Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 91 (2016) 547 – 555
Procedia Computer Science

Information Technology and Quantitative Management (ITQM 2016)

# Mapping the research trends by co-word analysis based on keywords from funded project

Xiuwen Chen[a,b], Jianming Chen[a], Dengsheng Wu[a,] *, Yongjia Xie[a], Jing Li[c]

*a. Institute of Policy and Management, Chinese Academy of Sciences, Beijing 100190, P. R. China*
*b. University of Chinese Academy of Sciences, Beijing 100190, P. R. China*
*c. School of Management, Anhui University, Hefei 230039, P. R. China*

**Abstract**

In this paper, a co-word method based on keywords from funded project is proposed to map the research trends. Firstly, the keywords of funded project are used to describe the rsearch topic statistically. Then, co-word analysis, including cluster analysis, social network analysis, is adopted to study the relationship of each research topic. The projects of Management Science and Engineering in National Natural Science Foundation of China during 2011-2015 are collected as the empirical data. The data is composed of General Project, Youth Project, and Regional Project. The results show that the focus of researches are Game Theory, Supply Chain Management, Complex Network, Data Mining, Optimize, Risk Management, and Data Envelopment Analysis. Moreover, Game Theory, Supply Chain Management, and Data Mining are hot topics. The research fields in Management Science and Engineering in China are varied, and the well developed and core research fields are fewer.


Peer-review under responsibility of the Organizing Committee of ITQM 2016

*Keywords*: co-word analysis; Fund project; social network analysis; Management Science and Engineering

## 1. Introduction

Bibliometric analysis method, which has the basic characteristic of quantitative, has been applied by many researchers in their studies [1]. Recently, people pay more attention to reference management, and the research scope expands constantly. We can find the research distribution of one field by the Bibliometric analysis

* Corresponding author. Tel.: +86-10-59358813.
E-mail address: wds@casipm.ac.cn.


Peer-review under responsibility of the Organizing Committee of ITQM 2016
doi:10.1016/j.procs.2016.07.140

method. It also has significance influence on adjusting our research direction and discovering the hot topics. At present, the more popular method is co-word analysis, proposed as early as the late 70s in 20th century by the French bibliometric scientists [2]. Its principle can be summarized as follows: when two or more professional keywords representing a special research topic appeared in the same dissertation have essential relationships [3]. And the more co-occurrence between two keywords, the closer their relationship is. To reveal the structure and development of research fields, some co-word analysis methods will be applied based on co-word matrix, which consist of factor analysis, cluster analysis, multivariate analysis and social network analysis. These methods are helpful for researchers to learn overview of a field. So it plays an important role in identify the value of academic discipline [4, 5].

Co-word analysis has been used to study conceptual work in different domains by many researchers. Dehdarirad et.al select 959 full text articles included in the journal Scientometrics to mapping the intellectual structure of scientometrics using text-mining and co-word analysis [6]. Some social researchers make use of co-word analysis to analyze the structure and development of the scientific literature on gender differences in science and higher education rely on 652 published articles, concentrating on factors related to differences between 1991 and 2012 [6]. There are researchers studied some other topics, for instance, library and information science [7]; recommendation system in China [8]; robot technology in Korea [9]. Is it an effective way of basing on the published papers to study the academic trends of one territory? Generally speaking, the article published in 2012 accepted in 2011, in other words, this article accomplished before 2011 which concerned with hot topics between 2008 and 2010. There is serious possibility that the hot topics analyzed from published papers in 2012 couldn't include the latest trends. How to solve this problem? We find that the articles published in 2012 obtained funding in 2008. That is to say, the fund project was applied in 2008. In this way, there is adequate reason to use the keywords in funding project to analysis research trends. These keywords may be more timeliness. Almost researchers concerned about the relationship between science funding and research output [10, 11], and funding ratios in a research field [10]. Thus, we will be the first to analysis the trends of a field using keywords in funding project.

## 2. Method and data

### *2.1. Data collection and data process*

For this study, several important steps will be followed to accomplish our research: counting the keywords frequency, selecting high-frequency keywords, making co-occurrence matrix, clustering keywords, interpreting intellectual structure of topics represented by social network analysis.

