

# Author Co-citation Analysis of Information Science in China based on the CSSCI

Wang Feifei<sup>1</sup> and Qiu Junping<sup>2</sup>

<sup>1</sup> wangfei0256@sina.com <sup>2</sup> jpqiu@whu.edu.cn

School of Information Management, Wuhan University, Wuhan, China 430072

## Abstract

As an important method of information researches, author co-citation analysis (ACA) has a very wide range of applications in exploring subject structure and research status and trends. To the 70947 cited articles in the information science field included by CSSCI from 2000-2009, this paper makes an author co-citation analysis which is different from the traditional mode, that is, using a combination of citation analysis, factor analysis and social network analysis (SNA) to further explore the relationship of the highly-cited authors in Chinese information science field and attempt to explore of the situation of authors influences and subject structure in the field.

## Introduction

In 1990, McCain divided the process of Author Co-citation Analysis (ACA) into 6 steps: selecting authors, retrieving co-cited frequency, constructing co-cited matrix, conversion of Pearson correlation coefficient matrix, multivariate analysis and results interpretation, validity analysis <sup>[1]</sup>. The above 6 steps can be called traditional ACA model which provides a good example in revealing the present and changing situation of subject structure for the later scholars. To the cited articles in the information science field included by CSSCI from 2000-2009, this paper makes an author co-citation analysis which draws on this model but is different from the traditional idea, that is, automatically generating the co-cited matrix through the program, then using a combination of citation analysis, factor analysis and social network analysis (SNA) to further explore the relationship of the highly-cited authors in information science field of China and attempt to explore of the situation of authors influences and subject structure in the field.

## Data and methodology

### *Basic Data*

The study is based on bibliographic data retrieved from the Chinese social sciences citation index (CSSCI). The data contains all types of documents with G35 as Chinese class Numbers in the field of Information science and Intelligence work from 2000-2009. There were 9208 documents including 9189 articles with authors, 8603 articles with references. After deleting duplicate records and comments, we finally obtain data sample of 8567 articles.

The total number of first authors is 4806. The top 30 prolific authors (publishing 15 or more papers) are shown in Table 1. There are 31470 cited authors in 70947 references. The 24 most highly cited authors (cited frequency  $\geq 100$ ) are shown in Table 2.

Table 1 The top 30 prolific authors

<i>Author</i>	<i>Articles</i>	<i>Author</i>	<i>Articles</i>	<i>Author</i>	<i>Articles</i>
Qiu Junping	68	Ma Haiqun	21	Wen Youkui	18
Wang Zhijin	68	Han Yi	21	Yang Jianlin	18
Peng Jingli	35	Ma Feicheng	21	Zhou Ning	17
Wang Yuefen	26	Wen Tingxiao	21	Zhu Qinghua	16
Jia Junzhi	25	Jiang Chunlin	20	Hua Bolin	16
Zhou Jiuchang	24	Wu Xiaowei	19	Liu Yuzhao	16
Huang Xiaobin	23	Bi Qiang	18	Jiao Yuying	15
Chen Feng	22	Pang Jing'an	18	Zhao Rongying	15
Xia Lixin	22	He Shaohua	18	Sha Yongzhong	15
Xiao Yong	21	Qin Tiehui	18	Lai Maosheng	15

Table 2 The 24 most highly cited authors

<i>Author</i>	<i>Cited Frequency</i>	<i>Author</i>	<i>Cited Frequency</i>	<i>Author</i>	<i>Cited Frequency</i>
Qiu Junping	782	Lai Maosheng	202	Cheng Feng	115
Bao Changhuo	529	Hu Changping	186	Peng Jingli	114
Ma Feicheng	446	Liang Zhanping	170	Miu Qihao	114
Wang Zhijin	292	Su Xinning	167	Wu Weici	108
Yan Yimin	286	Liu Zhihui	132	Wen Youkui	107
Zhang Qiyu	219	Lu Taihong	119	Jiao Yuying	106
Zhang Xiaolin	216	Yue Jianbo	116	Qin Tiehui	102
Wang Chongde	214	Meng Guangjun	116	Ma Haiqun	100

### *Methodologies*

Author co-citation analysis (ACA) first introduced by WHITE & GRIFFITH <sup>[2]</sup> has attracted much attention from different fields since then. Many authors have been devoting themselves to working in its application in study analysis of knowledge structure of specific areas. White summarized the ACA as a standard set of ideas, that is, using factor analysis, cluster analysis and multidimensional scaling analysis methods to do this kind of study later on. In China, most scholars used to explore the classification of cited-authors based on relationship of co-citation combined using clustering analysis with multi-dimensional scale analyses in order to find out the research hotspots and trends in a given area. In view of data characteristics and research purpose, factor analysis was single employed to explore the relationship of the

highly-cited authors in information science field of China in this article <sup>[3]</sup>, which will provide a new perspective and useful supplement for the traditional ACA model. This process will be completed by the SPSS.

