Network Positions, Similarities, Prior Collaboration Experiences and the Evolution of Co-authorship Networks

Zhi-hong Song[1]
Qianyun Jia[2]
Dong-mei Lee[3]

[1]Institute of Management and Decision, Shanxi University, Taiyuan City, Shanxi Province, P.R.China.
[1]*Email: songzhihong@sxu.edu.cn*
[2]School of Economics and Management, Shanxi University, Taiyuan City, Shanxi Province, P.R.China.
[3]School of Mathematical Sciences, Shanxi University, Taiyuan City, Shanxi Province, P.R.China.

Abstract

Based on the articles published by Chinese inland scholars in 35 SCI and SSCI journals in the field of information science, the paper examines the impact of the network positions, similarities and prior collaboration experiences of core authors on the evolution of international co-authorship networks. The empirical results show that, (1) central positions of authors in co-authorship networks may lead to the establishment co-authorships; (2) The similarities and prior collaboration experiences may also lead to the formation of co-authorships. Specifically, authors within the same affiliations are more likely to establish co-authorships, and prior collaboration experiences will lead to the continuation co-authorship.

1. Introduction

Collaborations between scholars have become an increasingly more important mode of knowledge generation in almost all scientific disciplines (Wuchty, Jones, & Uzzi, 2007) particularly in science-based fields (Cantner & Rake, 2014). Collaborations may contribute to the pool and exchange of knowledge and resources as well as the generation of new ideas (Powell & Brantley, 1992; Powell & Grodal, 2005; Powell, Koput, & Smith-Doerr, 1996). It may also help to reduce the negative effects on scientific research output due to the limitations of individual scholar's cognitive ability, and improve the scientific research quality and research output. As an important form of collaboration between scholars, the quantity and quality of co-authored papers may reflect the strength and intensity of collaboration, which can be used to measure the development of academic network and research output (see (Hoekman, Frenken, & Oort, 2009; Katz & Martin, 1997; Laudel, 2002)) for a discussion).

Extant literature on co-authorship networks mainly focuses on three different streams of research: (1) the first research stream focuses on the patterns and characteristics of co-authorship networks. such as small world phenomenon (Newman, 2001) i.e. the network frequently follows, at least asymptotically, a scale-free power law (Barabási, 2003; Barabasi & Albert, 1999); Newman (2001) uses social network analysis to investigate the network of scientists in the field of biomedical science, physics and computer science. The results show that collaboration networks among scholars are characterized by small world and clustering phenomena; (2) the second research stream focuses on the influence of co-authorship on scientific research output. Katz and Martin (1997) show that collaborative research improves the quality of academic papers. (Goldfinch, Dale, & Jr, 2003) further point out that periphery scholars in co-authorship networks may have access to communicate with core scholars in the networks by collaboration and thus can improve their research output and scientific impact; (3) the third research stream mainly investigates the driving factors of co-authorship. Boschma and Frenken (2010) argue that authors with higher degree of connectedness will be more attractive, and individual authors prefer to collaborate with those who are proximate or similar with

themselves. The decisions to form new ties or break up old ties are driven by the agents' ambitions to maintain or reconfigure their network positions in order to benefit from opportunities created by the network (Zaheer & Soda, 2009). The driving factors for establishing co-authorship include accessing complementary resources, sharing research equipments and improving the academic status (Archibugi & Coco, 2004).

In conclusion, extant literature on the evolution of co-authorship networks mainly focuses on the driving factors which form or break up ties and such research are usually qualitative. Few researches thus far have investigated quantitatively the evolution of co-authorship networks, especially the endogenous factors driving the network evolution. This article uses MRQAP (Multiple Regression Quadratic the Assignment Procedure) analysis method and discusses how network positions, similarities, and prior collaboration experiences influence the evolution of co-authorship networks.

2. Theoretical Background and Research Hypotheses

One form of collaborations among authors can be measured by co-authored publications. If two or more authors collaborate to publish an article, co-authorships are established among these authors In order to describe factors driving the evolution of international co-authorship networks, we take co-authorship as the dependent variable and focuses on three factors which drive the formation of co-authorships.