We must determine the analysis unite before co-word analysis. Researchers almost select keywords extracted from papers as basic analysis unite. In this study, we select keywords extracting from the projects of Management Science and Engineering in National Natural Science Foundation of China as our research data, which include Youth Science Fund Project, General Science Fund Project and Regional Science Fund Project. The time interval is 5 years, spanning from 2011 to 2015. We totally acquire 7304 fund projects.

In our fund project data, some projects have two application code (application code 1 and application code 2) while others only have application code "1". Thus, we select project, which the first three of application code 1 is "G01", as the research data. Finally, we extracted 6153 keywords from 2054 fund project with total 9348 frequencies.

The keywords with frequencies lower than 8 implied that there are few researchers pay attention to them. So we remove the keyword which has less than 8 frequencies, 100 related keywords with an individual frequency greater than 8 were used for this study at last. The steps of co-word analysis of Management Science and Engineering were shown in Fig.1.

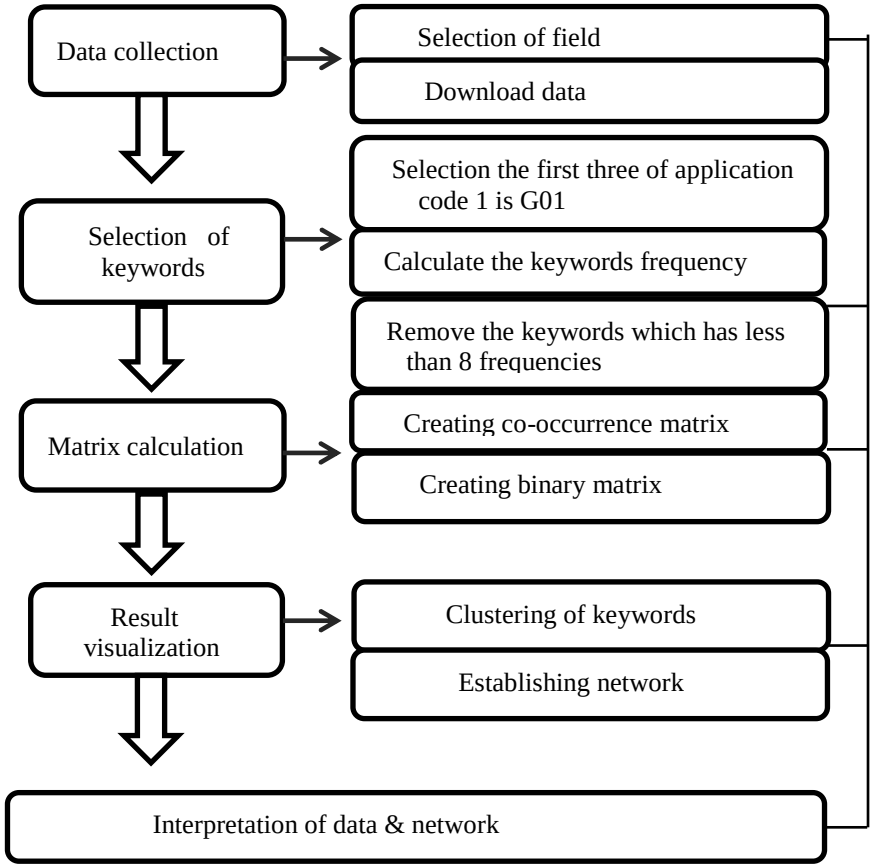


Fig. 1 step of co-word analysis used in Management Science

### *2.2. Method*

After selecting 100 keywords which frequency greater than 8 as our data, the next step is to acquire co-occurrence matrix. Computing the frequency of two keywords appeared together in the same paper, we will get a symmetrical co-occurrence matrix based on the word co-occurrence [12]. In this matrix, values in the diagonal cells were keyword frequencies, and values of non-diagonal cells were co-word frequencies. Two keywords occur in a same article is an indication of connection between the topics which they represent. The higher frequency of co-occurrence between keywords, the closer research theme is [4]. Currently, Bibliographic Items Co-occurrence Matrix Builder (BICOMB), Bibexcel, Ucinet , Citespace are popular instruments to co-word analysis [6, 8]. Their data source are PubMed, Web of Science, CNKI, with the format are txt or html. Under the help of these instruments, co-occurrence matrix will be easily acquired. However, the format of data we obtained is xls. So there is no available related tool that can be put into use directly. To realize our aim, the first step of counting word frequency achieved in Pivot Table, co-occurrence matrix

received by writing a program in Matlab. Co-correlation matrix or other similarity matrix is conversed from co-occurrence matrix using Ucinet, which is a basis of co-word analysis.

