Structures and dynamics of scientific knowledge networks: An empirical analysis based on a co-word network

WANG Xiaoguang1,3*, JIANG Tingting2 & LI Xiaoyu1

1 Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China
2 School of Information Sciences, University of Pittsburgh, Pittsburgh, PA, United States
3 School of Information Management, Wuhan University, Wuhan 430072, China

Abstract  Co-word networks are constructed with author-provided keywords in academic publications and their relations of co-occurrence. As special form of scientific knowledge networks, they represent the cognitive structure of scientific literature. This paper analyzes the complex structure of a co-word network based on 8,190 author-provided keywords extracted from 3,651 papers in five Chinese core journals in the field of management science. Small-world and scale-free phenomena are found in this network. A large-scale co-word network graph, which consists of one major giant component and many small isolated components, has been generated with the GUESS software. The dynamic growth of keywords and keyword co-occurrence relationships are described with four new informetrics measures. The results indicate that existing concepts always serve as the intellectual base of new ideas as represented by keywords.

Keywords  Knowledge network, Co-word network, Complex network, Visualization, Dynamics, Small-world, Scale-free

1 Introduction

According to the sociology of science, scientific knowledge is created during the collaboration of a group of scientists rather than by a given program automatically. Stable universal core concepts, clear arrangement of research work, continuous internal and external communication all will facilitate the development of an academic discipline[1]. In particular, formal communication through publications within scholarly communities has become a crucial factor in knowledge creation and innovation[2]. New concepts, vocabularies, research methods and paradigms are
Proposed, standardized, accepted and recognized in the process of formal scholarly communication.

Publication-based communication will give birth to several types of self-generated networks, including citation, co-citation, co-author, and co-word networks. The first three are of high-level knowledge networks that represent the intellectual structures of any of a given disciplinary field. In contrast, the nodes of co-word networks are author-provided keywords, which are of basic knowledge units instead of being packaged as journals, articles or authors. Being at low-level of knowledge development, co-word networks reflect not only the collective understanding of scholarly communities, but also reveal the patterns of knowledge growth[3].

Previous studies have investigated the citation, co-citation, or co-author networks in various subject fields of academic writings. Both small-world[4] and scale-free[5] phenomena have been identified in the networks under research. Owing to a lack of empirical studies about co-word networks, we know little about their structural characteristics, which prevented us from understanding the development of academic disciplines with multiple perspectives. This research study intends to fill this gap by visualizing the structures and dynamics of co-word networks. The findings are expected to be beneficial in clarifying the scope of any of a given academic discipline and detecting emerging topics in that particular subject field area.

2 Research background

2.1 Mapping knowledge domains

During the 1990s, the development of scientific databases, statistical analysis software and computer graphic technologies dramatically increased the research interests in mapping knowledge domains. A wide variety of knowledge mapping algorithm and software applications, e.g. SOM, PFNET, Landscape, VxSight, Timeline, Crossmap, and Citesea, have been created and applied in generating scientific maps based on the citation or co-citation relations among journals or articles. Most of these studies are limited to specific knowledge domains, while there are extraordinary efforts made in mapping the backbone architecture of the entire system of academic disciplines[6]. Since 2002, Ledesdorff[7–9], Börner[10–12], Boyack[13–15], Chen[16–18], Morris[19–21] and a few others have made impressive progress in this research area.

There are five dimensions in the scientific knowledge system, namely, vocabulary, paper, author, journal and discipline[22]. In the past, the middle three are the major dimensions considered in the exploration of scientific knowledge networks, such as paper-based co-citation networks, journal-based co-citation networks, author-based
Structures and dynamics of scientific knowledge networks: An empirical analysis based on co-citation networks and co-author collaboration networks. Comparatively, co-word networks have drawn much less attention. The latter is vocabulary-based networks, focusing on presenting scholars’ understanding of their own subject specialties[14].

2.2 Co-word analysis

Co-word analysis is a content analysis method and a scientific mapping technique. Callon et al. introduced co-word analysis because they thought that co-citation analysis was inadequate in directly describing the changes of subject contents in academic writings or in the development of academic disciplines[23]. Co-word analysis has been extensively applied to information science research due to its ability to expose the embedded subject matters in articles. It excels in visualizing the knowledge structures of any of a given disciplinary field as well as their evolutionary development along with the passing of the time.