The first driving factor which affects the formation of co-authorship is the focal author's position in co-authorship networks. The network position of an agent implies his/her attractiveness as a potential partner. Archibugi and Coco (2004) finds that an individual author is more willing to collaboration with core authors with the aim of improving their academic status. Rivera, Soderstrom, and Uzzi (2010) argues that authors who are in central positions in the network during previous period are more likely to attract new relations in later period. Over time, the phenomenon appears which are characterized by "rich-get-richer" (i.e, authors who are early entrants in the network may form a higher number of co-authorships than late entrants) (Cantner & Rake, 2014). In this case, network incumbents may increase their connectivity at the expense of newcomers. Therefore, the hypothesis is stated as follows.

Hypothesis 1: The central position of an individual author in co-authorship network in period t-1 may establish more collaborations in period t.

Secondly, authors who are more similar may form collaborations (Rivera et al., 2010). Luo and Deng (2009) argue that similarity of affiliations between authors may lead to the formation of collaborations among them. Jiang, Tao, and Santoro (2010) argue that authors who are in the same sector and carry out similar activities may share knowledge and experiences, therefore may lead to the formation of collaborations (Cohen & Levinthal, 1990). Therefore, individual authors who are in the same affiliations are more likely to establish collaborations, which leads to the following hypothesis

Hypothesis 2: Authors with the same affiliations in co-authorship network during period t-1 period may form collaborations during period t.

Finally, the closely-connected network may enhance trust among individual authors and increases the likelihood of collaboration in the future (Coleman, 1988). An individual author with prior collaboration experiences may have a clear understanding of the ability and reliability of partners, which may increase the chance of repeated collaboration. Gulati and Gargiulo (1999) confirms that previous link between two organizations are more likely to form new alliances. As for individual authors, prior collaboration experience may increase the possibility of later collaboration. So, we argue that the focal author in the network may maintain prior collaborations. Therefore, the hypothesis is stated as follows.

Hypothesis 3: The authors who have prior collaborations during period t-1 are more likely to establish co-authorships in period t.

3. Research Design
3.1. Date Collection

The paper chooses articles published by Chinese inland scholars in 35 SCI and SSCI journals in the field of information science as the data source, including MIS Quarterly, Journal of Information Technology, Journal of the American Medical Informatics Association, Information Systems Research, European Journal of Information Systems, International Journal of Information Management, Journal of Strategic Information Systems, Information Systems Journal, Telematics and Informetrics, Journal of the Association for Information Science and Technology, Journal of The Association for Information Systems, Journal of Knowledge Management, Information and Organization, Information Society, Journal of Librarianship and Information Science, Online Information Review, Information Technology & People, Aslib Proceedings, Information Technology and Libraries, Program-Electronic Library and Information Systems, MIS Quarterly Executive, Journal of Information Science, Information Technology for Development, Information Development, Journal of Global Information Technology Management, Data Base for Advances In Information Systems, Aslib Journal of Information Management, Information Technology & Management, Knowledge Management Research & Practice, Knowledge Organization, Information Research-An International Electronic Journal, Journal of Global Information Management, Information & Culture, Library and Information Science, Canadian Journal of Information and Library Science. Using the advanced search

option in Web of Science, the search tactics are as follows: "SO(journal's name)" and "AD" (author's address) for joint retrieval, such as "SO = MIS Quarterly AND AD = China". A total of 2032 articles were retrieved. According to the search results, the bibliographic information is collected, including the title, the name of author(s), affiliations and year of publication.

Because the paper chooses co-authorships by Chinese inland scholars(excluding Chinese Hong Kong, Chinese Macao and Chinese Taiwan) , we exclude 125 solo-authored papers; According to the authors' affiliations, 482 papers were deleted which are written by scholars at institutions in Chinese Hong Kong, Chinese Macao and Chinese Taiwan, leading to 1425 papers. Then, we use the Bibexcel to extract the authors' bibliographic information. We manually check the authors' names and affiliations to ensure that no duplicated authors exist in the sample, leading to 2870 authors.

Due the lack of enough data before 2006, the paper investigate factors which drive the evolution of co-authorship networks, we examine the network evolution during five periods (2006~2008, 2009~2010, 2011~2012, 2013~2014, 2015~2016).

Since the co-authored papers do not involve the direction of collaboration, the network is undirected. In addition, we do not distinguish the order of authors in publications. We take co-authorship as authors who are collaborate to publish in the same article.

3.2. Co-Authorship Network

Ucinet software is used to calculate the overall network, which involve 2870 nodes and 12716 sides Figure 1. The density of the co-authorship network is 0.0015, indicating the network is loosely connected. Accordingly, we select core authors and those tied with the core authors as the research sample.