In our study, we also apply cluster and social network analysis to reveal the intellect structure of Management Science and Engineering by using Ucinet and VOSviewer. Clustering is a method which have a merit of setting objects into a group by similarity or dis-similarity [13]. Thus, generally speaking, keywords with high correlation with each other having tendency to put into the same cluster. Social network analysis (SNA) evaluating the unique structure of interrelationships among individual has been extensively used in social science, psychological science, management science and scientometrics [13]. We can get a sociogram from social network analysis. The main character of sociogram is providing information about the relationship between members of a network. In the sociogram, each member of a network is described by a "vertex" or "node". Vertex stands for high-frequency words, and the sizes of nodes shows the occurrence frequency. The smaller size about a node, the lower occurrence frequency is. Line depicts the connection relationship between two words, which exists between two keywords representing they appeared in the same dissertation. The thickness of which is drawn proportionally to the correlation between keywords. The thicker line between two keywords, the stronger the connection is. Using this rational, the map visualization and network characteristic (centrality, density, the core-periphery structure, a strategic diagram and the network chart) were obtained by analyzing Pearson's correlation matrix or other similarity matrix [8]. In this study, network analysis conduct on binary matrix to display the trend and current statues of research in management science and engineering in China.

Meanwhile, k-core analysis is commonly used in SNA. A k-core is a maximal group of nodes, all of which are connected to at least k other nodes in the group (Eschenfelder 1980; Maimon and Rokach 2005). By varying the value of k (that is, how many members of the group do you have to be connected to), different pictures can emerge. As the value of k becomes larger, group sizes will decrease, and the relationship among the members will be tighter. Line between two nodes indicates that the two keywords have appeared in a same dissertation.

We transformed co-occurrence matrix to binary matrix by Ucinet. The threshold is 1 and value "0" will be assigned in new binary matrix if the co-occurrence frequency of keywords was less than one in co-occurrence matrix. Value "1" stands for the opposite side. Network visualization accomplished in VOSviewer.

## 3. Result and discussion

The keyword of an article can represent its main content, and the frequency of occurrence and co-occurrence can reflect themes focuses in a special field to some extent [4]. The top ten keywords with high frequency of occurrence are game theory (48), supply chain management (43), complex network (41), data mining (39), optimize (29), supply chain (29), risk management (29), Data Envelopment Analysis (25), computational experiment (24), and pricing (24). The top ten keywords with high frequency of co-occurrence are game theory (45), supply chain management (30), complex network (31), data mining (31), optimize (17), supply chain (16), risk management (18), Data Envelopment Analysis (19), asset pricing (14), uncertain (14). We may easily find that game theory, supply chain management, complex network, data mining, optimize, supply chain, risk management, Data Envelopment Analysis not only have individual high frequency but co-occurrence frequency, and reflecting that these research topics attract more attention and have a close relationship with other research topics.

### 3.1. Clustering

We got the graph of co-occurrence relationships among the 100 keywords using VOSviewer (Fig.1). There is a merit which can be combined cluster and map together visualization by VOSviewer, and with the help of VOSviewer offing zoom and scroll functionality, we may acquire more detail information of a map. We may easily find that “behavior decision” and decision support system” are not connected with others, so we use the rest of 98 items to instead of all items. In the network visualization (Fig.1), keywords in the different clusters have display different colors. The 98 keywords of Youth Science Fund Project, General Science Fund Project and Regional Science Fund Project were divided into 8 clusters. If keywords are grouped into a same cluster, they are more likely to reflect identical topics. Each cluster has different number of subject keywords. It indicates that the research fields of management in China are varied. The detail information of clusters is shown in Table 1.