As a set of theory and method to analyze the structure and properties of various relations in social network, Social Network Analysis (SNA) was mainly used to analyze the relationship mode among the actors, which was also proved to be successful in studies of citation relationship research <sup>[4]</sup>. We will apply the SNA in the analysis of co-cited network of highly cited authors in information science field, which will be completed by the software of UCINET and Netdraw. The co-occurrence analysis in this paper will be completed by the VBA program.

Essentially, factor analysis, clustering analysis and multi-dimensional scale analyses are all based on the classification ideology. The combination of later methods in traditional ACA model can be regarded as a graphical classification and limited visualization. However, the combination of factor analysis and visualization software (such as Netdraw) in SNA can produce a more appropriate and more detailed classification model, in which we can get the discrimination and relevance between the different categories in order to help researchers to do more accurate analysis on knowledge structure of a particular field. In addition, author co-cited relationship can be considered as a network and help to explore the situation of author influences by the analysis of network structure, coherency and centrality drawing on the idea of social networks, which will be wished to provide a reference and help to the research promotion and discipline development.

## Results and discussion

### *Co-cited matrix construction*

According to the “80/20” law, we select 101 highly cited authors (cited frequency  $\geq 40$ ) as the object of ACA, whose cited frequency accounts for 20% (a total of 48053 times). Firstly, the original data should be converted to co-citation matrix. As we know, the diagonal values in author co-citation matrix make great impact on the correlation coefficient calculation such as Pearson's  $r$ , however, for which explanations varies and some are even subjective. We defined maximum+1 as diagonal value in order to emphasize the intimacy degree of an author with himself <sup>[5]</sup>.

Further on the analysis of author co-citation pairs, we find that 5050 author pairs formed 9886 times co-citation relationship, the average number close to 2 times / pair which shows that the overall co-citation intensity between authors is relatively low. In addition, there are 2680 pairs of authors without co-citation relationship, accounted for 53%; there are 234 pairs of authors with more than 10 co-cited frequencies, only accounted for 4.63%, including 7 pairs of authors with more than 50 co-cited frequencies. This indicated that co-citation intensity of highly cited authors is relatively dispersed and there is no unified model of research influences in Chinese information science field; the specialties impacted by authors are also more dispersed which lack of relative stability.

### *Analysis of author co-cited network structure*

According to network structure analysis by the Ucinet, we find that the density of author co-cited network is 1.9576, which indicates that the network connectivity lies in the middle level of general social networks. There are some nodes highly connected each other, whereas the majority nodes don't connected others. According to the centrality measure (i.e., table 3), Ma Feicheng and Qiu Junping have most co-cited frequencies with others, which indicates that they get the highest degree of centrality, who locate in the heart of the network and have the most significant impact on Chinese information science in the last decade. Both of them get

the highest centrality of betweenness and closeness which indicates that these two scholars have extremely high control over the network connectivity and information flow.

Table 3 The centrality measure of author co-cited network nodes

No.	Author	NrmDegree	Author	nBetweenness	Author	nCloseness
1	Ma Feicheng	14.048	Ma Feicheng	3.795	Ma Feicheng	96.154
2	Qiu Junping	13.940	Qiu Junping	2.954	Qiu Junping	92.593
3	Yan Yimin	10.542	Wang Zhijin	2.385	Wang Zhijin	90.090
4	Wang Zhijin	9.952	Bao Changhuo	1.920	Yan Yimin	84.034
5	Bao Changhuo	8.048	Yan Yimin	1.689	Bao Changhuo	82.645
6	Liang Zhanping	7.554	Wang Chongde	1.655	Lai Maosheng	82.645
7	Wang Chongde	6.229	Lai Maosheng	1.645	Wang Chongde	81.301
8	Lai Maosheng	6.000	Su Xinning	1.468	Yue Jianbo	81.301
9	Yue Jianbo	5.687	Yue Jianbo	1.397	Hu Changping	80.645
10	Liu Zhihui	5.205	Hu Changping	1.248	Liang Zhanping	80.000

The result of correlation analysis of author cited frequencies and three centrality indexes shows that there are significant correlations between them (i.e., table 4). The above analysis may indicate that highly cited authors have great impacts in co-cited network of Chinese information science and play an important role for the promotion of discipline development.

