

Comparative study on structure and correlation among author co-occurrence networks in bibliometrics

Jun-Ping Qiu · Ke Dong · Hou-Qiang Yu

Received: 17 November 2013 / Published online: 7 May 2014
© Akadémiai Kiadó, Budapest, Hungary 2014

Abstract This paper introduces author-level bibliometric co-occurrence network by discussing its history and contribution to the analysis of scholarly communication and intellectual structure. The difference among various author co-occurrence networks, which type of network shall be adapted in different situations, as well as the relationship among these networks, however, remain not explored. Five types of author co-occurrence networks were proposed: (1) co-authorship (CA); (2) author co-citation (ACC); (3) author bibliographic coupling (ABC); (4) words-based author coupling (WAC); (5) journals-based author coupling (JAC). Networks of 98 high impact authors from 30 journals indexed by 2011 version of Journal Citation Report-SSCI under the Information Science & Library Science category are constructed for study. Social network analysis and hierarchical cluster analysis are applied to identify sub-networks with results visualized by VOSviewer software. QAP test is used to find potential correlation among networks. Cluster analysis results show that all the five types of networks have the power for revealing intellectual structure of sciences but the revealed structures are different from each other. ABC identified more sub-structures than other types of network, followed by CA and ACC. WAC result is easily affected and JAC result is ambiguous. QAP test result shows that ABC network has the highest proximity with other types of networks while CA network has relatively lower proximity with other networks. This paper will provide a better comprehension of author interaction and contribute to cognitive application of author co-occurrence network analysis.

Keywords Co-authorship analysis · Author co-citation analysis · Author bibliographic coupling · Correlation analysis · Bibliometrics

J.-P. Qiu · K. Dong (✉) · H.-Q. Yu (✉)
School of Information Management, Wuhan University, Wuhan, China
e-mail: dongke40@gmail.com

H.-Q. Yu
e-mail: yuhouq@yeah.net

Introduction and background

Co-occurrence in bibliometrics has been studied for a long history. Different kinds of co-occurrence analysis were used to investigate a wide spectrum of questions such as scholarly communication, research front, and intellectual structure of research fields. Paper is usually the basic unit in co-occurrence analysis, and higher levels of aggregation include authors (White and Griffith 1981; Zhao and Strotmann 2008b), journals (Ni et al. 2013a), institutions (Thijs and Glänzel 2010; van Rijnsouwer et al. 2008), fields (Fox 2008), and countries (He 2009; Wagner 2005). Author-level is importantly dominated at aggregation levels, because for journals, institutions and countries, every type of network is formed through authors' academic relations such as collaboration or citation among authors.

Author co-occurrence in bibliometrics

Author co-occurrence analysis is a prevailing approach in analyzing scholarly communication and structure of science (Chen and Lien 2011; Ding and Cronin 2011; White and Griffith 1982; White and McCain 1998). Author collaboration network is a typical author co-occurrence network. Back to year 1966 when Price and Beaver (de Solla Price and Beaver 1966) did analysis on the author collaboration among 'invisible collages', they found that prolific authors lead the way in their research group and bridged the way among different groups. Later in the twentyfirst century, with the development of network analysis and especially the accumulated researches in complex network on scientific collaboration (Newman 2001a, b, c, 2004), collaboration has been increasingly pervaded in mining intellectual structure (Acedo et al. 2006; Eslami et al. 2013; Kretschmer 2004; Otte and Rousseau 2002; Thijs and Glänzel 2010). In addition, some studies showed that research collaboration can bring co-authors greater research productivity (Lee and Bozeman 2005) and research impact (Gazni and Didegah 2011). With the development of social network analysis, some techniques from social network analysis are also used to explain the dynamics of co-authorship networks (Acedo et al. 2006; Yan and Ding 2009). These investigations have provided insights to study the collaboration patterns and scholarly communication.

Author co-citation analysis (White and Griffith 1981) assumes that two authors are related if they are cited together by later works, and the higher the frequency, the more similar these two authors are. Since its introduction, author co-citation (ACC) analysis has become a popular and widely used technique to analyze the intellectual structure of an academic discipline, such as library and information science (White and McCain 1998; Zhao and Strotmann 2008b), bioinformatics (Song and Kim 2013), and e-learning research (Chen and Lien 2011). Meanwhile, studies on ACC analysis methodology itself have kept obtaining much attention, for example, studies on common procedures of ACC analysis (McCain 1990), similarity measure (Ahlgren et al. 2003; Egghe and Leydesdorff 2009; Van Eck and Waltman 2008), visualization (Chen et al. 2001; White 2003; Zhao and Strotmann 2008c), ACC analysis in web environment (Leydesdorff and Vaughan 2006), first and all-ACC analysis (Schneider et al. 2009; Zhao and Strotmann 2008a).

Author bibliographic coupling is extended from bibliographic coupling (Kessler 1963) which assumes that the more references two authors have in common in their oeuvres, the more similar their research is. Zhao and Strotmann (2008b) proposed the idea of author bibliographic coupling (ABC) and implemented it in the field of library and information science. Rousseau (2010) did research in calculating the author coupling strength theoretically and proposed a method of calculating the simple coupling strength and the

relatively simple coupling strength. Ma (2012) found that ABC can not only discover the intellectual structure of a discipline but also reflect the research frontier of the discipline. On top of the proposed couplings, authors can also be coupled through other forms of academic communication. For example, Cabanac (2011) considered author-venue-coupling to measure inter-researcher similarity through the conferences they might have jointly attended (Ni et al. 2013b; Tang et al. 2008). In fact, the current extension of author co-occurrence network is mainly on different types of coupling relationship among authors.

