

Study on Author Cooperation Relationship Based on Data Mining

Hongfei Sun

School of Economics and Management
Northeast Dianli University, NEDU, Jilin, China
Sunny_bird@126.com

Fang Wang, Wei Hou

School of Economics and Management
Northeast Dianli University, NEDU
Jilin, China

Abstract—In order to explore authors' cooperation relationship of application of research methods of information science in our country, this paper made a detailed statistics and analysis based on data mining on all the papers of 19 kinds of core journals about Library and Information Science from 2008 to 2012. Through the analysis of existing core authors' cooperative relationship and potential relationship, the author summarizes the characteristics and the deficiencies of the authors' cooperative relationship and provides an important basis for cooperation and exchanges in the field.

Keywords- data mining; author cooperation relationship; social network analysis

I. INTRODUCTION

With the increasingly frequent scientific cooperation, exploring the cooperative relationship in various fields has become an important research content of literature metrology. Yun Fu etc used the small group analysis, the degree of cluster analysis and centrality analysis of the social network to explore the authors' cooperative relationship in the field of science from 2004 to 2008^[1]. Dan Wang explored the role of the network structure analysis in the field of authors' cooperative relationship analysis^[2]. Hongyong Yang and Siying Zhang established a complex network model of the authors' cooperation, and conducted the simulation research with the data of authors' cooperation of "information science" from 2001 to 2006; finally they demonstrated the effectiveness of the model^[3].

In recent years, with the rise of literature metrology and other quantitative methods, the application of information science research methods is more and more brought to the attention of the scholars. Therefore, through the analysis of authors' cooperative relationship in the field of application of research methods of Information Science, it could reveal the existing and potential cooperation relationship of the core authors in this area, and it also could explore the characteristic, deficiencies and cooperation potentiality of authors' cooperation. As a result, it could provide an important basis for the promotion of cooperation and communication and for the promotion of the healthy and rapid development of authors' cooperation in the field.

II. AUTHORS' COOPERATIVE SITUATION ANALYSIS

A. Authors' Cooperation Statistics

Through the statistical analysis, we found that the rate of authors' cooperation in the field of application of research methods of information science reached 73.17% in recent five years, indicating that authors' cooperation was

more generally in this area, and it also showed information exchange and knowledge sharing among authors were more frequently. The results showed that: In all of 4292 authors' cooperation papers, 2 to 3 authors' cooperation accounted for 89%, 2 to 4 authors' cooperation accounted for 97.5%, which reflects the way of authors' cooperation is small team cooperation for outstanding performance in the field of application of research methods of information science in China.

B. Core Authors

In recent five years, research achievements in the field of research methods applications of information science, among which were 6,946 authors in total and co-authors were 6208, accounting for 89.37% of all authors. This indicates that authors have a higher degree of cooperation in the field of research methods applications of information science. Due to the number of co-authors is relatively high, according to Pryce's theory: someone published N papers or more papers would be called outstanding scientist, namely core authors. Calculating formula is $N = 0.749(n_{\max})^{1/2}$, in this formula, n_{\max} means the largest volume Number of papers involved in co-author posting papers. In this paper, the maximum amount of posting papers is Junping Qiu and the number of his papers of co-author is 81, whereby we get $N = 6.741$. Therefore, this article selects whose posting papers is 7 or above 7 as the core authors, as a result, we get the core authors in the field of research methods applications of information science.

In the group of core authors, Junping Qiu who works in Wuhan University ranked first, and he published 81 papers of this field in total in recent 5 years. In addition, well-known scholars in the field of information science, such as Qinghua Zhu, Yufeng Zhang, Feicheng Ma and so on. They are all in front of others. These also indicate that their teams make a great contribution to the applications of research methods of information science.

III. ANALYSIS OF CORE AUTHORS' COOPERATION

This paper targeted to choose the core authors' cooperation in this field as the research object. At the same time, it constructed the core authors' cooperation matrix, and used the method of social network to do visual analysis. As a result, we know the area of the authors' cooperation network, network structure and the important authors in the network.