Nevertheless, the vocabulary-based mapping technique is inherently a threat to the validity of co-word analysis. A word may have different meanings within different contexts. So Leydesdorff has doubted that this method could scientifically describe the development of a particular disciplinary field[24]. Courtial countered such negative opinion by considering the vocabulary-based mapping technique, such as the co-word analysis, is applicable to academic assessment and knowledge discovery[25]. A comprehensive review of the co-word method by Courtial suggests that it has been proved successful in studying the development process of knowledge discovery. Nevertheless, there are some problems with this co-word method that are not ignorable, such as the apparent problems of indexing, the ambiguous definitions and correlations of co-words, simple clustering algorithms, and multiple similarity calculation approaches, etc[26].

Co-word analysis is usually used in the visualization of knowledge structure by means of multidimensional scaling analysis (MDS). MDS has a couple of inherent weaknesses, first of which is the limited sampling size. If the number of samples exceeds 1,000, MDS will consume significantly more time, and the legibility of the resulting knowledge map will also be significantly reduced. The second weakness is its limited applicability in analyzing the macro-structures of scientific knowledge. It is not easy to choose from the Pearson, Jaccard, or Cosine coefficients or to apply other related indexes to calculate similarities among objects[27–28].

3 Methodology

Thanks to the development of data analysis techniques and complex network mapping technologies, knowledge mapping based on a large vocabulary set
has been made possible, and co-word research studies are on the rise. Mane\cite{11}, Onyancha\cite{29}, and Su\cite{30} have analyzed the co-word relations among article keywords and built co-word networks for different subject domains.

### 3.1 Co-word network construction

Figure 1 illustrates how a co-word network is constructed. Let us assume that Articles A, B, C, and D respectively have 4, 3, 4, and 3 keywords. The keywords in A, B, and C respectively constitute a fully-connected sub-network. The three isolated sub-networks are linked together because keywords a3, b1, and c4 co-occur in D, and they form one co-word network. In this network, not all the relations among the keywords has the same strength. The more the two keywords co-occur frequently, the greater the strength of their relationship. Since the co-occurrence of keywords has no direction, the co-word network is of an undirected weighted network.

<table>
<thead>
<tr>
<th>Article</th>
<th>Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a1 a2</td>
</tr>
<tr>
<td>B</td>
<td>b1 b2</td>
</tr>
<tr>
<td>C</td>
<td>c1 c2</td>
</tr>
<tr>
<td>D</td>
<td>a3 b1 c4</td>
</tr>
</tbody>
</table>

![Fig. 1](image)

The basic building process of a co-word network.

Co-word networks value semiotics more than semantics. They are not able to accurately represent the semantic relationship between any two words in a specific co-occurrence. But this does not influence the usefulness of co-word networks in revealing the structures of scientific knowledge domains, which has been clarified in the academic exchanges between Courtial and Leydesdorff\cite{25}.

### 3.2 Analysis of the co-word network

We employed in our study three common measures of complex network analysis as introduced by Watts – density, clustering coefficient, and average path length\cite{31}. A co-word network with a higher density will have more keywords co-occurring. The clustering coefficient implies that the concepts within an academic discipline aggregate. A higher clustering coefficient means closer relations among different research subjects or directions. The distance between any two nodes is a measure
Structures and dynamics of scientific knowledge networks: An empirical analysis based on a co-word network

WANG Xiaoguang et al.

Research Papers

In general, the longer the distance between two keywords, the fewer the chance that they will co-occur.

3.3 The growth of co-word network

Sciences evolve forward constantly, so do co-word networks. When scientific concepts come into being, new nodes will be added to the networks. When scientific propositions are made, new links will be generated in the networks. When more researches are conducted to find out how one phenomenon is related to another, the relations among the corresponding keyword nodes will become stronger. When scientific domains arise, new subject categories will be associated with the nodes in the networks. Out of these considerations, we propose the following four new measures to examine the growth of co-word networks.

