



Important institutions of interinstitutional scientific collaboration networks in materials science

Yang Li^{1,2,3} · Huajiao Li^{1,2,3}  · Nairong Liu^{1,2,3} · Xueyong Liu^{1,2,3}

Received: 30 October 2017 / Published online: 13 July 2018
© Akadémiai Kiadó, Budapest, Hungary 2018

Abstract

Interinstitutional scientific collaboration plays an important role in knowledge production and scientific development. Together with the increasing scale of scientific collaboration, a few institutions that positively participate in interinstitutional scientific collaboration are important in collaboration networks. However, whether becoming an important institution in collaboration networks could be a contributing factor to research success and how these important institutions collaborate are still indistinct. In this paper, we identified the scientific institutions that possess the highest degree centrality as important institutions of an interinstitutional scientific collaboration network in materials science and examined their collaboration preferences utilizing several network measures. We first visualized the appearance of these important institutions that had the most positive collaborations in the interinstitutional scientific collaboration networks during the period of 2005–2015 and found an obvious scale-free feature in interinstitutional scientific collaboration networks. Then, we measured the advantages of being important in collaboration networks to research performance and found that positive interinstitutional collaborations can always bring both publication advantages and citation advantages. Finally, we identified two collaboration preferences of these important institutions in collaboration networks—one type of important institution represented by the Chinese Academy of Science plays an intermediary role between domestic institutions and foreign institutions with high betweenness centrality and a low clustering coefficient. This type of important institution has better performance in the number of publications. The other type of important institution represented by MIT tends to collaborate with similar institutions that have positive collaborations and possess a larger citation growth rate. Our finding can provide a better understanding of important institutions' collaboration preferences and have significant reference for government policy and institutional collaboration strategies.

Keywords Network · Scientific collaboration · Materials science · Important institutions · Collaboration preference

✉ Huajiao Li
babyproud@126.com

Extended author information available on the last page of the article

Introduction

Global scientific collaboration is particularly interesting to study since it has grown at a remarkable rate since 1990 (Adams 2012). Attempts to describe the patterns of global scientific collaboration, in terms of who collaborates with others and how they collaborate, have been a preoccupation of bibliometric researchers (Wagner et al. 2017). The most commonly used approach is the bibliometric analysis of scientific co-authorship networks, as the data can be easily and accurately extracted from publication databases (Pike 2010). Co-authorship networks constitute an important class of social networks. A co-authorship network is defined as a collection of individuals, each of whom is acquainted with some subset of the others by one or more different types of relationships, such as friendship, kinship and co-authorship. There is a substantial body of literature in information science that addresses both co-authorship patterns (Egghe and Rousseau 1990; Melin and Persson 1996). In this paper, we constructed interinstitutional scientific collaboration networks (ISCNs) in which two institutions are considered connected if they have co-authored a paper in the field of materials science.

Scientific collaboration is defined by Sonnenwald as “interaction taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal” (Sonnenwald 2014). It is mainly identified by authoring records (Wuchty et al. 2007) and project records (White 1992). Due to the development of information and communication technologies, most of the publication records could be accessed through the Internet (Asadi et al. 2017). Increasingly more researchers are analyzing scientific collaborations based on co-authoring records (Bhattacharyya and Bandyopadhyay 2015; Shahadat et al. 2013) because they are criteria to determine the structures of social networks among scholars (Fuchs 2017).

It is widely understood that collaborative activity differs by field and sector (Choi et al. 2015). Wuchty examined many fields over decades and observed growth in collaborative activity in Physics, Environmental Science and Medicine using both article and patent records of the Web of Science database (Wuchty et al. 2007). Furthermore, Carrington examined collaboration networks in six fields: Physics, Biology, Environmental, Engineering, Agriculture and Mathematics (Carrington et al. 2005). Taşkın and Aydinoglu (2015) analyzed co-authored publications in astrobiology and investigated journal references. As scientific disciplines become increasingly diversified, there also appears to be an increasing amount of “communication” among various fields, thereby connecting professionals from different research backgrounds (Mattsson et al. 2008). More recent studies focus on interdisciplinary collaborations (Porter and Youtie 2009; Wagner et al. 2011). Karlovčec and Mladenčić (2015) investigated interdisciplinarity scientific fields and their evolution based on a graph of project collaboration and co-authorship. Avkiran (2013) investigated the impact of collaborative research in academic finance literature and found that collaboration can lead to articles of high impact. Although interdisciplinary collaborations were found to have a higher potential to foster research outcomes (Sigelman 2009), to the best of our knowledge, there has been a relative lack of research on collaborations in materials science, which is a syncretic discipline hybridizing metallurgy, ceramics, solid-state physics, and chemistry (Kockelmans 1979).