A. Based on the Undirected Matrix Analysis of Core Authors' Cooperation Analysis

First of all, according to social network analysis method, this paper built undirected matrix S^1 (190×1190) of the core authors' cooperation by using Excel VBA code. The matrix's row i and column j represent the core authors, numerical number $S^1_{i,j}$ in the matrix represents the number of cooperation of corresponding row and column. The specific calculation formula is $S^1_{i,j} = f(k_{i,j})$. Among them $k_{i,j} \in (\{k_i\} \cap \{k_j\})$, $\{k_i\}$ are the papers collection published by the author i , $\{k_j\}$ are the papers collection published by the author j , so $k_{i,j}$ are the papers published by the author i and the author j together. Then let the matrix into Netdraw to do visual analysis. Finally, we concluded undirected network diagram of the core authors' cooperation, as shown in Fig. 1.

Each dot in the figure represents a core author, and the size of the dot represents the number of papers, which cooperated by core authors and other core authors; The color of the dot is the result of the K-cores analysis, the same color on behalf of the same class K nuclear; the thickness of the line connecting between dot represents the number of papers which published by different core authors' cooperation. Diagram layout uses the Netdraw's peculiar "Spring-embedding" layout, and the distance between the points can be approximately regarded as the cooperation distance of the core authors.

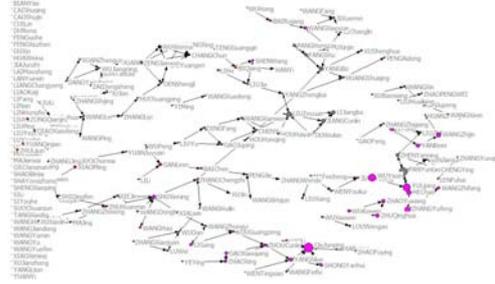


Figure 1. Undirected network diagram of the core authors' cooperation

As can be seen in Fig. 1, the partnership of the core authors' group is a non-connected network, and there are 51 core authors for outlier among them. This illustrates the 51 core authors have no cooperative relationship with other core authors (not on behalf of the core authors have no cooperative relationship with non-core authors). From the figure, we also can clearly see that the average density of the entire network is 0.0297, this indicates the cooperation between core authors often limited to local, especially for small academic team cooperation, and lacking of large and broad cooperation.

Then, we put undirected matrix of the core authors' cooperation to the Ucinet, and analyze the centrality of core authors in cooperation network. Analysis from three aspects: point centrality, closeness centrality and betweenness centrality [4]. Three kinds of centrality to the author of the top 10, as shown in table I.

TABLE I. NODE CENTRALITY OF UNDIRECTED NETWORK OF THE CORE AUTHORS' COOPERATION (TOP 10)

Point Centrality	Degree	Closeness Centrality	Closeness	Betweenness Centrality	Betweenness
Junping Qiu	3.704	Xinning Su	0.691	Xinning Su	3.901
Zeyuan Liu	3.704	Chengzhi Zhang	0.691	Chengzhi Zhang	3.372
Xinning Su	3.175	Lixin Xia	0.691	Lixin Xia	2.617
Bo Yang	3.175	Lin He	0.69	Zhongyi Wang	2.522
Chunlei Zhou	3.175	Hengmin Zhu	0.69	Hanqing Hou	2.522
Yuntao Pan	2.646	Zhongyi Wang	0.69	Lin He	2.415
Yuangen Lai	2.646	Jing Xie	0.69	Chun Lei	2.308
Jianxun Zeng	2.646	Guangzeng Kou	0.69	Guangzeng Kou	2.297
Xinjin Fu	2.646	Huilin Wang	0.69	Bo Yang	2.077
Yanning Zheng	2.116	Hanqing Hou	0.69	Xinjin Fu	1.351

From the point centrality, we can see Junping Qiu and Zeyuan Liu ranking first of the same column. This shows their scope of cooperation more extensive, and had a relationship with many core authors. From the closeness centrality, we can see the distance between Xinning Su, Chengzhi Zhang, Lixin Xia and all the other core authors is nearest, and combined with Fig. 1, we can see that they are in the middle of a large ring network, connecting the key points in each cooperation sub-network. From the betweenness centrality, we can see Xinning Su, Chengzhi Zhang, Lixin Xia still located in the top three, this indicates that their control degree is higher for the entire network cooperation, and only passing them can be connected with other sub-network, this also can be verified from the Fig. 1.