Annual keyword (AK): the collection of all the distinct keywords appearing during a specific year;
Annual new keyword (ANK): the collection of all distinct new keywords appearing during a specific year, which have never been seen in previous years;
Annual relationship (AR): the collection of all the distinct relationship of keyword co-occurrences appearing during a specific year; and
Annual new relationship (ANR): the collection of all the distinct new relationship of keyword co-occurrences appearing during a specific year, which have never been seen in previous years. This collection can be further divided into three categories: namely, 1) The relations among new keywords (ANR-NN), 2) the relationship between new and existing keywords (ANR-NE), and 3) the relations among existing keywords (ANR-EE).

4 Data

It is important to select a proper scientific domain as the context of our co-word network analysis. As most high-quality research papers on traditional natural sciences by Chinese researchers are published in international journals in English, we found that Chinese language journals that contain such comparable quality papers and research topics are indeed few and far in between. Besides, we also have taken those disciplines in the humanities and social sciences into consideration such as economics, sociology, history, and so on. The majority of the most significant research findings in these well-established disciplines are usually seen in published books or book chapters rather than in journal articles. After taking into account the issues of research publications in certain academic disciplines that involved with the above mentioned two variables, namely, the foreign language issue and the monographic format of presentation, we finally decided to concentrate
our study on management science so as to avoid the entanglement with those mentioned variables. A younger academic discipline in China, management science shows a clear track record of growing path. Relevant Chinese language journals in this disciplinary area have accumulated abundant keyword data, which make them ideal for our study.

Since it can be very costly to obtain sufficient historical data of any academic discipline for an analysis of the knowledge growth in that discipline, we have to determine a time frame for our data collection that is most cost-effective. The half-life phenomenon has been seen in article citing. Most disciplinary fields have 5-to-8-year half-life, meaning that every year half of the cited articles are published within the last 5 to 8 years. In addition, there are usually a small number of core journals in each subject field, which publish most of all the important articles. It is for this reason that core journals are considered to contain substantially more subject knowledge than those of other journals of lower academic ranking. Therefore, we used 8-year keyword data from 5 core journals to investigate the evolutionary structural development of management science in China.

4.1 Data acquisition

The five Chinese language core journals that we selected are *Chinese Management Science*, *Nankai Business Review*, *Management Journal*, *Journal of Engineering Management*, and *Journal of Management Sciences*. We chose these journals because they are all chronically listed in Chinese Social Science Citation Index (CSSCI) and the articles contained in them are of high quality. We roughly extracted 8-year (1999–2006) keyword data from the Wanfang Database, which is one of the three major periodical databases based in Mainland China, containing about 5,900 academic journals in the Chinese language.

As can be seen in Table 1, the time spans of the data from the five journals are in slightly difference. We obtained 8-year data from *China Management Science*, i.e. 1999–2006. But our data coverage from *Nankai Business Review*, *Journal of Management Engineering*, and *Journal of Management Sciences* is only for 7 years.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Time span</th>
<th>Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>China Management Science</em></td>
<td>1999–2006</td>
<td>1,219</td>
</tr>
<tr>
<td><em>Management Journal</em></td>
<td>2004–2006</td>
<td>400</td>
</tr>
<tr>
<td><em>Journal of Management Engineering</em></td>
<td>2000–2006</td>
<td>918</td>
</tr>
<tr>
<td><em>Journal of Management Sciences</em></td>
<td>2000–2006</td>
<td>494</td>
</tr>
<tr>
<td>Total</td>
<td>1999–2006</td>
<td>3,651</td>
</tr>
</tbody>
</table>

Table 1: The statistical data of journal articles
Structures and dynamics of scientific knowledge networks: An empirical analysis based on a co-word network

due to a lack of data in these journals in the year of 1999 in the Wanfang Database. Besides, our data coverage for Management Journal is only for 3 years as it started publication in 2004. The articles in these five journals are all research papers during the years surveyed, which have an aggregated number amount of 3,651.

4.2 Keywords extraction

In building the co-word network, we used author-provided keywords attached with the articles rather than using those automatically assigned by the Wanfang database system. The former type of keywords represents the way that those authors understand them within the context of their thematic discussions and so they have their concepts labeled accordingly. They are essential for the authors to conduct a clustering analysis. We need to consider synonyms, antonyms, abbreviations, bilingual words, the tenses in the English grammar, and so on. However, there are no clear standard guidelines yet for us to follow[26]. In actual practice, therefore, we had to skip this step due to two considerations. On the one hand, we assumed that researchers had adequate knowledge about their own disciplinary fields. They were supposed to be familiar with the established vocabularies in their own field of learning and were capable of selecting appropriate keywords in denoting the exact concepts within their disciplinary context. On the other hand, the scale of our co-word network was very large, which may defy an easy explanation about the impacts of some of the non-standard, esoteric and obscure keywords.