From the Table 1, we can see that cluster 4 has the largest number of keywords, indicating that theme-cluster 1 absorb more attention from the researcher in China. That also means cluster 4 is the most centralized fields [8]. In other words, the keywords in cluster 4 have been pay more attention in the field of Management Science and Engineering [7].

Table 1. 8 cluster of management science in China

| Cluster | Number of keywords | Selected keywords |
|---|---|---|
| 1 | 9 | game theory; supply chain coordination; supply chain management; inventory management; behaviour operations management; inventory control; dynamic programming; robust optimization; dynamic pricing |
| 2 | 10 | risk management; risk measurement; asset pricing; market microstructure; final market; risk control; liquidity risk; high frequency data; option pricing; credit risk |
| 3 | 12 | coordination mechanism; supply chain; coordination; game; pricing; closed loop supply chains; remanufacture; system dynamic; performance evaluation; operation management; risk preference; equilibrium |
| 4 | 19 | complex network; big data; optimize; complexity; knowledge management; computational experiment; social media; evolution; social network; urban traffic; information fusion; decision support system; complexity; evolution mechanism; knowledge transfer; multi-attribute decision-making; group decision making; evolution mechanism; prospect theory |
| 5 | 12 | uncertainty; data envelopment analysis; systemic risk; computing experiments financial; efficiency evaluation; heterogeneity; resource allocation; simulation; reliability; variable selection; predict; panel data |
| 6 | 18 | Integrated optimization; heuristic algorithm; combinatorial optimization; optimization model; multi-objective optimization; scheduling; production scheduling; dynamic; algorithm; model; intelligence algorithm; decision-making; Nash equilibrium; Portfolio; bounded rationality; decision analysis; decision-making method; optimization |
| 7 | 11 | emergency management; machine design; incentive mechanism; cooperative game; modelling; cloud computing; trust; creative commons; emergency; the social network; unconventional emergencies |
| 8 | 7 | The information system; data mining; business intelligence; disruption management; Internet of things; electronic commerce; risk evaluation |

### *3.2. Social network analysis*

*Social network map of co-occurrence matrix* we first got the social network map of co-occurrence matrix (Fig.2). It could intuitively reveal the relationship of research themes of management. The size of nodes can reflect the frequency of keywords: the higher frequency of keyword, the larger size of node. The thickness of line is proportional to the closeness of connections between two keywords, the thicker line between two words, the closer relationship is. As shown in Fig 1, the Game Theory node has the biggest size representing Game Theory has the highest frequency of keywords. Supply Chain Management, Data Mining also have higher frequency.

The thinker lines between two keywords, such as Game Theory, Supply Chain Management, Data Mining and Business Intelligence, represent closer relationship. Conversely, we also easily find some keywords, such as Complex System, Knowledge Transform, has lower frequency whose size of node is small. And there is no line connecting to other keywords, indicating these keywords are in the margin of research field.