Table 4 Correlation analysis of author cited frequencies and centrality indexes

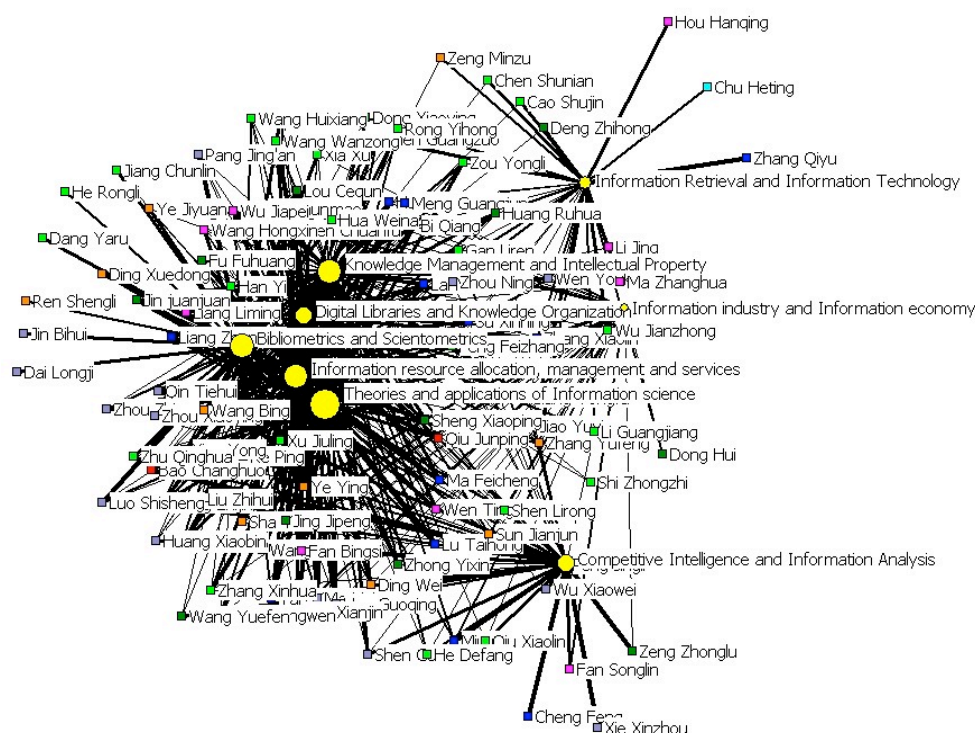
<i>Pearson Correlation</i>	<i>Cited Frequency</i>	<i>Degree</i>	<i>Betweenness</i>	<i>Closeness</i>
<i>Cited Frequency</i>	1	0.8537**	0.8492**	0.7073**
<i>Degree</i>	0.8537**	1	0.9098**	0.8814**
<i>Betweenness</i>	0.8492**	0.9098**	1	0.8970**
<i>Closeness</i>	0.7073**	0.8814**	0.8970**	1

\*\*. Correlation is significant at the 0.01 level

### Factor analysis

8 factors explained 85.591% of the total variance were extracted by principal component analysis (PCA) and oblique rotation (SPSS Oblimin). An oblique rotation can produce two distinct matrices: the structure matrix and the pattern matrix. Loadings in the pattern matrix represent the unique contribution of individual authors (variables) to specialties (factors), whereas loadings in the structure matrix, which are "simple correlations between variables and factors," are determined both by an author's unique contribution to each factor and by the correlation among factors [6]. Therefore, we will select the structure matrix to do further analysis. We visualized the factor loading matrix as two-dimensional map by the Netdraw. On the map (i.e., figure 1), authors are represented by square nodes and factors by circular nodes. The size of a factor node corresponds to the sum of the loadings on this factor by all authors

who load sufficiently on it (i.e., with a value of 0.2 or higher in this case). The width of a line that connects an author with a factor is proportional to the loading of this author on this factor. The color of an author node indicates the number of factors that this author loads on with a value of at least 0.2 each. Figure 1 shows that Shi Zhongzhi becoming isolated nodes doesn't belong to any factor after limited the loading.



**Figure 1. Two-dimensional map of factor analysis results**

According to the results of factor analysis, we can divide information science research in China into 8 specialties: Theories and applications of Information science, Bibliometrics and Scientometrics, Competitive Intelligence and Information Analysis, Information Retrieval and Information Technology, Information resource allocation, management and services, Digital Libraries and Knowledge Organization, Knowledge Management and Intellectual Property, Information industry and Information economy. But different specialties involved in different highly-cited authors and author contributions to the specialty are quite different, too (i.e., table 5). Specialties 1 and 2 have the most of primary loading authors, followed by specialties 3 and 4. As a cross research topic, specialty 8 has the least of primary loading authors and their loadings are relatively low, which indicates that this specialty lying in a marginalized scope is not the research focus in the field.