4.3 Building co-word matrix

The co-word matrix is the basis on which the co-word network is built. In traditional co-word analysis, researchers often use an inclusive factor to indicate the strength of the relationship between two words. But in our analysis relationship strength was not considered if it did not affect the co-word network topology. We converted the co-word matrix into a binary adjacency matrix, and the value in each cell was either 1 or 0, with 1 representing co-occurrence of two words and 0, no co-occurrence between two words. During the conversion, the values of all the non-zero $E_{ij}$ were reset to “1” and the rest remained “0”. The co-word network diagram based on such binary adjacency matrix was a one-mode, undirected, and unweighted graph.

5 Results

From our surveyed journals, there are 3,651 articles that contain 13,488 author-provided keywords in total. That is to say, each article has 3.7 keywords in them on average. From this pool of keywords, we detected 8,190 distinct ones.
5.1 The distribution of word frequency

Keywords are special data of academic writings. Ma has used the keywords and subject terms of a group of articles to validate the Bradford Law[32–33]. We examined the distribution of keyword occurrence frequencies and found that it followed the Bradford Law, as shown in Fig. 2. In a broad sense, such distribution obeys the Zipf’s Law which in informatics is actually a power law phenomenon with the exponent equaling 1[34].

\[ P_n = n^{-a} \]

\( P_n \) – the usage frequency of a keyword; and \( n \) – the ranking of the keyword

According to our results, the exponent \( a \) in the above formula is 2.1199 (\( R^2 = 0.9501 \)), which conforms to the Usability Principle in the system of academic exchange[34]. Scholars, in order to facilitate scientific exchange, are inclined to index their articles with widely recognized keywords that can be used by others when searching and assessing these articles.

![Fig. 2 The distribution graph of keyword occurrence frequency.](image)

5.2 The degree of relationship intensity for the distribution of nodes

In a network, the degree of relationship intensity of a node is the number of relations it has. Since the 8,190 keywords in our co-word network exhibit a total of 19,512 relations, the average intensity degree of relations for each of all the keywords is 4.7. We also analyzed the distribution of such degrees of keyword relationship intensity, and it obeys a power law (Fig. 3). The exponent \( a \) in the following formula is 1.7719 (\( R^2 = 0.8478 \)).

\[ f(k) = nk^{-a} \]

\( k \) – the degree of a keyword; and \( f(k) \) – the occurrence probability of keywords whose degrees equal \( k \)
In Fig. 3, the degree of relational intensity of many keywords in the raw data equals 1, which implies that some articles have only one keyword. This phenomenon is abnormal since research papers are usually indexed with at least two keywords. If a paper has only one keyword, it is possible that the author(s) fails to provide more of them. This one-keyword situation leads to some deviations in the fitting. After removing these singular keywords, a new fitting result is shown in Fig. 4, where the exponent is 1.9479. Apparently, the fitting result is improved to some extent ($R^2=0.8927$) as shown in the following two figures.

![Original intensity degree of keyword distribution.](image1)

![The adjusted degree of keyword distribution.](image2)

As we can see so far, the degree of keyword intensity distribution presents the power-law distribution instead of the Poisson distribution, indicating that this co-word network is a scale-free network. A small number of core keywords, which have large degrees, generate indirect relations among many keywords. And apparently, these core keywords are the key concepts during the development of an academic discipline. Given their significance to our research, we extracted 40 top-ranked keywords in terms of their ranking of occurrence frequency (Table 2).
5.3 Density, cluster coefficient and the shortest path length

Table 3 displays the values of several network measures calculated for our co-word network. Since $C (0.86)$ is much higher than $C_{\text{rand}} (0.00039)$ but $<d>$ is close to $<d_{\text{rand}}>$, it can be concluded that this co-word network is a small-world network[20]. Actually we may predict such finding when building the co-word network. The keywords in each article will form a fully-connected sub-network. This co-word network, based on 3,651 articles, comprises 3,651 fully-connected sub-networks, resulting in a high cluster coefficient. The low density (0.00056) indicates that the nodes in the network are sparsely connected. It further implies that management science is an academic discipline with many branches.