B. Based on the Directed Matrix Analysis of the Core Authors' Cooperation

In the analysis of the authors' cooperation, many scholars tend to ignore the problem of the authors' ranking sequence, and regard the author cooperation as undirected network. Undirected network can only identify cooperation's scale and frequency of each author's cooperation team, but it is hard to see the author's information such as position, contribution of the team cooperation. However, directed network diagram can make up for the deficiency in this area. In this paper, we base on the core authors' partnership to build the matrix S^2 , S^3 (190×190) in this paper. Row i and column j of the matrix represent the core authors, and the numerical number $S^2_{i,j}$,

$S_{i,j}^3$ in the matrix represent the number of authors' cooperation of the corresponding row and column. The specific calculation formula is $S_{i,j} = f(k_{i,j})$ (Regardless of $S_{i,j}^2$ or $S_{i,j}^3$ are all applicable to the formula). Among them $k_{i,j} \in (\{k_i\} \cap \{k_j\})$, $\{k_i\}$ are the papers collection published by the author i, and $\{k_j\}$ are the papers collection published by the author j. In $S_{i,j}^2$, $k_{i,j}$ are the papers published by the author i and the author j, and the sequence is the author i in the first, then the author j; In $S_{i,j}^3$, $k_{i,j}$ are the papers published by the author i and the author j, and the sequence is the author j in the first, then the author i; $f(k_{ij})$ is the number of. Then we put the matrix into Netdraw to do visual analysis, and get directed network diagram of two core authors' cooperation, as shown in Fig. 2, 3.

In Fig. 2 and Fig. 3, the meaning of dot size, color, thickness of line connecting and the location is the same as shown in Fig. 1. Through comparing Fig.2 and Fig. 3, which can be more clearly and objectively to analyze the authors' position, contribution, cooperation scale and frequency of each author's team cooperation.

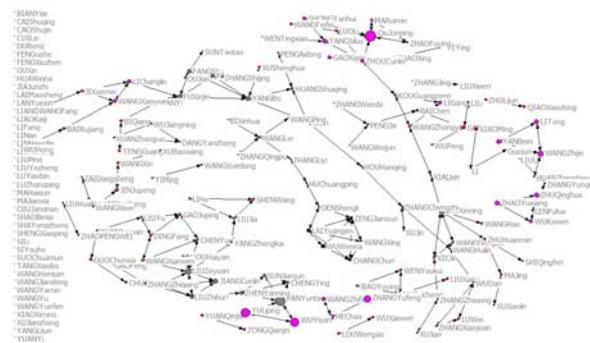


Figure2. Directed network diagram of the core authors' cooperation (1)

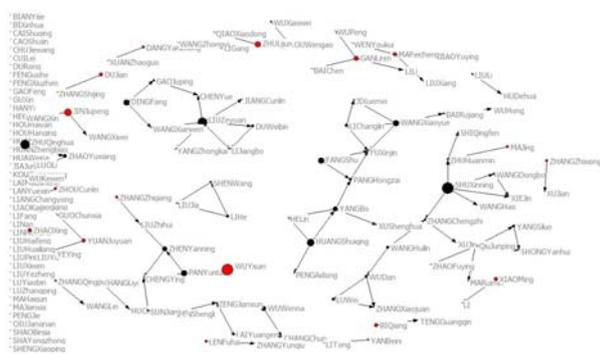


Figure3. Directed network diagram of the core authors' cooperation (2)

IV. THE POTENTIAL PARTNERSHIP MINING OF THE CORE AUTHORS

The above analysis are all based on the partnership established of the papers published by authors together, but

there are still many authors with similar research field, research content can't cooperate with each other, but there exists a potential cooperation between them. So this article is based on the keywords published by the core authors, and the core authors-keywords coupling matrix is established to mine potential relationship between the core authors.

First of all, establishing the core authors-keywords coupling matrix S^4 . Row i and column j of the matrix represents the core authors, the numerical value of the matrix $S^4_{i,j}$ represents the times of core authors-keywords coupling of the corresponding row and column. The specific calculation formula is $S^4_{i,j} = f(k_{i,j})$. Among them $k_{i,j} \in (\{k_i\} \cap \{k_j\})$, but too many keywords are not good for mining of potential cooperation network, therefore this article selects the high frequency keywords whose frequency keywords are greater than 15 keywords [5], $\{k_i\}$ are the high frequency keywords sets published by author i, $\{k_j\}$ are the high frequency keywords sets published by author j, $k_{i,j}$ are the high frequency keywords sets used by author I and the author j together, $f(k_{ij})$ is the number of $k_{i,j}$ [6]. Based on the core authors-keywords

coupling matrix S^4 , we can draw potential partnership network diagram of the core authors, as shown in Fig. 4. The meaning of the color of the dot, location in the Figure is the same as Fig. 1. Just the size of the dot represents the number of the current core authors and other core authors' keywords coupling, this also means that the potential cooperation opportunities exist between the current core authors and other core authors. And thickness of line connecting represents the number of having the same keywords of the core authors and other core authors.