The above studies have investigated author co-occurrence from several aspects, which greatly developed the author co-occurrence analysis methodology from both theory and practice. As studies on these author co-occurrence analysis have habitually used only one type of network, the corresponding findings may not be complete. A few scholars have become aware of the problem, and bring out the current trend of combining different co-occurrence networks (Boyack and Klavans 2010; Groh and Fuchs 2011; Strotmann and Bleier 2013; Zitt et al. 2011). Adaption motivation and analysis result of different author co-occurrence networks may be different, questions are: how is the difference reflected during the structure analysis, which network shall be adapted to solve the specific problem? (e.g. scholarly communication, intellectual structure)? To find out the answer is the first research goal. Meanwhile, different co-occurrence relationships may have mutual influence (Ding 2011; Lin and Huang 2012; Sooryamoorthy 2009) and scientific communication reflected by different co-occurrence networks may share similarity. So, what is the similarity and how does the similarity bring mutual influence among these networks? To reveal the inner principle becomes the second research goal.

Correlation among co-occurrence networks

Correlation analysis in bibliometrics has gained increasing attention in recent years. White et al. (2004) analyzed the correlation between social networks and intellectual networks of authors, and found that intellectual ties based on shared content perform better as predictor than content-neutral social ties like friendship. Johnson and Oppenheim (2007) identified and discussed the similarities between citation patterns and social closeness. Ding (2011) analyzed co-authorship and citation networks to show whether prolific authors tend to collaborate with or cite other researchers, and whether highly cited authors tend to collaborate with or cite each other. Wallace et al. (2012) examined the influence of collaboration networks on citation practices. Ni et al. (2013b) examined the proximity of journals based on producer, artifact, concept, and gatekeeper. However, these researches compared only a few networks, and few researches, with scattered discoveries, tap into the relationships among author co-occurrence networks.

Yan and Ding (2012) used cosine distance to explore the similarity among six types of scholarly networks aggregated at institution level, including bibliographic coupling networks, citation networks, co-citation networks, topical networks, co-authorship networks, and co-word networks. Despite the similar research design at first glance, it is completely different research compared with our study because authors and institutions are at different levels of grouping. For institutions, each type of co-occurrence network is formed through academic relations such as collaboration or citation among authors. Author is the subject of co-occurrence relationship at higher aggregation levels (e.g. institutions, countries), thus, analysis on author co-occurrence relationship can reflect more the truth of scientific research and scholarly communication. What's more, in Yan's study, citation network was compared with other networks. Given that citation network is directed network and the

other co-occurrence networks are undirected networks, whether or not the same method can be applied to compare citation network with other networks, remains to be verified by more researches. In this paper, the quadratic assignment procedure (QAP) is used to measure correlations between networks, which contributes to the full landscape of relationships among different author-level co-occurrence networks.

To achieve these research goals, five types of author-level co-occurrence networks, namely co-authorship (CA), ACC, ABC, words-based author coupling network (WAC) which provokes the same assumptions of ABC and journal-based author coupling network (JAC) which bases on authors' publications in the same journals are constructed for structure and correlation analysis. The analysis consisted of two main steps, First of all, structure analysis results of different types of co-occurrence networks were obtained, the results were used to compare the structure of different types of author co-occurrence networks and find out their difference; After that, correlations among networks were calculated to reveal the mutual influence. For the first time, structure comparison and correlation analysis were done on five author co-occurrence networks, We suppose this study will provide a better comprehension of author interaction and contribute to cognitive application of author co-occurrence network analysis.

Methodology

Data collection and cleaning

It's been a convention to select high impact author group (prolific authors, highly cited authors etc.) as object for co-occurrence analysis. As technologies develop, contemporary related tools and technologies begin to support large volume data analysis. The primary goal of this research is to reveal the correlation among different networks. Thus, if too many authors were included in the study, co-occurrence networks will be too sparse to do correlation analysis with value of many cells being 0. If too many authors neither collaborate nor are co-cited, it's difficult to tell what on earth the mutual influence is between these two relationships. As a result, data of high impact journals and high impact author groups were collected to construct the networks.

Firstly, 30 relatively high impact journals (e.g. $IF \geq 1$) in the Information Science & Library Science (IS&LS) category were identified according to the 2011 version of Journal Citation Report. Then, journal titles were used to retrieve in Web of Science. The situation in which journal title altered has been taken into account during the data collection process. For example, JASIS altered its name to JASIST. Three kinds of documents: articles, proceedings and review, were all downloaded. The time span is from the past to the data collection date Dec 12 2012.

Author analysis must consider disambiguation of author names (Torvik et al. 2005). Data cleaning is a progressive process when different situations cropped out in dealing with author names and keywords. Author name data cleaning includes two steps: firstly, Thomson Data Analyzer software is applied to clean the data. Secondly, the preliminary cleaned data were further manually cleaned. Table 1 shows the basic information of the cleaned data. Citing authors' names have no obvious labeling problem and the error rate is less than 5 % compared with the raw data. But serious problem lied in cited authors' name labeling and the error rate is upto 20 %. It is because cited authors are labeled by different citing authors and that different journals do not share the same name labeling rules. For example, James J. Cimino is found to be Cimino J J, Cimino James J, Cimino JJ or other

Table 1 Summary of dataset

	Value	Cleaned
Number of papers	25,652	–
Number of cited references	499,986	–
Number of first authors	14,857	14,151
Number of authors	32,268	30,908
Number of cited authors	214,887	172,295

forms in papers’ references. Four major forms of labeling rules influence author names, they are: (1) whether space character is being used, (2) whether names use abbreviations, (3) whether a dot is marked at the right side of abbreviations, and (4) first name and second name are reversed out of order like White HD and White DH. Name labeling is even more complex in Chinese and Korean (Kim and Cho 2013). Since Chinese or Korean authors with high impact are relatively the minority in the research, a detailed discussion about it is omitted.