**Table 5 Factors and their labels of ACA**

No.	Label	Size	Number of primary loading	Total loadings	Average loadings	Highest loading	Primary loading authors
1	Theories and applications of Information science	62	29	33.8863	0.5466	0.9536	Liu Zhihui, Yan Yimin, Liang Zhanping, Xin Juanjuan, Ma Feicheng
2	Bibliometrics and Scientometrics	53	24	24.0618	0.4655	0.8478	Luo Shisheng, Pang Jing'an, Qiu Junping, Ding Xuedong, Wang Chongde
3	Competitive Intelligence and Information Analysis	37	15	17.2248	0.4540	0.9309	Bao Changhuo, Chen Feng, Peng Jingli, Xie Xinzhou, Miu Qihao
4	Information	21	14	10.5508	0.5024	0.8561	Zhang Qiyu, Ma Zhanghua, Chen

	Retrieval and Information Technology						Shunian, Cao Shujin, Houh Hanqing
5	Information resource allocation, management and services	56	6	26.2713	0.4691	0.8756	Huo Guoqing, Yue Jianbo, Wu Jiawei
6	Digital Libraries and Knowledge Organization	43	6	15.7458	0.3662	0.8631	Jiang Yongfu, Sheng Xiaoping, Zhang Xiaolin
7	Knowledge Management and Intellectual Property	66	5	25.3585	0.3842	0.7526	Ke Ping, Chen Chuanfu, Ma Haiqun
8	Information industry and Information economy	16	1	4.2314	0.2645	0.2801	Sun Jianjun

## Conclusions

According to analysis of co-cited relationship of highly-cited authors in Chinese information science field, we can draw some conclusions as follows:

- Qiu Junping, Ma Feicheng, Yan Yimin, Wang Zhijin, Liang Zhanping and Wang Chongde who have the highest co-cited frequencies with each other have produced the greatest impact in the development of the field. The co-citation intensity of highly cited authors is relatively dispersed and there is no unified model of research influences in the field; the specialties impacted by authors are also more dispersed which lack of relative stability.

- Ma Feicheng and Qiu Junping have most co-cited frequencies with others and the highest centrality of degree, betweenness and closeness which indicates that they have extremely high control over the network connectivity and information flow. There are significant correlations between authors cited frequencies and three centrality indexes which indicates that highly cited authors have great impacts in co-cited network of Chinese information science and play an important role for the promotion of discipline development.

- According to factor analysis, information science research in China can be divided into 8 specialties involving into library and information research located in the left part and information technology and application research located in the right part. This structure is similar to the research results of the word analyzed by White in 2003<sup>[7]</sup>. The former refers to 5 specialties, such as Theories and applications of Information science, Bibliometrics and Scientometrics, Information resource allocation, management and services, Digital Libraries and Knowledge Organization, Knowledge Management and Intellectual Property. The later refers to 3 specialties, such as Information Retrieval and Information Technology, Competitive Intelligence and Information Analysis, Information industry and Information economy.

## References

- MCCAIN, K. W. Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, 1990,41 (1) : 433–443.
- WHITE, H. D., GRIFFITH, B. Author cocitation: A literature mesaure of intellectual structures. *Journal of the American Society for Information Science*, 1981,32 (3) : 163–171.
- Dangzhi Zhao, Andreas Strotmann. Information Science during the First Decade of the Web: An Enriched Author Cocitation Analysis. *Journal of the American Society for Information Science and Technology*, 2008,59(6):916–937.
- XU,Y.Y.,ZHU,Q.H. Demonstration Study of Social Network Analysis Method in Citation Analysis. *Information Studies:Theory & Application*,2008,31(2):184-188.

- QIU, J. P., MA, R. M., LI, Y. J. New thoughts about co-citation analysis, *Journal of the China Society for Scientific and Technical Information* (in Chinese), 2008,27(1):70-74.
- Hair,J.F.Anderson,R.E.,Tatham,R.L.,&Black,W.C. *Multivariate data analysis*(5th ed.). Upper Saddle River,NJ:Prentice Hall,1998.
- White H D. Pathfinder networks and author co-citation analysis: A re-mapping of paradigmatic information scientists [J]. *Journal of the American Society for Information Science and Technology*, 2003, 54(5): 423-434.