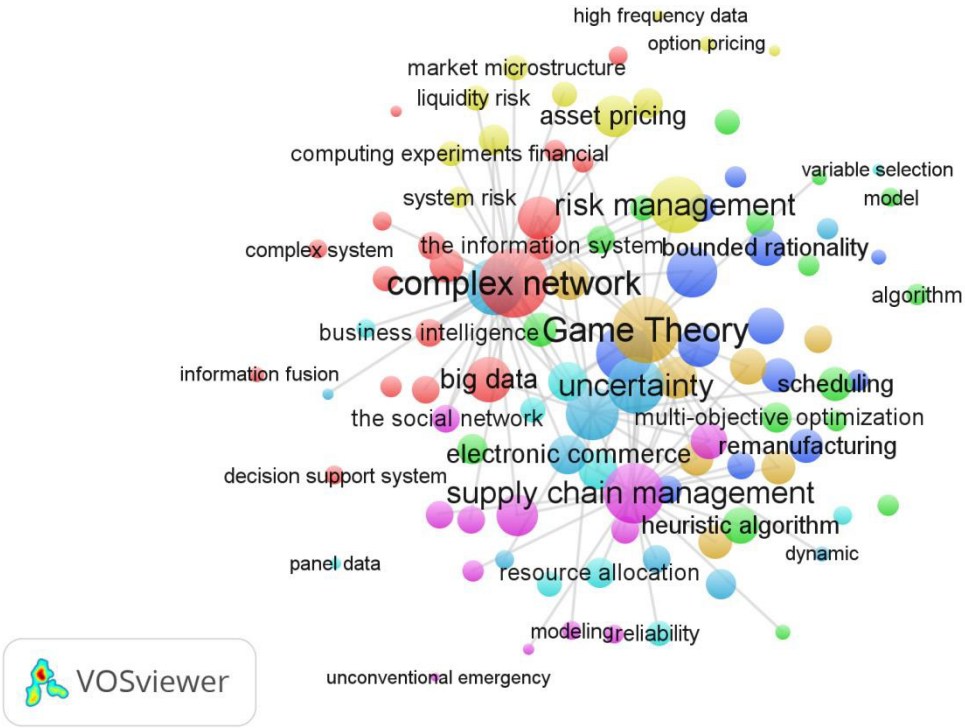


Fig. 2 social network map of co-occurrence matrix

To acquire more information of these keywords, we got a graph of density visualization in Fig. 3. In the item density visualization, items are indicated by their label in a similar way as in the network visualization. Each point in a map has a color that depends on the density of items at that point. By default, this color is somewhere in between red and blue. The larger the number of items in the neighborhood of a point and the higher the weights of the neighboring items, the closer the color of the point is to red. Conversely, the smaller the number of items in the neighborhood of a point and the lower the weights of the neighbouring items, the

closer the color of the point is to blue. From the density visualization, we may directly find that Game Theory, Supply Chain Management, Data Mining and Risk Management have a high density representing these keywords have a strong relationship with other keywords. And it is reasonable to assume that the higher the density, the more mature and well-developed study on the theme [3, 8]. However, we couldn't ignore a point that keywords in red area are fewer, most of keywords in green area. It transforms information to us that the core research fields are fewer, and many of these research fields tend to be immature.