After author data cleaning, first author, all authors and cited authors are ranked according to the number of papers published or times cited. According to Price’s Law, 216 authors who have published over $0.749 (n_{max})^{1/2}$ papers are identified as prolific authors. These prolific authors are further filtered by their total citations. At last, 98 authors who published 10 or more papers and simultaneously are cited higher than 160 times are identified in IS&LS domain. Keywords data cleaning shares basically the same process with that of author name data. Finally, 1,484 words with frequency over (or equal to) 5 times are used to construct words-based author coupling network.

Network construction

In social network analysis, one mode network is constructed by measuring only one actor set, two mode network measures two actor sets or the relationship between one actor set and one event set (de Nooy et al. 2005). All 5 types of network studied here are one mode network. The construction process can be described uniformly using the transformation from two mode network to one mode network. Suppose A is a two mode network constructed by actor group multiplied by event group, the rows are actor group and the columns are event group. Define A^T as transposed matrix and P as one mode network among actors. Then we got:

$$P = A * A^T$$

In fact, both CA and ACC can be generated directly by Thomson Data Analyzer. For ABC, WAC and JAC, basic matrixes which include authors*cited reference matrix, authors*keywords matrix and authors*journals matrix were generated first, and then transformed into ABC, WAC and JAC networks via formulas.

Another key problem in co-occurrence network construction is to calculate the author coupling strength. Ma (2012) summarized the author coupling strength calculation process into three methods: simple method, minimum method, and combined method. He also suggested the minimum method is the most appropriate. Here we endorse his view and use the minimum method to calculate the coupling strength.

Methods

Social network analysis (SNA) and traditional bibliometric analysis are based on different perspectives and methods for structural analysis purpose. For social network analysis, cohesion analysis is used with methods such as component, k-core, p-cliques etc., and it treats the network as a whole to explore the structure by analyzing sub-networks. In bibliometric analysis, commonly used method is hierarchical clustering which categorizes samples into different types and aggregates them by calculating similarity distance. In this paper, component analysis from SNA is applied because of its simple and intuitive effect; if component analysis cannot produce the sub-networks, then hierarchical clustering is used to do structural analysis where the clustering option is Squared Euclidean Distance and Ward's Method. The final results are visualized with VOSviewer software (Van Eck and Waltman 2010). A comparison on proximity results was conducted using the QAP which is always used in measuring correlations between two networks (Hubert and Schultz 1976). QAP statistics are annotated in the documentation of software Ucinet (Borgatti et al. 2002). More detailed description can be found in another previous paper (White et al. 2004).

Results

Summary of networks

Density refers to the percentage of number of existing lines to number of all potential lines. Table 2 shows the number of lines and general density of five types of network. Because all five types of networks consist of 98 author nodes, the influence on network density from network scale can be eliminated. All five types of networks except from author collaboration network show very big density and large number of lines. And VOSviewer can well abate the influence on visualization result brought by overwhelming number of lines.

Structure analysis and visualization

Figure 1 is the visualization result of CA network. The size of vertices marked by using loops represents the number of papers published. The largest cluster of this network is found to be composed by 38 authors (cluster 1), which has three distinct groups. First group is for the authors who are highly productive such as Egghe L, Leydesdroff L, Braun T, Rousseau T, Thelwall M etc. Most of them focus on bibliometrics, informetrics, scientometrics, and webometrics. The second group is centered around Ingwersen P, Croft W B with other members including Borgman CL, Belkin NJ, Rebertson SE and so on. Their research topics are related with information seeking and retrieval. The third group is centered around Spink A with members Wilson TD, Ford N, Ellis D, Cole C etc., and the main research focuses are related with information behavior. The ties connected three groups indicate the research contents are inter-crossing. The first and second group is linked by Ingwersen P who is the winner of Derek de Solla Price Medal for his research in scientometrics and webometrics in 2005, and he is also a professor in information retrieval. The second and third group is linked by Saracevic T who has an extensive research interests mainly resting on digital library, information seeking and retrieval. Moreover, cluster 2 is also quite large in scale with medical informatics as its research topic. This cluster is centered around Friedman C and Cimino JJ. Cluster 3 is centered around four

Table 2 Scale and density of networks

Networks	No. of vertices	No. of edges	Density
CA	98	81	0.0170
ACC	98	3092	0.6505
ABC	98	3457	0.7273
WAC	98	2518	0.5298
JAC	98	3344	0.6964

authors who are professors of business school with management information system as their research domain. Cluster 4 includes four authors centered around Grover V whose research topics obviously emphasized on information technology. There are three authors in cluster 5 who do the research work about government information.

The ACC clustering result yields four clusters as displayed in Fig. 2. The size of vertices is proportional to author's cited times. For example, large vertices include authors like Garfield E, Leydesdorff L, Thelwall M, Egghe L, Salton G, Braun T and Kostoff RN etc. The boundary between different clusters is clearly displayed. The biggest group is labeled as cluster 1 in the center of the graph but actually comprises isolated vertices. It contains many vertices that do not belong to other clusters like Cimino JJ, who is an expert in bioinformatics field, and Venkatesh V, who is an expert in management information system. Cluster 2 on the top of the figure is the second largest in scale and is composed of experts who are connected more closely with each other in bibliometrics, informetrics, scientometrics and citation analysis. Cluster 3, where Salton G lies in the center, is clustered with members in information retrieval field while cluster 4 was composed of authors in text mining and knowledge discovery fields where Kostoff RN is positioned in the center. Except for cluster 1, the research boundaries of other three clusters can be easily identified. Although degree is used to represent the size of each vertex, citation distribution among different authors is different as to cited papers. For example, Garfield E, Small H, and Salton G are quite similar in that most citations come from their few classic works (Garfield 1972; Garfield and Merton 1979; Salton 1989; Small 1973). But cases are different for vertices such as Leydesdorff L, Thelwall M, and Egghe L whose citation distribution is clustered more loosely than the previous three authors.