Previous research has shown that collaboration at different levels, such as the author level, institution level, and country level, have different characteristics (Li and Li 2015). At the author level, Wallace showed that an increasing share of citations come from collaborators, as well as the collaborators of collaborators (Wallace et al. 2011). Krongerger’s

study of researchers from four scientific disciplines was designed to understand which kinds of incentives, perceptions and personal strategies help account for collaboration from the perspectives of individual scientists (Kronegger et al. 2015). The researchers who occupy different positions in the network have their own characteristics and play different roles (Ebadi and Schiffauerova 2015). At the country level, territorial borders have become less important in international scientific collaboration (Hoekman et al. 2010). Furthermore, collaborations at the institutional level, which can be considered the authors' larger social network, have attracted much attention. Avkiran (2013) found that collaboration on high impact articles is mostly cross-institutional. In addition, it was found that institutions with similar research profiles were more likely to collaborate with each other (Thijs and Glänzel 2010). Notably, as the number of collaborations between two institutions increases, the likelihood of citations between them also increases (Yan and Ding 2012). Unfortunately, not every institution has the same importance in the scientific collaboration network, as a small number of institutions (which we called important institutions in the network) are very central, while many others are at the periphery (Chang and Huang 2013). For instance, there was an obvious heterogeneity among the distribution of scientific collaboration resources. Approximately 20% of the institutions occupied core positions and processed 80% of the collaboration relationship (Drożdż et al. 2017). Owing to the special roles that important institutions play, this paper performs further investigation to obtain more insight into these important institutions of ISCNs in materials science.

Studying scientific collaboration networks is important not only for understanding the entire picture of global scientific collaboration but also for uncovering the underlying mechanisms (Barabási et al. 2001; Dorogovtsev and Mendes 2002; Ghosh et al. 2015; Newman 2001). Athen et al. (2015) found important functional information characteristics of the system of scientific collaboration by examining topological structures of co-authorship networks. Çavuşoğlu and Türker (2014) identified small-world structures in co-authorship networks. Barabasi and Albert (1999) examined a formal model of cumulative advantage in terms of preferential attachment as the driving mechanism of co-authorship. They found a common property among many large networks: the node degrees of each network followed a scale-free power-law distribution. The implications of scale-free distributions have been widely used to understand scientific co-authorship networks (Moody 2004). Furthermore, this feature was found to be a consequence of two generic mechanisms: networks expand continuously with the addition of new nodes, and new nodes attach preferentially to nodes that are well connected. This preferential attachment mechanism gives superior nodes an accumulated advantage, which results in the appearance of hub nodes (Said et al. 2008).

The important institutions in a scientific collaboration network describe nodes which play an important role in structural and functional properties (Ma and Mondragón 2015). Several traditional network centrality measures such as degree, closeness, betweenness and PageRank have been applied in collaboration networks to estimate the importance of nodes (Porter and Youtie 2009). Also some complex methods such as the SIR model, HITS algorithm and structural holes theory have been used to measure the importance of nodes (Latora et al. 2013; Zhou et al. 2018). These studies describe importance from different perspectives (Zhou et al. 2018). In this paper, we define important institutions in inter-institutional scientific collaboration networks as institutions with high degree centrality in collaboration networks. Recent research revealed that collaborations among the 100 highest-impact institutions have increased faster than those among lower-impact institutions (Gazni and Thelwall 2016). They also found that the top-100 institutions cite the papers of other top-100 institutions more than those of lower-ranked institutions, which

indicates that they prefer to cite a narrower set of high-quality works, instead of the research that they wrote jointly with lower-impact institutions. Similarly, Athen et al. (2015) examined the global interconnectedness among the leading institutions, which corresponds to the nodes with the largest degrees, i.e., the so-called hubs, by detecting the rich-club phenomenon in the co-funding networks, and found an extremely heterogeneous funding distribution. Notwithstanding focal interests, only a few studies have applied network analysis to analyze the performance of important academic and research institutions (Abbasi et al. 2011). However, the configuration of the interinstitutional collaboration network in materials science remains unclear. Little attention has been given to discover which institutions occupy prominent positions in the interinstitutional scientific collaboration networks. Could becoming important in interinstitutional scientific collaboration networks be a contributing factor toward success in research? Do these important institutions have any collaboration preference? To answer these questions, in this paper, we first examined the heterogeneity of the ISCNS in materials science through degree distribution and visualized the appearance of important institutions. Second, the research performance of these important institutions' is discovered to determine whether being important in ISCNS can be a contributing factor toward success in research. Additionally, MIT and the Chinese Academy of Science are shown as outliers which have better research performance. Finally, we analyze the collaboration preference of MIT and the Chinese Academy of Science to explain the reason for their outstanding performance.

Data and method

Constructing collaboration networks

The data were extracted from the Web of Science database, one of the largest paper citation indices in the world. The attributes of data that were processed included the time of publication, the institutions and the titles. All data including the above three attributes which were indexed using the subject search "materials science" were downloaded. The results contained 32,592 records published between 2005 and 2015.