From the Fig. 4, we can see the average density of the entire network is 0.8026. This suggests wide cooperation with larger possibility and potential between the core authors. Through the analysis of three kinds of centrality: point centrality, closeness centrality and betweenness centrality, it is concluded that the authors of the top 10, as shown in table II.

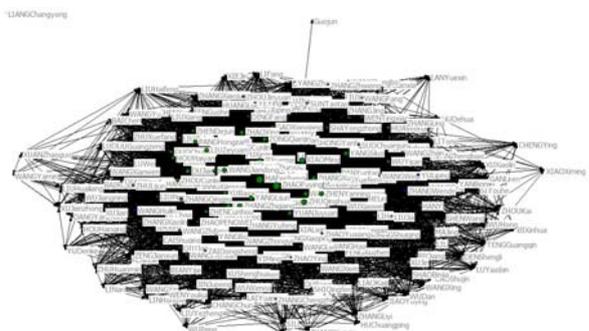


Figure4. Potential partnership network diagram of the core authors

TABLE II. NODE CENTRALITY OF POTENTIAL PARTNERSHIP NETWORK OF THE CORE AUTHORS (TOP 10)

Author	Degree	Author	Closeness	Author	Betweenness
Junping Qiu	88.889	Junping Qiu	47.487	Junping Qiu	1.894
Feicheng Ma	83.069	Feicheng Ma	46.21	Yufeng Zhang	1.329
Qinghua Zhu	79.894	Qinghua Zhu	45.542	Feicheng Ma	1.292
Yufeng Zhang	78.836	Yufeng Zhang	45.324	Qinghua Zhu	1.208
Jianjun Sun	78.307	Jianjun Sun	45.215	Rongying Zhao	1.175
Rongying Zhao	78.307	Rongying Zhao	45.215	Gang Li	1.165
Yuefen Wang	76.72	Yuefen Wang	44.893	Jianjun Sun	0.973
Gang Li	75.132	Gang Li	44.575	Yuefen Wang	0.93
Zhiqiang Zhang	75.132	Zhiqiang Zhang	44.575	Chunhou Zheng	0.89
Fuhai Leng	74.603	Fuhai Leng	44.471	Fuhai Leng	0.831

Taking three centrality indexes (point centrality, closeness centrality and betweenness centrality), we can see that Junping Qiu, Feicheng Ma, Qinghua Zhu, Yufeng Zhang, Jianjun Sun, Rongying Zhao, Yuefen Wang, Gang Li and Fuhai Leng are the key core authors of potential partnership network of the core authors.

V. CONCLUSION

In a word, high frequency, stable state and huge potential are the characteristics in the field of application of research method of information science. At the same time there are also existing deficiencies of large cooperation limitations, insufficient span of cooperation, fragile networks of cooperation and so on. Therefore, we suggest that academic institutions should be through platforms such as fund projects, academic conferences and so on, to promote authors' cooperation and exchanges in the field of application of research methods of information science.

REFERENCES

- [1] Fu Yun, Niu Wenyuan, Wang Yunlin, Li Ding. Analysis on the author cooperation network in the field of science of science: A case of Science Research Management from 2004 to 2008[J]. Science Research Management, 2009(6):41-46.
- [2] Wang Dan. Research of Ego-centric Network Analysis in Scientific Collaboration-----A Case Study for Carbon Nanotubes Domain[J]. Information Science, 2008(11):1735-1739.
- [3] Yang Hongyong, Zhang Siying. Author Collaboration Complex Network Model[J]. Information Science, 2008(05):774-778
- [4] Liu Jun. Lectures on Whole Network Approach----A Practical Guide to UCINET[M]. Shanghai: Shanghai People's Publishing House, 2009
- [5] Qiu Junping, Chen Mupei. Study on the relationship between the cooperation in the field of the science of National Metrology[J]. Information Studies: Theory & Application, 2012(11):56-60
- [6] Liu Zhihui, Zhang Zhiqiang. Research methods and empirical analysis of the authors' keywords coupling[J]. Information Science, 2010(02):268-275