Figure 3 is the mapping result of ABC network. Unlike ACC network, the research group on bibliometrics, informetrics and scientometrics is partitioned into two parts; one is composed of Leydesdorff L, Glanzel W etc. (cluster 1), the other one contains six authors including Egghe L, Rousseau R and Bernmann L etc. (cluster 2). The six authors in cluster 2 mostly cited papers on h-index and bibliometric laws; while Leydesdorff L and Glanzel W and other authors in cluster 1 focus more on citation analysis, visualization and the application of bibliometric methods. Small H, White HD and other authors in cluster 3 are from universities or research institutes in US such as Drexel University, ISI, and Indiana University at Bloomington. They are clustered together through large number of papers with theme on co-occurrence. The three groups above are all doing related researches on bibliometrics, informetrics and scientometrics, but research topics are diversified at micro level. The cluster 4 formed by Thelwall M and other four authors is quite small in scale, but due to its extensive influence, it has a relatively good visualization result of cohesion analysis. And this group is clustered by papers in webometrics, link analysis, and application of informetrics under web environment. Authors in cluster 8 are researchers mainly involved in management science. The high density of ABC network is due to the large size of overlapped references in the authors' published papers. These authors have written

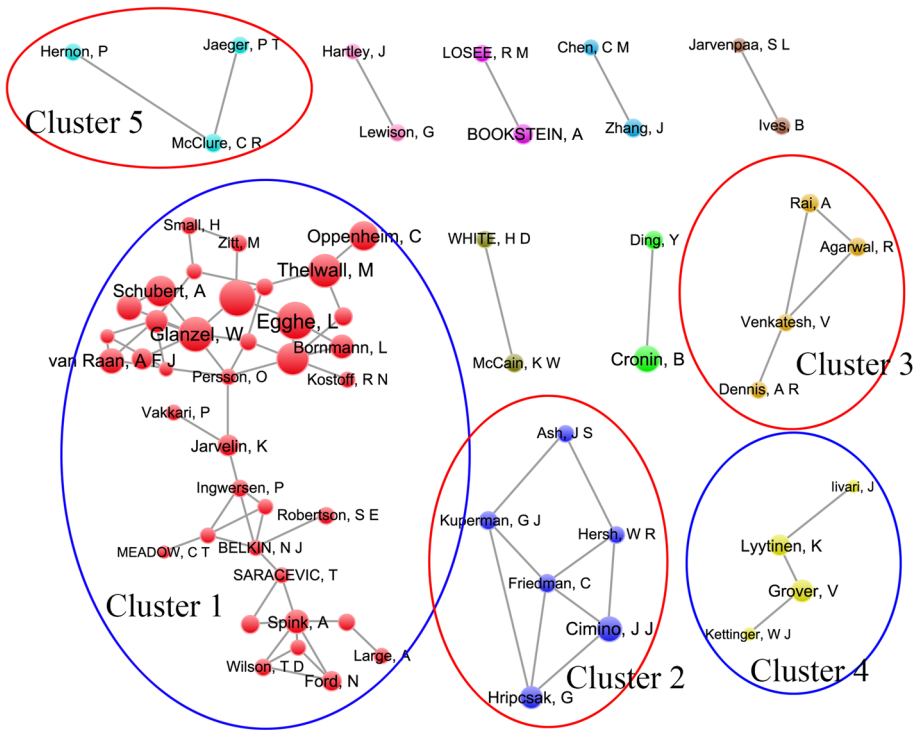


Fig. 1 Mapping result of CA network

articles with substantial number of references, so the vertices turn out to be large in size in the visualization result. It is showed in the right part of the figure (cluster 5) that a research community on information retrieval is centered around Salton G. The Cimino JJ-centered community on medical informatics (cluster 6) and the community with member Spink A etc. who focus on users and information behavior (cluster 7) are partitioned clearly.

Figure 4 is the visualization result of WAC network. The analysis process is very alike to the process in the previous ABC network in which research proximity is reflected by intersection of cited references. In WAC network, research proximity is manifested by intersection of academic terms which are used to indicate research content. The six research groups are partitioned and clearly presented. Compared to the previous network analysis results, the number of authors in information system (cluster 6) in this network is even larger, containing researchers who focus on information system from information science and those who focus on management information system from management science. Cluster 1 and cluster 2 from ABC are merged into cluster 1 in WAC network with research interests in bibliometrics, informetrics and scientometrics. For cluster 2, Thelwall M himself is mainly involved in webometrics, but the boundary of this group is not as clear as the one showed in ABC analysis result because papers published by Thelwall M as a co-author involve research content such as citation analysis and impact factor etc. Thus, the scale of this group becomes larger. The analysis results for clusters such as medical informatics (cluster 4), information seeking and retrieval (cluster 3), information behavior (cluster 5) are quite similar to the analysis result in ABC.