### Table 3 Features of the co-word network

<table>
<thead>
<tr>
<th>Item</th>
<th>Value</th>
<th>$N$</th>
<th>$E$</th>
<th>$&lt;k&gt;$</th>
<th>$C$</th>
<th>$C_{\text{rand}}$</th>
<th>$&lt;d&gt;$</th>
<th>$&lt;d_{\text{rand}}&gt;$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>8190</td>
<td>19512</td>
<td>4.7</td>
<td>0.86</td>
<td>0.00039</td>
<td>4.9</td>
<td>6.0</td>
<td>0.00056</td>
<td></td>
</tr>
</tbody>
</table>

* $N$: Node; $E$: edge; $<k>$: average degree; $C$: cluster coefficient; $C_{\text{rand}}$: cluster coefficient of corresponding random network; $<d>$: average shortest path length; $<d_{\text{rand}}>$: average shortest path length of corresponding random networks; $d$: density.
5.4 The visualization of relations

The global structure of our co-word network was mapped and analyzed with the GUESS software, an exploratory data analysis and visualization tool for graphs and networks[36].

As shown in Fig. 5, this co-word network is made of 442 isolated components. The largest component consists of full-connected 6,536 nodes. The rest components are much smaller, with each containing 3 to 5 nodes. The bottom part of Fig. 5...
zooms in some small components. Most of the keywords in these components are rarely used words or phrases, e.g. “increase of ability”, “insurance pricing”, “traffic and transportation”, “mining enterprise”, “pricing formula”, “technician”, “development target”, “practice restraint”, “the privatization of state-owned enterprises”, etc. Such keywords are seldom seen in the standard indexing practices. Therefore, it is possible that they cause the failure in connecting the small components with the main component.

5.5 The growth of relationship between keywords and co-words

Our analyses of the growth of relations between keywords and co-words relationships are demonstrated in Tables 4 and 5 respectively.

<table>
<thead>
<tr>
<th>Year</th>
<th>AK</th>
<th>ANK</th>
<th>ANK/AK(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>127</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2000</td>
<td>806</td>
<td>792</td>
<td>98.3</td>
</tr>
<tr>
<td>2001</td>
<td>1022</td>
<td>891</td>
<td>87.2</td>
</tr>
<tr>
<td>2002</td>
<td>1867</td>
<td>1550</td>
<td>83.0</td>
</tr>
<tr>
<td>2003</td>
<td>1575</td>
<td>1165</td>
<td>73.9</td>
</tr>
<tr>
<td>2004</td>
<td>1902</td>
<td>1379</td>
<td>72.5</td>
</tr>
<tr>
<td>2005</td>
<td>1960</td>
<td>1291</td>
<td>65.8</td>
</tr>
<tr>
<td>2006</td>
<td>1595</td>
<td>979</td>
<td>61.4</td>
</tr>
</tbody>
</table>

In the above table, we can find a clear downtrend in the percentage of annual new keywords in the annual keywords category from 1999 to 2006. However, this ought not to be interpreted as a decline of innovation in management science. Instead, it is caused by data collection. Many of the keywords seen after 1999 had possibly already appeared before 1999, which affected the later years to have a lower percentage of appearance. On account of the paper citing in half-life, the percentage would have probably stabilized at a fixed level which was unattainable in this paper because of the limited time span of the data collected.