The core authors in the co-authorship network are identified by network centrality measures, including degree centrality, betweenness centrality and closeness centrality. Among the above three measures on network centrality, we selected the top 20 authors in each measures, resulting 199 authors. The information on the authors' centrality measures, affiliations, and co-authors are used in the data analysis.
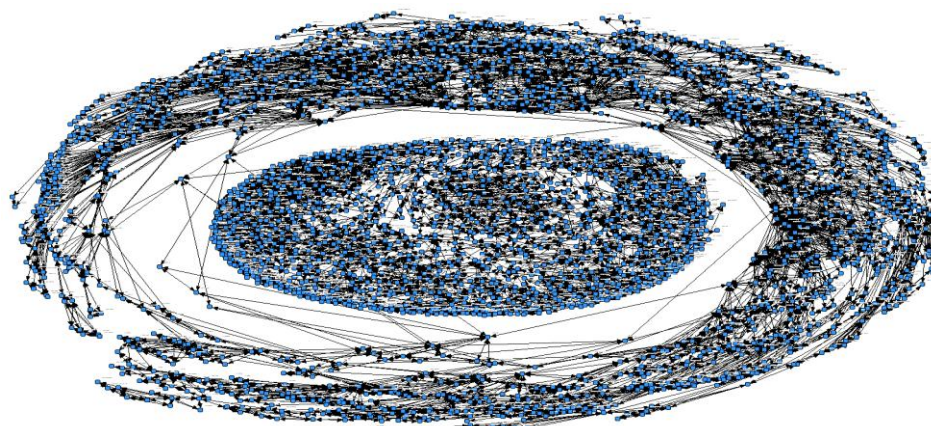
Figure-1. Co-authorship network by Chinese inland Scholars in the field of information science.
Source: Visualization by UCINET6.

3.3. Variable Description and Measurement

In order to examine the driving factors for Chinese inland scholars' co-authorship network evolution, we take the collaborations by Chinese inland scholars as the dependent variable, i.e. in the collaborations matrix, the value is 1 if two authors have co-authored publications, the value is 0 otherwise.

As for network position of the authors, we use the centrality measure, namely degree centrality and betweenness centrality. We measure the similarity if two authors are in the same affiliation. In co-authorship network matrix, the value is 1 if two authors are in the same affiliations, the value is 0 otherwise. As for prior collaborations, the value is 1 if two authors collaborated during pried t-1 , the value is 0 otherwise.

4. Network Regressions Analysis
4.1. Descriptive Statistics and Correlation Matrix of Variables

The descriptive statistics and correlation matrix are shown in Table 1, Table 2, Table 3 and Table 4. From Table 1, Table 2, Table 3 and Table 4, it can be seen that the network positions of authors in the co-authorship network is positively correlated with collaborations, but the correlation is increasingly weaker over time. The similarity between authors is positively correlated with collaborations, indicating that if the two authors belong to the same affiliation, they are more likely to collaborate in later period. Prior collaboration experience is positively correlated with the formation of co-authorships. However, the correlation is relatively weak in the fifth period, indicating that part of the authors will continue to maintain

prior co-authorships. As the expansion of the network over time, the correlation between prior collaboration experience and network formation become increasingly insignificant.

Table-1. Descriptive Statistics and Correlation Matrix for Period 2.

	Degree centrality	Betweenness centrality	Similarity	Prior collaboration experience	Collaboration
Degree centrality	1				
Betweenness centrality	0.236	1			
Similarity	0.074	0.099	1		
Prior collaboration experience	0.046	0.053	0.043	1	
Collaboration	0.275	0.215	0.146	0.162	1

Source: Output by Eviews 5.

Table-2. Descriptive Statistics and Correlation Matrix for Period 3.

	Degree centrality	Betweenness centrality	Similarity	Prior collaboration experience	Collaboration
Degree centrality	1				
Betweenness centrality	0.421	1			
Similarity	0.135	0.095	1		
Prior collaboration experience	0.051	0.048	0.146	1	
Collaboration	0.256	0.267	0.187	0.177	1

Source: Output by Eviews 5.

Table-3. Descriptive Statistics and Correlation Matrix for Period 4.

	Degree centrality	Betweenness centrality	Similarity	Prior collaboration experience	Collaboration
Degree centrality	1				
Betweenness centrality	0.275	1			
Similarity	0.065	0.056	1		
Prior collaboration experience	0.017	0.050	0.187	1	
Collaboration	0.206	0.150	0.170	0.137	1

Source: Output by Eviews 5.