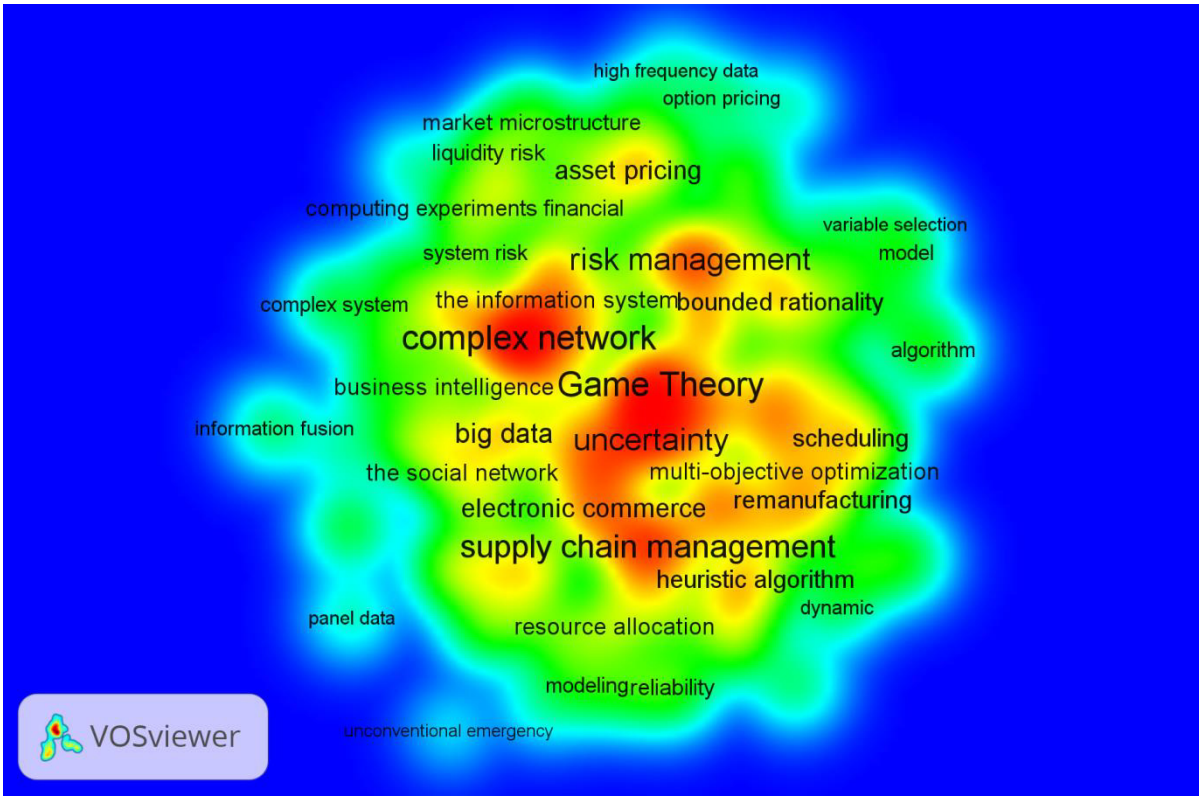


Fig. 3 density visualization

*K-core analysis of binary matrix* we got the second network by using the binary matrix, which was converted from the co-occurrence matrix in Ucinet with K-core analysis to find core-verge topics. As show in Fig. 4, five cores are identified by k-cores analysis. To display the cores clearly, different shapes mark degree of core or edge: up triangle nodes (k = 5) represent core themes of the network. Square nodes (k = 4), the secondary core themes. Down triangle nodes (k = 3) are between core and edge themes. Circle nodes (k=o) are edge themes. We can easily find that core themes include Game Theory, Supply Chain Management, Data Mining, which accord with the above analysis.

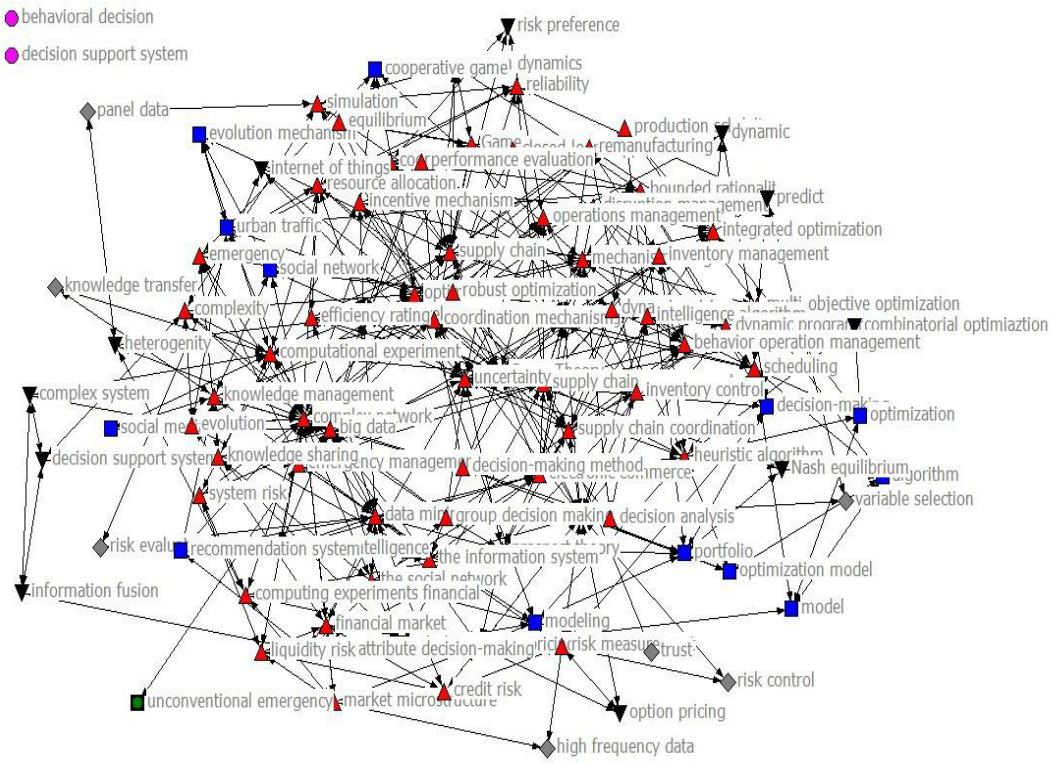


Fig. 4 K-core analysis of binary matrix

## 4. conclusion

In this study, we achieved research hot topics analysis of Management Science and Engineering in Chian using co-word analysis, which based on the data of Fund Project. Utilizing the tools of VOSviewer and Ucinet, we obtain a clear understanding on the development of Management Science and Engineering research in China from 2011 to 2015.