<table>
<thead>
<tr>
<th>Year</th>
<th>AR</th>
<th>ANR</th>
<th>ANR/AR (%)</th>
<th>ANR-NN</th>
<th>ANR-NN/ANR (%)</th>
<th>ANR-NE</th>
<th>ANR-NE/ANR (%)</th>
<th>ANR-EE</th>
<th>ANR-EE/ANR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>162</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2000</td>
<td>1246</td>
<td>1246</td>
<td>100</td>
<td>1160</td>
<td>93.1</td>
<td>85</td>
<td>6.8</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>2001</td>
<td>1717</td>
<td>1712</td>
<td>99.7</td>
<td>1182</td>
<td>69.0</td>
<td>467</td>
<td>27.3</td>
<td>63</td>
<td>3.7</td>
</tr>
<tr>
<td>2002</td>
<td>3428</td>
<td>3392</td>
<td>98.9</td>
<td>1874</td>
<td>55.3</td>
<td>1244</td>
<td>36.7</td>
<td>274</td>
<td>8.1</td>
</tr>
<tr>
<td>2003</td>
<td>2925</td>
<td>2856</td>
<td>97.6</td>
<td>1243</td>
<td>43.5</td>
<td>1264</td>
<td>44.3</td>
<td>349</td>
<td>12.2</td>
</tr>
<tr>
<td>2004</td>
<td>3355</td>
<td>3289</td>
<td>98.0</td>
<td>1359</td>
<td>41.3</td>
<td>1424</td>
<td>43.3</td>
<td>506</td>
<td>15.4</td>
</tr>
<tr>
<td>2005</td>
<td>3573</td>
<td>3458</td>
<td>96.8</td>
<td>1096</td>
<td>31.7</td>
<td>1591</td>
<td>46.0</td>
<td>771</td>
<td>22.3</td>
</tr>
<tr>
<td>2006</td>
<td>2930</td>
<td>2799</td>
<td>95.5</td>
<td>858</td>
<td>30.7</td>
<td>1303</td>
<td>46.6</td>
<td>638</td>
<td>22.8</td>
</tr>
</tbody>
</table>
Table 5 shows a similar downtrend. The percentage of annual new relations (ANR) versus total annual relations (AR) also decreases from 1999 to 2006. Moreover, the percentages of the three different types of new relations (i.e., ANR-NN, ANR-NE, and ANR-EE) in annual new relations have changed. Specifically, the percentage of ANR-NN declined greatly, whereas those of ANR-NE and ANR-EE showed an obvious rise. In 2006, the percentages of ANR-NN, ANR-NE, and ANR-EE were 30.7%, 46.6%, and 22.8% respectively. These percentages may vary with different time spans of data collection. The percentage of ANR-NN, especially, is in a positive correlation with the length of time span.

6 Discussion

This paper reviewed the mapping of knowledge domains and along with it a co-word analysis. It also put forward the concept and theory of co-word network. Our qualitative analysis argued that co-word networks are realistic knowledge networks. From the vocabulary’s angle, they account for the logical consistency among mapping, literature measurement, knowledge organization, knowledge discovery, and other concepts in academic research practices. In the meanwhile, they introduce an original theoretical perspective into the study of knowledge growth, which enables us to observe the formation and evolution of knowledge networks stemming from low-level and microcosmic knowledge units.

6.1 Macro-features of the co-word network

The small-world phenomenon, which exists in co-citation networks and co-author networks, is also evident in co-word networks. The average path length of our co-word network is 4.9, and that of human language networks and syntactic dependency networks is 3 in accordance with the research findings of Wei[37], Ferrer-i-Cancho[38], Newman[39], Liu[40], Liu[41], etc. Several factors may explain this difference: (1) author-provided keywords in research papers are usually in the form of phrases rather than in single words; (2) our sample size is not large enough and there are not plenty of connections between keywords; and (3) many keywords are synonymous, but they have not been normalized in this paper. In our previous analysis of a co-word network representing a specific academic discipline, the average path length is about 2.8[42]. Averaging it with the value in our current study, we will get 4, which means that in the keyword-based knowledge space, three to five keywords are adequate to locate an article.

Being scale-free, our co-word network has a small number of keywords with many relationships. They play an important role in helping researchers understand the overall structure of a knowledge domain and navigate it within that knowledge space.
We collected the keyword data from Chinese language journals in the field of management science, but our results show noticeable complex network features in the co-word network. Studies on complex networks have indicated that many real world complex networks are self-similar\cite{143}. Boyack, who investigated the backbone structures of scientific networks, claimed that different disciplines were interrelated and bore a strikingly high similarity to one another\cite{6}. So we assume that other disciplines and the whole range of academic disciplines had such similar network features as those of management science. Specifically, the networks of all sciences are interconnected, with each discipline occupying its own territory (or community). The relations among the knowledge units within a particular discipline are very dense, whereas the relations among disciplines in general are sparse. Different disciplines and specialties have different situations. As a result, the all-inclusive network of scientific knowledge presents the small-world and scale-free features.