Table-4. Descriptive Statistics and Correlation Matrix for Period 5.

	Degree centrality	Betweenness centrality	Similarity	Prior collaboration experience	Collaboration
Degree centrality	1				
Betweenness centrality	0.224	1			
Similarity	0.032	0.043	1		
Prior collaboration experience	0.002	0.036	0.170	1	
Collaboration	0.241	0.144	0.157	0.088	1

Source: Output by Eviews 5.

4.2. Results of Network Regressions

Due to the fact that variables in the model are not completely independent, the use of conventional statistical methods may lead to inaccurate results because the variables are assumed to be independent of each other. Therefore, The Quadratic Assignment Procedure (QAP) is used to investigate the relationship between relations by comparing the similarity of values in two matrices (i.e. the correlation coefficient between two matrices) with non-parametric test coefficients based on the replacement of matrix data. The regression analysis with Multiple Regression Quadratic Assignment Procedure (MRQAP) has the advantage of avoiding the problem of multi-collinearity, and truly reflecting the explanation and accuracy of independent variables on dependent variable.

The empirical results are shown in Table 5, Table 6, Table 7 and Table 8. In the second period, the coefficient of determination is 0.132, and the adjusted coefficient of determination is 0.132, indicating that the

independent variable can be explained by 13.2% with the dependent variable. The coefficient of determination and the adjusted coefficient of determination in the third period are both 0.138. As for the fourth and fifth period, the linear relationship between the independent variables and the dependent variable is weaker, which means that the explanatory powers of dependent variables are well reflected in the early stages of network evolution, but the influence of independent variables on dependent variable getting smaller and smaller as the expansion of the network in the later period of network evolution.

For Hypothesis 1, the results show that authors who are at the center in period t-1 in the co-authorship network can form more collaborations during period t. This is in line with the findings of Archibugi and Coco (2004) where core authors are more likely to establish new partnerships and have more partners.

For Hypothesis 2, the similarities between authors may contribute to more collaborations (Rivera et al., 2010) indicating that authors belong to the same affiliations are more likely to collaborate and the impact of similarity on collaborations continue to strengthen with the evolution of network.

Table-5. Network regression results for Period 2.

Independent	Un-stdized Coefficient	Stdized Coefficient	Sig.
Intercept	-0.0015	0.000	
Degree centrality	0.0653	0.2268	0.000
Betweenness centrality	0.1679	0.1438	0.000
Similarity	0.0335	0.1096	0.000
Prior collaboration experience	0.2426	0.1388	0.000
R-square	0.132		
Adj R-square	0.132		
Probability	0.000		

Source: Output with MRQAP method by UCINET 6.

Table-6. Network regression results for Period **3**.

Independent	Un-stdized Coefficient	Stdized Coefficient	Sig.
Intercept	-0.0022	0.000	
Degree centrality	0.0430	0.1536	0.000
Betweenness centrality	0.1129	0.1838	0.000
Similarity	0.0517	0.1282	0.000
Prior collaboration experience	0.1874	0.1420	0.000
R-square	0.138		
Adj R-square	0.138		
Probability	0.000		

Source: Output with MRQAP method by UCINET 6.

Table-7. Network regression results for Period 4.

Independent	Un-stdized Coefficient	Stdized Coefficient	Sig.
Intercept	-0.0028	0.000	
Degree centrality	0.0124	0.1707	0.001
Betweenness centrality	0.06793	0.0900	0.006
Similarity	0.5795	0.1345	0.000
Prior collaboration experience	0.1111	0.1040	0.000
R-square	0.086		
Adj R-square	0.086		
Probability	0.000		

Source: Output with MRQAP method by UCINET 6.

For hypothesis 3, there exists positive influence of prior collaboration experience on collaborations. However, the influence gradually becomes weaker with the expansion of the network size. If two authors have prior collaborations in the previous period, the collaboration may continue in the next period. In our sample, a small number of authors have maintained prior collaborative relations. however, due to the expansion of network, the influence gradually becomes weaker. Prior good contact between organizations can form a more strong connection. For an individual author, previous collaboration experience provides the possibility of later collaboration.

Table-8. Network regression results for Period 5.