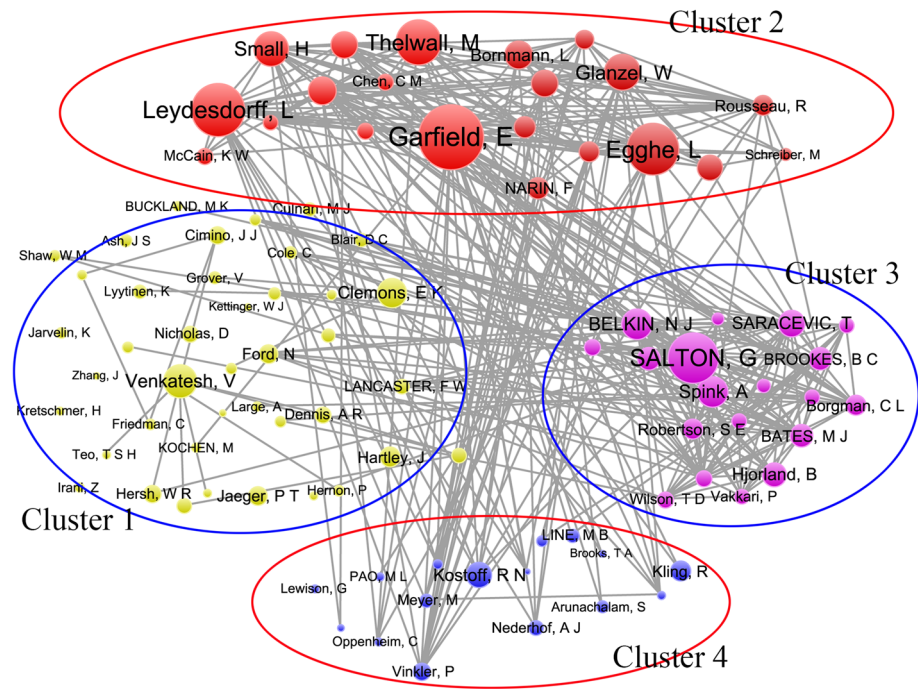


Fig. 2 Mapping result of ACC network

Figure 5 is the clustering result for JAC. Authors from medical informatics and management information system fields are distinctively partitioned at the upper part of the figure. Author group centered around Cimino JJ have papers published on *Journal of the American Medical Informatics Association*. Three authors in cluster 7 do researches in government information and most of the papers are published on *Government Information Quarterly*. The group at the bottom of the figure (cluster 2) is mainly composed of authors who publish on *Journal of Information*. The boundaries of other groups can be roughly distinguished but the authors fail to be divided into specific research communities.

Quadratic assignment procedure test

Table 3 is the result of QAP correlation test calculated among networks. It is noted that CA network is most significantly correlated with ABC network, while ACC network is its second most significantly correlated network followed by JAC network and WAC network.

QAP analysis results show that CA network is least related with other networks. Generally, if collaboration among authors has happened, authors must have been acquainted with each other and have communicated and researched on shared research topics, so the CA network among authors are closest to social network where desire for collaboration is the strongest. We found that CA network is the most loosely constructed, which indicates that if two authors are in the same research domain, it doesn't necessarily mean they collaborate with each other. So CA network is powerful in revealing the structure of academic communication but slightly weak in discovering disciplinary structure. Except from CA network, other four types of network are constructed based on

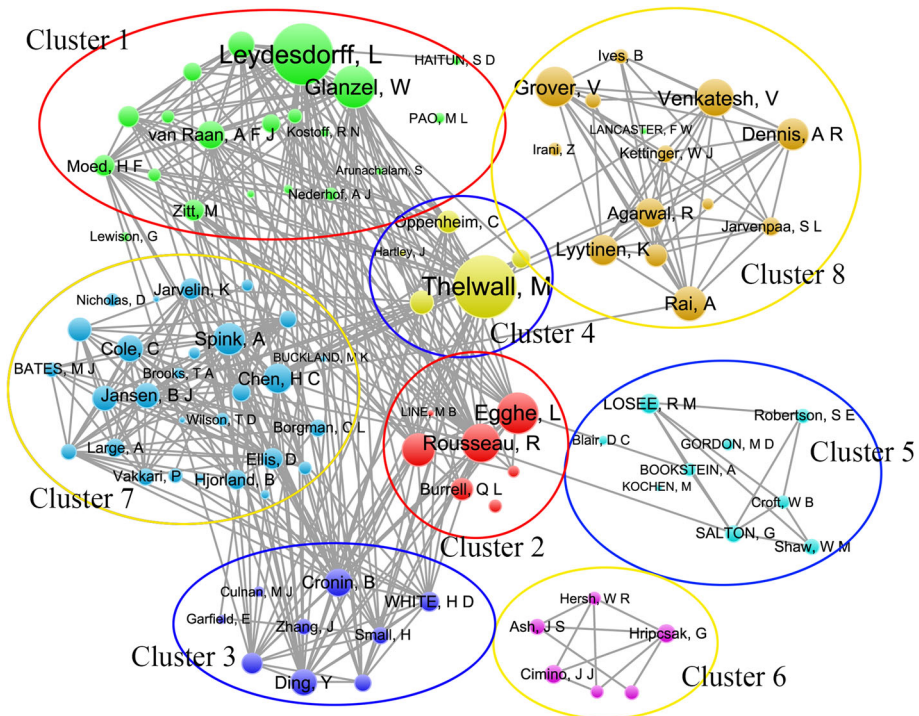


Fig. 3 Mapping result of ABC network

indirected connections and it may be one reason why CA network is least related with other networks.

ACC network is far different from CA network. This can simply be explained in that ACC network indicates authors' degree of recognition by other scholars while the real situation may be that the authors have never had direct academic communication at all. So ACC network may not be an appropriate way to analyze academic communication, rather it can be used to reveal disciplinary structure. With respect to the network proximity between ACC network and other networks, ACC is most related with ABC in that ACC reveals structure of relation in research paper references while author coupling tends to show the structure of relation among authors at research frontier. For author groups of the same community, ACC and ABC are linked and have higher proximity.