6.2 The structures of components in a co-word network

Sciences and technologies develop gradually, with the creation of new knowledge building upon previously accumulated knowledge. Therefore, newly published scientific articles will be related to the existing literature to different extent through citations or references, despite the fact that they in fact develop or deduce existing scientific concepts and repeat, refute, or correct prevalent conclusive assertions and beliefs. In co-word networks, the majority of all the nodes should be reachable to one other, which means that they are either directly or indirectly connected.

The visualization of our co-word network supports the above assumption. This network consists of a huge main component and a large number of isolated small components. The keywords in the latter components are mostly neither standard nor frequently used ones. We hence think that these small components can be ignored in representing the scientific knowledge network.

6.3 The growth of co-words network

From the viewpoint of complex networks, the scale-free feature of our co-word network originates in the Preferential Attachment mechanism during the growth of the network\cite{30}. As knowledge grows, the increase of new relationships preferentially starts with two specific types of keywords. One includes the fundamental concepts of a subject, which always elaborates on the thematic tenet or research directions of this particular research paper. The other type comprises new concepts representing the leading edge of each discipline. Every year we will see some new relations between keywords and co-words appear, but these new scientific concepts or topics do not come into being abruptly and they can help discover the developing trends of all academic disciplines\cite{17}.
7 Conclusion and limitations

This study analyzed a large-scale co-word network based on the keyword data from five Chinese management science journals which covered roughly a time span of 8 years (from 1999 to 2006). We disclosed the complex structure of the co-word network and found that the growth of author-provided keywords followed the power law. Despite the abnormal keywords and limited data coverage, our research results are intriguingly complex. The co-word network, as a whole, presents the small-world and scale-free features. There exists a large principal component in this network and many very small components. The keywords did not increase greatly during these surveyed years, as the percentage of annual new keywords during this period was not high. ANR-NE and ANR-EE account for the vast majority of new co-word relations. In conclusion, the development of academic disciplines is a gradual progress in which existing knowledge is the foundation for creating new knowledge.

Given that the data collected in this study is confined only to a specific discipline, we are not able to describe how knowledge flows between different disciplines, i.e. how the knowledge and concepts of one discipline are being used in another. Besides, we concentrated on the nodes of the co-word network and the growing patterns of their relations, instead of examining the structural changes of the network, such as the increase or decrease in network density, the shrinkage of the gaps in the network, the rise and fall of network communities, and so on.

It should be mentioned that we did not normalize the keywords when constructing the co-word network. This might cause the loss of relations among some keywords and reduce the degrees of the nodes and the diversity of the entire network. Moreover, the time span and the size of the collected data vary from journal to journal. While the data from some journals covered a period of 8 years, yet some of those in others, only 3 years. Because different journals may focus on different topics, their keyword data could result in topical communities with different density. There are both dense and sparse aspects in the network, even though their impacts on the growth of keywords and keyword relations can be ignored.

The “sleeping beauty” phenomenon has been seen in scientific domains. Some concepts and topics or themes begin to revive even after having been ignored for years. We set the year 1999 as the division line to differentiate newly emerging and existing keywords. However, the fact is that some keywords were possibly already used before 1999, rather than being totally “new.” This is a problem we can hardly avoid in this study. As a matter of fact, the longer the time spans for the data collection, the lower the value of ANK/AK. Due to the half-life phenomenon, this value may finally stabilize at a certain level which we cannot elaborate them here in detail because of our time restraint.
Nevertheless, our study is truly inspirational. It helps us understand the static structures of knowledge network that is made up of micro knowledge units, i.e. author-provided keywords, and their growing patterns. In the future, we will employ the above presented research method to analyze other co-word networks so as to verify the findings of this paper. Besides, we are planning to normalize the keywords automatically with a PLSA (probabilistic latent semantic analysis) and to conduct diachronic tracking studies on the flow of knowledge among different disciplines. We shall also pay special attention to the growth, integration, and decomposition of topical communities within the co-word network so that a more precise evaluation of the academic topics at the intermediate level can be made. Ultimately we aim to find out other patterns of the evolutionary development of knowledge networks.

References

Structures and dynamics of scientific knowledge networks: An empirical analysis based on a co-word network

26 He, Q. Knowledge discovery through co-word analysis. Library Trends, 1999, 48(1):133–159.


(Copy editor: Ms. Jing CAO; Language revision: Prof. Charles C. YEN)