. The conceptual domain of GEQR clearly encompasses both natural and social environmental quality. This paper identified topological patterns of the co-word network and divided them into six clusters, from which the assessment strategy map/table of GEQR was created to provide key factors or indexes for each cluster. Deriving this assessment strategy map/table based on the dynamic co-word network is a new development in co-word network functions.

Briefly, this work provides a comprehensive macro-level view of GEQR and offers new perspectives on global environmental quality from a total environment standpoint. As such, it can be instructive for selecting future research topics. Further, the evolutionary mechanisms of GEQR, namely hot topics, self-organized emergence, small-world link effects, and scale-free topical distributions, may allow scientists and policy-makers to efficiently learn and manage the GEQR KS to inform funding decisions and environmental policies. Finally, by using the strategy map/table, scientists and managers can conveniently extract different factors to construct an assessment index system. Although the practical uses and instructtions of the proposed method are intended for researchers, managers and policy-makers, they are currently more interesting as academic guides. These guides, however, may become more operational and empirical in the future by adding more detailed network analyses to the GEQR clusters. The universalities and rigorous mathematical-physical foundations of the informetric dynamic co-word network method has can be applied to evaluate KS evolution in other fields. Thus, the informetric dynamic co-word network method may prove a powerful tool for evolutionary analysis of many other environmental problems.

Acknowledgments. The paper was supported by Key Science and Technology Projects of Tibet Autonomous Region (No. XZ201703-GA-01) and Plateau remote sensing application technology innovation team.

References

Brookes, B.C. (1980). The foundations of information science: Part I. Philosophical aspects. *J. Inf. Sci.*, 2(3/4), 125-133. https://doi.org/10.1177/016555158000200302

Caldarelli, G. (2007). *Scale-free networks: complex webs in nature and technology*. Oxford University Press, Oxford.

Callon, M., Courtial, J.P., Turner, W.A., and Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Soc. Sci. Inf.*, 22(2), 191-235. https://doi.org/10.1177/053901883022002003

Coulter, N., Monarch, I., and Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *JACS*, 49(13), 1206-1223. https://doi.org/10.1002/(SICI)1097-4571(1998)49:13<1206::AID-ASI7>3.0.CO;2-F

Dehdarirad, T., Villarroya, A., and Barrios, M. (2014). Research trends in gender differences in higher education and science: a co-word analysis. *Scientometrics*, 101(1), 273-290. https://doi.org/10.1007/s11192-014-1327-2

Ding, Y., Chowdhury, G.G., and Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Inf Process Manag*, 37(6), 817-842. https://doi.org/10.1016/S0306-4573(00)00051-0

Feng, Q.Y. and Chai, L.H. (2008). A new statistical dynamic analysis on vegetation patterns in land ecosystems. *Physica A*, 387(14), 3583-3593. https://doi.org/10.1016/j.physa.2008.01.118

Ferrat, L., Pergent-Martini, C., and Romeo, M. (2003). Assessment of the use of biomarkers in aquatic plants for the evaluation of environmental quality: application to seagrasses. *Aquat. Toxicol.*, 65(2), 187-204. https://doi.org/10.1016/S0166-445X(03)00133-4

Gao, X., Zhou, F., and Chen, C.T.A. (2014). Pollution status of the Bohai Sea: An overview of the environmental quality assessment related trace metals. *Environ Int*, 62(4), 12-30. https://doi.org/10.1016/j.envint.2013.09.019

Karimipour, F. Delavar, M.R., and Kinaie, M. (2005). Water quality management using GIS data mining. *J. Environ. Inform.*, 5(2), 61-72. https://doi.org/10.3808/jei.200500047

Khan, M.A. and Ho, Y.S. (2012). Top-cited articles in environmental sciences: Merits and demerits of citation analysis. *Sci. Total Environ.*, 431, 122-127. https://doi.org/10.1016/j.scitotenv.2012.05.035

Kolokotsa, D. and Santamouris, M. (2015). Review of the indoor environmental quality and energy consumption studies for low income households in Europe. *Sci. Total Environ.*, 536, 316-330. https://doi.org/10.1016/j.scitotenv.2015.07.073

Larsson, M. and Hanberger, A. (2016). Evaluation in management by objectives: A critical analysis of Sweden's national environmental quality objectives system. *Evaluation*, 22(2), 190-208. https://doi.org/10.1177/1356389016638751

Li, W. and Zhao, Y. (2015). Bibliometric analysis of global environmental assessment research in a 20-year period. *Environ. Impact Assess. Rev.*, 50, 158-166. https://doi.org/10.1016/j.eiar.2014.09.012

Luo, R., Li, J., Zhao, Y.M., Fan, X.W., Zhao, P., and Chai, L.H. (2017). A critical review on the research topic system of soil heavy metal pollution bioremediation based on dynamic co-words net-work measures. *Geoderma*, 305, 281-292. https://doi.org/10.1016/j.geoderma.2017.06.019

Newman, M.E.J. (2003). The structure and function of complex networks. *SIAM Rev.*, 45(2), 167-256. https://doi.org/10.1137/S003614450342480

Penenko, V.V and Tsvetova, E.A. (2014). Methods of environmental quality assessments in conditions of climate variability. *20th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*, 9292, 929245. https://doi.org/10.1117/12.2075127

Robati, M., Monavari, S.M., and Majedi, H. (2015). Urban environment quality assessment by using composite index model. *Environ. Prog. Sustain. Energy*, 34(5), 1473-1480. https://doi.org/10.1002/ep.12125

Singh, K.P., Mohan, D., Sinha, S., and Dalwani, R. (2004). Impact assessment of treated/untreated wastewater toxicants discharged by sewage treatment plants on health, agricultural, and environmental quality in the wastewater disposal area. *Chemosphere*, 55 (2), 227-255. https://doi.org/10.1016/j.chemosphere.2003.10.050

Vrscaj, B., Poggio, L., and Marsan, F.A. (2008). A method for soil environmental quality evaluation for management and planning in urban areas. *Landsc Urban Plan*, 88(2-4), 81-94. https://doi.org/10.1016/j.landurbplan.2008.08. 005

Wang, X.Y., Chen, Y.B., Yang, L., Liu, G.L., Zhu, D.W., and Li, Z. (2015). Water environment quality assessment of three gorges reservoir based on the distance evaluation theory and SVM. *2015 4th International Conference on Energy and Environmental Protection*, (ICEEP 2015) 2960-2966.

Wang, X.G. (2009). Formation and evolution of science knowledge network (I): A new research method based on co-word network. *Journal of the China Society for Scientific and Technical Information,* 28(4), 599-605.

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of "small-world" networks. *Nature*, 393(6684), 440-442. https://doi.org/10.1038/30918

Xiong, Y., Zeng, G.M., Chen, G.Q., Tang, L., Wang, K.L., and Huang, D.Y. (2007). Combining AHP with GIS in synthetic evaluation of eco-environment quality - A case study of Hunan Province, China. *Ecol. Modell.*, 209(2-4), 97-109. https://doi.org/10.1016/j.ecolmodel.2007.06.007

Yi, S. and Choi, J. (2012). The organization of scientific knowledge: the structural characteristics of keyword networks. *Scientometrics*, 90(3), 1015-1026. https://doi.org/10.1007/s11192-011-0560-1

Zhang, Q.R., Li, Y., Liu, J.S., Chen, Y.D., and Chai, L.H. (2017). A dynamic co-word network-related approach on the evolution of China's urbanization research. *Scientometrics*, 111(3), 1623-1642. https://doi.org/10.1007/s11192-017-2