ABC network is the most proximate with the other four types of networks. Its proximity value with WAC is as high as 0.7. This result is reasonable since paper references which represent the existing knowledge base are chosen closely related to author's own research content during scientific research; while for WAC, authors' description of their own researches through usage of words or academic terms, can be regarded as summarization when investigating further from previous researches. ABC network is also highly correlated with ACC network. Being cited is usually deemed to reflect the authors' important role in the whole development process of research topics. So discipline structure revealed by ACC network is retrospective. ABC network tends to reveal the frontier structure. For the same author group, their research content usually keep significant consistency from historical view, so these two types of networks are highly correlated.

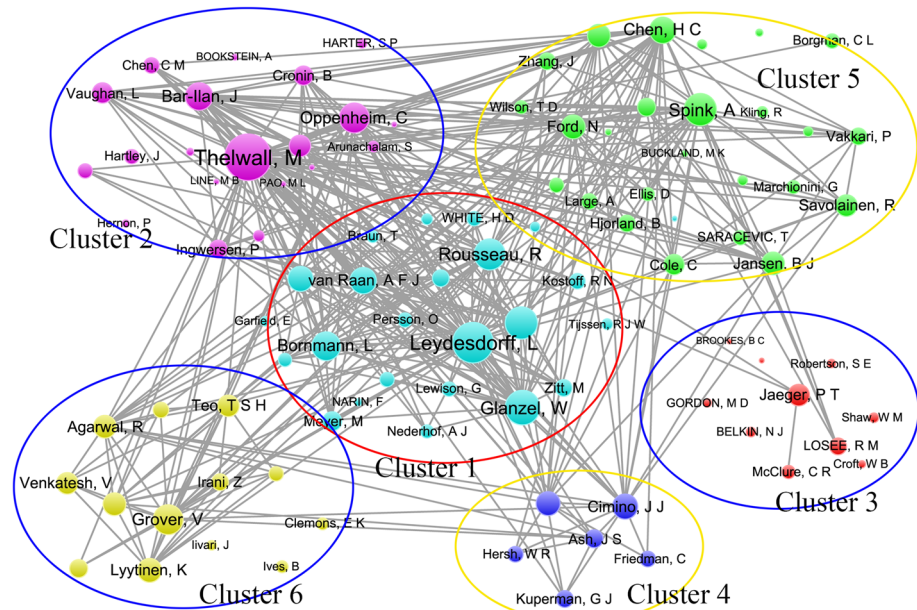


Fig. 4 Mapping result of WAC network

QAP test in Table 3 shows that WAC is weakly correlated with CA and ACC. The limitation of WAC is that its analysis result and ability of revealing disciplinary structure can be easily affected by words which can cause ambiguity when they have different meanings. For example, authors from information science who study information system are clustered together with those from management science by using words related to information system. Literally they are both doing researches on information system, but the fact is that their researches may have totally different perspective or pattern. In case of low ambiguity, WAC network becomes more efficient in analyzing the scientific structure.

According to the researches by Garfield (Garfield 1996; Garfield and Merton 1979), a relatively small number of journals publish the majority of significant scholarly results. Papers in these journals reflect the disciplinary affiliation of the journals. With respect to the topics of journals, they are generally not confined to single topic. Meanwhile, only a few groups which are obviously different can be partitioned in the visualization result of JAC network. So, JAC network is not closely correlated with other networks and we think it is a good way to analyze neither the scholarly communication nor the intellectual structure.

Discussion

As an explorative study, the analysis method of the study remain to be discussed from several aspects. Correlation analysis among networks is an important way to intuitively compare different networks. QAP method applies very well in analyzing different types of author co-occurrence relationships. In substructure analysis, cohesive group analysis in social network analysis is widely used which includes components, k-core and cliques. Centrality degree is used to aid to do structure explanation. The paper also tried to use

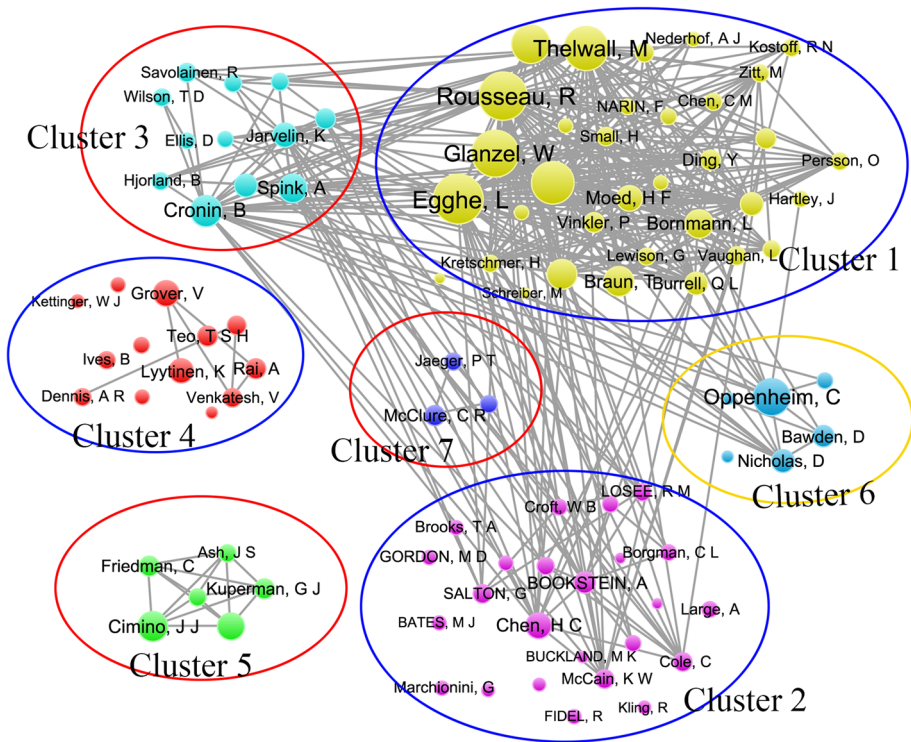


Fig. 5 Mapping result of JAC network

Table 3 QAP correlation test of networks

Network	CA	ACC	ABC	WAC	JAC
CA	1	0.353	0.473	0.220	0.306
ACC	0.353	1	0.641	0.364	0.513
ABC	0.473	0.641	1	0.700	0.586
WAC	0.220	0.364	0.700	1	0.584
JAC	0.306	0.513	0.586	0.584	1