According to keyword frequency and k-cores, we identify the focus of research in Management Science and Engineering in China, including game theory, supply chain management, complex network, data mining, optimize, supply chain risk management, Data Envelopment Analysis, computational experiment, and pricing. The cluster analysis for 100 keywords shows that the research fields of Management Science in China are varied. We identify 8 topics cluster of Management, which represent a research direction. Also, we studied the relationship between keywords directly from social network analysis. Coincide with above analysis, “Game Theory”, “Supply Chain Management” and “Data Mining” are hot topics and developed more matured. However, the evolution trends in the density visualization suggest that many research fields in Management Science in China are still immature. It should be noted that the well-developed and core research field in Management Science in China are fewer. Thus, many research fields need to be further studied.

This research is just a preliminary and still has limitations. A further perfect research need to action in the future. For example, there is need to increase the number of research to make our research result precisely. And some research methods, such as cluster, the network characteristics of centrality, density, and the core-periphery structure, will be a necessary to apply in the future research.

## Acknowledgements

This research has been supported by grants from the Youth Innovation Promotion Association of the Chinese Academy of Sciences and the Key Research Program of Institute of Policy and Management, Chinese Academy of Sciences.

## References

[1] D.S. Wu, Y.J. Xie, Q.Z. Dai, J.P. Li, A systematic overview of operations research/ management science research in Mainland China: Bibliometric analysis of the period 2001-2013, Asia-Pacific Journal of Operational Research, forthcoming (2016).
[2] Y. Yang, M. Wu, L. Cui, Integration of three visualization methods based on co-word analysis, Scientometrics, 90 (2012) 659-673.
[3] X.Y. An, Q.Q. Wu, Co-word analysis of the trends in stem cells field based on subject heading weighting, Scientometrics, 88 (2011) 133-144.
[4] Q.J. Zong, H.Z. Shen, Q.J. Yuan, X.W. Hu, Z.P. Hou, S.G. Deng, Doctoral dissertations of Library and Information Science in China: A co-word analysis, Scientometrics, 94 (2013) 781-799.
[5] P.B. Musgrove, R. Binns, T. Page-Kennedy, M. Thelwall, A method for identifying clusters in sets of interlinking Web spaces, Scientometrics, 58 (2003) 657-672.
[6] T. Dehdarirad, A. Villarroya, M. Barrios, Research trends in gender differences in higher education and science: a co-word analysis, Scientometrics, 101 (2014) 273-290.
[7] G.Y. Liu, J.M. Hu, H.L. Wang, A co-word analysis of digital library field in China, Scientometrics, 91 (2012) 203-217.
[8] J. Hu, Y. Zhang, Research patterns and trends of Recommendation System in China using co-word analysis, Information Processing & Management, 51 (2015) 329-339.
[9] B. Lee, Y.I. Jeong, Mapping Korea's national R&D domain of robot technology by using the co-word analysis, Scientometrics, 77 (2008) 3-19.
[10] X. Xu, A.M. Tan, S.X. Zhao, Funding ratios in social science: the perspective of countries/territories level and comparison with natural sciences, Scientometrics, 104 (2015) 673-684.
[11] X. Wang, D. Liu, K. Ding, X. Wang, Science funding and research output: a study on 10 countries, Scientometrics, 91 (2012) 591-599.
[12] L. Leydesdorff, L. Vaughan, Co-occurrence Matrices and their Applications in Information Science: Extending ACA to the Web Environment, Journal of the American Society for Information Science & Technology, 57 (2006) 1616-1628.
[13] S. Ravikumar, A. Agrahari, S.N. Singh, Mapping the intellectual structure of scientometrics: a co-word analysis of the journal Scientometrics (2005–2010), Scientometrics, 102 (2015) 929-955.