Independent	Un-stdized Coefficient	Stdized Coefficient	Sig.
Intercept	–0.0032	0.000	
Degree centrality	0.0609	0.2170	0.000
Betweenness centrality	0.0941	0.0875	0.006
Similarity	0.0672	0.1656	0.000
Prior collaboration experience	0.0704	0.0612	0.006
R-square	0.092		
Adj R-square	0.092		
Probability	0.000		

Source: Output with MRQAP method by UCINET 6.

5. Conclusion

Based on the articles published by Chinese inland scholars in 35 SCI and SSCI journals in the field of information science, the paper examines the impact of the network positions, similarities and prior collaboration experiences of core authors on the evolution of international co-authorship networks. The empirical results show that, (1) central positions of authors in co-authorship networks may lead to the establishment co-authorships; (2) The similarities and prior collaboration experiences may also lead to the formation of co-authorships. Specifically, authors within the same affiliations are more likely to establish co-authorships, and prior collaboration experiences will lead to the continuation co-authorship.

The empirical results indicate that, (1) the co-authorship network may evolve when an individual author strengthen the cooperation with central authors in the network, (2) the co-authorship network may evolve when an individual author maintain the prior collaboration ties with other authors; and (3) the co-authorship network may evolve when an individual author seek cooperation with authors belong to the same affiliation.

There are still some limitations with the article. First, the empirical results are based on central authors in the co-authorship network, which neglects the peripheral authors in the network. Secondly, the hypotheses are tested in the information science, which may not be applied to other disciplines. Future research could test the influence of peripheral authors on the evolution of co-authorship network as well as other disciplines.

References

Archibugi, D., & Coco, A. (2004). International partnerships for knowledge in business and academia: A comparison between Europe and the USA. *Technovation, 24*(7), 517-528.

Barabási, A. (2003). *Emergence of scaling in complex networks. Handbook of graphs and networks: From the genome to the internet.* Wiley-VCH: Weinheim.

Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509-512.

Boschma, R., & Frenken, K. (2010). *The spatial evolution of innovation networks: A proximity perspective. In: Boschma, R., Martin, R. (Eds.), Handbook of Evolutionary Economic Geography.* Cheltenham: Edward Elgar Publishing.

Cantner, U., & Rake, B. (2014). International research networks in pharmaceuticals: Structure and dynamics. *Research Policy, 43*(2), 333-348.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35*(1), 128–152.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94*(94), 95-120.

Goldfinch, S., Dale, T., & Jr, D. R. (2003). Science from the periphery: Collaboration, networks and 'periphery effects' in the citation of New Zealand crown research institutes articles, 1995-2000. *Scientometrics, 57*(3), 321-337.

Gulati, R., & Gargiulo, M. (1999). Where do inter organizational networks come from? *American Journal of Sociology, 104*(5), 1439-1438.

Hoekman, J., Frenken, K., & Oort, F. V. (2009). The geography of collaborative knowledge production in europe. *Annals of Regional Science, 43*(3), 721-738.

Jiang, R. J., Tao, Q. T., & Santoro, M. D. (2010). Alliance portfolio diversity and firm performance. *Strategic Management Journal, 31*(10), 1136-1144.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1-18.

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation, 11*(1), 3–15.

Luo, X., & Deng, L. (2009). Do birds of a feather flock higher? The effects of partner similarity on innovation in strategic alliances in knowledge-intensive industries. *Journal of Management Studies, 46*(6), 1005–1030.

Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical Review E Statistical Nonlinear & Soft Matter Physics, 64*(2), 25102.

Powell, W., & Brantley, P. (1992). Competitive cooperation in biotechnology: learning through networks? In: Eccles, R., Nohria, N. (Eds.), Networks and Organizations. Boston (pp. 366–394): Harvard Business School Press.

Powell, W., & Grodal, S. (2005). Networks of innovators. In: Fagerberg, J., Mowery, D.,Nelson, R. (Eds.), The Oxford Handbook of Innovation (pp. 56–85). Oxford: Oxford University Press.

Powell, W., Koput, K., & Smith-Doerr, L. (1996). Inter organizational collaboration and the locus of innovation: Networks of learning in biotechnology? *Administrative Science Quarterly, 41*(1), 116-145.

Rivera, M. T., Soderstrom, S. B., & Uzzi, B. (2010). Dynamics of dyads in social networks: Assortative, relational, and proximity mechanisms. *Annual Review of Sociology, 36*(36), 91-115.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039.

Zaheer, A., & Soda, G. (2009). Network evolution: The origins of structural holes? *Administrative Science Quarterly, 54*(1), 1–31.