Number of permutations is 5,000 times, with significant level at 0.001

Numbers in bold have shown the maximum and minimum correlation value between WAC and other types of networks

cohesive group analysis, but comparison work shows that cluster analysis results are relatively easier to explain. There are two problems in applying cohesive group analysis to author co-occurrence analysis:

- (1) Usability of indicators like k-core, cliques etc. is hard to guarantee. Traditional networks are mostly 0–1 network. Indicators like k-core are mainly used to explore structure features of nodes. They show significant advantage in analyzing such kind of network, because structure features of nodes in this type of network usually reflect features of nodes themselves. Co-occurrence network among authors are mostly

weighted network. The typical selection of high impact or highly productive authors for empirical studies makes the value of relationship in the network even higher. Whether or not structure features can truly reflect features of nodes remains a question. It's also not clear how distinguishing different weight values can be. So usability of indicators like k-core, cliques etc. is hard to guarantee.

- (2) Indicators that aid to explain structure like closeness centrality and betweenness centrality may not adapt to author co-occurrence analysis. Closeness centrality mainly reflects the capability of nodes in connecting with other nodes. The calculation basis is distance between one author and the other authors. Betweenness centrality mainly reflects the extent to which one node locates in connection paths among other nodes. In author co-occurrence relationship, distance and betweenness degree obtained from network analysis cannot truly reflect connections among authors, because the basis for connections among authors changes in academic circumstances. For example, when one author chooses to collaborate with the others, a mediator is not a necessity. In citing one's publication, the information access tunnel comes into priority in generating impact, but the information access tunnel is not based on distance among authors.

Conclusion

The paper constructed five types of author co-occurrence networks. Through hierarchical clustering and correlation analysis, capabilities of different types of author co-occurrence relationships in revealing scientific structure are compared. Hierarchical clustering result shows that author collaboration relationship has significant advantage in studying scientific communication while may leak some information in revealing discipline structure. Author co-citation network has moderate capability in revealing discipline structure. Author bibliographic coupling network shows significant advantage in revealing discipline structure with the highest accuracy and precision in structure analysis. Words-based author coupling is likely to be influenced by the characteristics of domain words in revealing scientific structure. Despite the poor capability of journal-based author coupling in revealing discipline structure, it can occasionally distinguish group difference that may easily be neglected.

From perspective of correlation among networks, result analysis of author collaboration network show the lowest correlation with that of the other networks. It indicates low substitution between result of author collaboration network and that of other networks. Author bibliographic coupling network presents the highest correlation with other networks, which demonstrates that analysis result of ABC may easily cover that of other networks. In the process of analyzing scientific structure, it's essential to analyze the features of selected samples themselves.

In summary, this study is useful for more appropriate combination of different author co-occurrence networks to understand scholarly communication and intellectual structure analysis. However, modeling author production from the Web of Science only seems to be an approximation of the diversity in scientific output. For instance, open access journals and conferences can also be taken into consideration to better reflect authors' true production. Some other methods, such as semantics can be employed to enhance WAC. Another interesting topic is to explore the influence of authors' disciplines and information

seeking behavior on selecting the type of author co-occurrence analysis during their studies. These further topics will be explored in the future researches.

Acknowledgments This paper is supported by the Major Program of the National Social Science Foundation of China (Grant No. 11&ZD152) and Humanities and Social Sciences project of Wuhan University (Grant No. 2012GSP032).

References

- Acedo, F. J., Barroso, C., Casanueva, C., & Galán, J. L. (2006). Co-authorship in management and organizational studies: an empirical and network analysis. *Journal of Management Studies*, *43*(5), 957–983.
- Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, *54*(6), 550–560.
- Borgatti, S., Everett, M., & Freeman, L. (2002). *UCINET 6 for Windows: Software for social network analysis (Version 6.102)*. Harvard: Analytic Technologies.
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389–2404.
- Cabanac, G. (2011). Accuracy of inter-researcher similarity measures based on topical and social clues. *Scientometrics*, *87*(3), 597–620.
- Chen, L.-C., & Lien, Y.-H. (2011). Using author co-citation analysis to examine the intellectual structure of e-learning: A MIS perspective. *Scientometrics*, *89*(3), 867–886.
- Chen, C., Paul, R. J., & O'Keefe, B. (2001). Fitting the jigsaw of citation: Information visualization in domain analysis. *Journal of the American Society for Information Science and Technology*, *52*(4), 315–330.
- de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek* (Vol. 27). Cambridge: Cambridge University Press.
- de Solla Price, D. J., & Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist*, *21*(11), 1011.
- Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of informetrics*, *5*(1), 187–203.
- Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing and Management*, *47*(1), 80–96.
- Egghe, L., & Leydesdorff, L. (2009). The relation between Pearson's correlation coefficient r and Salton's cosine measure. *Journal of the American Society for Information Science and Technology*, *60*(5), 1027–1036.
- Eslami, H., Ebadi, A., & Schiffauerova, A. (2013). Effect of collaboration network structure on knowledge creation and technological performance: The case of biotechnology in Canada. *Scientometrics*, *97*(1), 99–119.
- Fox, M. F. (2008). Collaboration between science and social science: Issues, challenges, and opportunities. *Research in Social Problems and Public Policy*, *16*, 17–30.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. American Association for the Advancement of Science.
- Garfield, E. (1996). Significant scientific literature appears in a small core of journals: Scientist, Incorporated.
- Garfield, E., & Merton, R. K. (1979). *Citation indexing: Its theory and application in science, technology, and humanities* (Vol. 8). New York: Wiley.
- Gazni, A., & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: A case study of Harvard University's publications. *Scientometrics*, *87*(2), 251–265.
- Groh, G., & Fuchs, C. (2011). Multi-modal social networks for modeling scientific fields. *Scientometrics*, *89*(2), 569–590.
- He, T. (2009). International scientific collaboration of China with the G7 countries. *Scientometrics*, *80*(3), 571–582.
- Hubert, L., & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, *29*(2), 190–241.
- Johnson, B., & Oppenheim, C. (2007). How socially connected are citers to those that they cite? *Journal of Documentation*, *63*(5), 609–637.

- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25.
- Kim, S., & Cho, S. (2013). Characteristics of Korean personal names. *Journal of the American Society for Information Science and Technology*, 64(1), 86–95.
- Kretschmer, H. (2004). Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web. *Scientometrics*, 60(3), 409–420.
- Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5), 673–702.
- Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology*, 57(12), 1616–1628.
- Lin, W.-Y. C., & Huang, M.-H. (2012). The relationship between co-authorship, currency of references and author self-citations. *Scientometrics*, 90(2), 343–360.
- Ma, R. (2012). Author bibliographic coupling analysis: A test based on a Chinese academic database. *Journal of Informetrics*, 6(4), 532–542.
- McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, 41(6), 433–443.
- Newman, M. E. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, 64 (1), 016131.
- Newman, M. E. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64 (1), 016132.
- Newman, M. E. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404–409.
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1), 5200–5205.
- Ni, C., Sugimoto, C. R., & Cronin, B. (2013a). Visualizing and comparing four facets of scholarly communication: producers, artifacts, concepts, and gatekeepers. *Scientometrics*, 94 (3), 1161–1173.
- Ni, C., Sugimoto, C. R., & Jiang, J. (2013b). Venue-author-coupling: A measure for identifying disciplines through author communities. *Journal of the American Society for Information Science and Technology*, 64 (2), 265–279.
- Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of information Science*, 28(6), 441–453.
- Rousseau, R. (2010). Bibliographic coupling and co-citation as dual notions. *A Festschrift in Honour of Peter Ingwersen, Special Volume of the e-Zine of the ISSI, 2010*, 173–183.
- Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Reading: Addison-Wesley.
- Schneider, J. W., Larsen, B., & Ingwersen, P. (2009). A comparative study of first and all-author co-citation counting, and two different matrix generation approaches applied for author co-citation analyses. *Scientometrics*, 80(1), 103–130.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.
- Song, M., & Kim, S. (2013). Detecting the knowledge structure of bioinformatics by mining full-text collections. *Scientometrics*, 96(1), 183–201.
- Sooryamoorthy, R. (2009). Do types of collaboration change citation? Collaboration and citation patterns of South African science publications. *Scientometrics*, 81(1), 177–193.
- Strotmann, A., & Bleier, A. (2013). Author name co-mention analysis: Testing a poor man's author co-citation analysis method. arXiv preprint arXiv:1309.5256.
- Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: extraction and mining of academic social networks. Paper presented at the Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining.
- Thijs, B., & Glänzel, W. (2010). A structural analysis of collaboration between European research institutes. *Research Evaluation*, 19(1), 55–65.
- Torvik, V. I., Weeber, M., Swanson, D. R., & Smalheiser, N. R. (2005). A probabilistic similarity metric for medline records: A model for author name disambiguation. *Journal of the American Society for Information Science and Technology*, 56(2), 140–158.
- Van Eck, N. J., & Waltman, L. (2008). Appropriate similarity measures for author co-citation analysis. *Journal of the American Society for Information Science and Technology*, 59(10), 1653–1661.
- Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.

- van Rijnsoever, F. J., Hessels, L. K., & Vandeberg, R. L. (2008). A resource-based view on the interactions of university researchers. *Research Policy*, 37(8), 1255–1266.
- Wagner, C. S. (2005). Six case studies of international collaboration in science. *Scientometrics*, 62(1), 3–26.
- Wallace, M. L., Larivière, V., & Gingras, Y. (2012). A small world of citations? The influence of collaboration networks on citation practices. *PLoS One*, 7(3), e33339.
- White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, 54(5), 423–434.
- White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163–171.
- White, H. D., & Griffith, B. C. (1982). Authors as markers of intellectual space: Co-citation in studies of science, technology and society. *Journal of Documentation*, 38(4), 255–272.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, 49(4), 327–355.
- White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure?: Longitudinal evidence from the “Globenet” interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, 55(2), 111–126.
- Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, 60(10), 2107–2118.
- Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and cword networks relate to each other. *Journal of the American Society for Information Science and Technology*, 63(7), 1313–1326.
- Zhao, D., & Strotmann, A. (2008a). Comparing all-author and first-author co-citation analyses of information science. *Journal of Informetrics*, 2(3), 229–239.
- Zhao, D., & Strotmann, A. (2008b). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, 59(13), 2070–2086.
- Zhao, D., & Strotmann, A. (2008c). Information science during the first decade of the web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology*, 59(6), 916–937.
- Zitt, M., Lelu, A., & Bassecoulard, E. (2011). Hybrid citation-word representations in science mapping: Portolan charts of research fields? *Journal of the American Society for Information Science and Technology*, 62(1), 19–39.