In this paper, an affiliation relationship can be found between articles and institutions which co-authored the articles. Networks constructed according to affiliation relationships are a typical two-mode member-network (Breiger 1974) or hyper-network (Mcpherson 1982). This network is composed of a set of actors (articles) and a set of events (co-authored institutions). According to Li et al., when there are two nodes, α and β , which have the same relationship with γ , then α and β are equivalent (Li et al. 2014). In the interinstitutional collaboration network, the institutions that co-authored one paper are equivalent. Therefore, we can construct a collaboration network through the co-authored relationship. We assumed that matrix A is the primitive $N \times M$ matrix of the affiliation relationships between articles and institutions. Then, we used formula (1) to obtain the derivative $N \times N$ matrix X of the collaboration network. Formula (1) shows the typical way to analyze two-mode networks and can be used to determine the number of the actors that the two events are co-containing as well as the number of the events at which the two actors are co-attendant (Breiger 1974). The demonstration of the constructing network is shown in Fig. 1.

$$X = A' \times A. \quad (1)$$

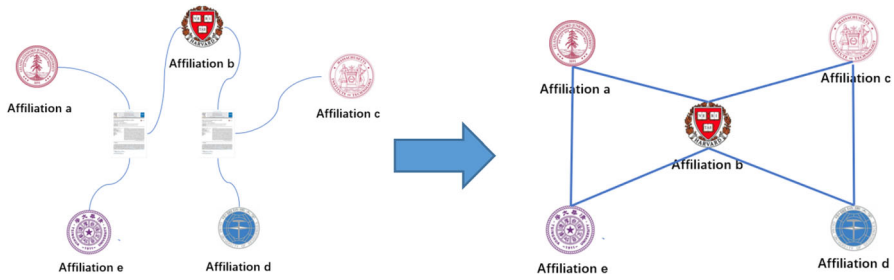


Fig. 1 Construction of a co-authorship collaboration network at the institutional level

Topological features of the network

Many indices can be used to analyze the topological features and evolution of scientific collaboration networks (Abbasi et al. 2011). In this article, we chose six major indices: degree, weighted degree, weighted degree assortativity, clustering coefficient, eigenvector centrality and betweenness centrality.

Degree

The node degree of a node, which is given by formula (2), indicates the number of nodes that are in contact with it (Freeman 1980). In our collaboration network, the node degree represents the number of collaborators of the institution.

$$d_i = \sum_{j=1}^n d_{ij}, \tag{2}$$

where d_i denotes the degree of node i ; $d_{ij} = 1$ if node i and node j have a co-relationship, and $d_{ij} = 0$ if node i and node j have no co-relationship. n is the total number of nodes in the network.

Weighted degree

The weighted degree of a node, which is given by formula (3), indicates the frequency with which this node is in contact with others (Freeman 1980). Therefore, the node weighted degree can well represent the positivity of participating in interinstitutional scientific collaborations.

$$w_i = \sum_{j=1}^n w_{ij}, \tag{3}$$

where w_i denotes the weighted degree of node i , and w_{ij} equals the number of co-relationships between node i and node j .

In a weighted network, if n_k nodes have the same weighted degree w and n is the total number of nodes in the network, the weighted degree distribution can be defined as

$$p(w) = n_w/n. \tag{4}$$

Moreover, if $p(w) \sim w^{-\lambda}$ 5), where λ is the power-law index and w is the weighted degree of the nodes, the weighted degree distribution satisfies the power law (Newman 2003b). Thus, most of the nodes have few connections, while a few crucial nodes have many connections.

Weighted degree assortativity

The weighted degree assortativity indicates the extent to which institutions in the network cooperate preferentially with other institutions with similar collaboration positivity (weighted degree). We calculate the weighted degree assortativity by formula (6 Newman 2003a):

$$r^w = \frac{H^{-1} \sum_{\phi} \left(w_{\phi} \prod_{i \in F(\phi)} k_i \right) - \left(\frac{H^{-1}}{2} \sum_{\phi} \left(w_{\phi} \sum_{i \in F(\phi)} k_i \right) \right)^2}{\frac{H^{-1}}{2} \sum_{\phi} \left(w_{\phi} \prod_{i \in F(\phi)} k_i \right)^2 - \left(\frac{H^{-1}}{2} \sum_{\phi} \left(w_{\phi} \sum_{i \in F(\phi)} k_i \right) \right)^2}, \quad (6)$$

where w_{ϕ} is the weighted degree of edge ϕ and $F(\phi)$ is the pair of nodes connected by edge ϕ . H is the sum of the weighted degrees of all edges in the network.

Finally, $r = -1$ when there is perfect disassortative mixing, $r = 0$ when there is no assortative mixing, and $r = 1$ when there is perfect assortative mixing. In other words, weighted degree assortativity indicates the level of homophily among institutions and their corresponding cooperators. For instance, if the coefficient r^w is equal to 1, institutions in the network only cooperate with cooperators with the same positivity. In contrast, if r^w is equal to -1 , scientific institutions with high positivity prefer cooperating with institutions with low positivity. Moreover, if r^w is equal to 0, there are no assortative collaboration patterns among these institutions.

Clustering coefficient

The clustering coefficient is always used to measure the connectivity of the neighbors of a given node. We use the average clustering coefficient, which is given by formula (7), to analyze the closeness and strength of the collaboration relationship (Goh et al. 2003).

$$C = \langle C_i \rangle_i = \left\langle \frac{2E_i}{P_i(p_i - 1)} \right\rangle, \quad (7)$$

where E_i is the number of links between the neighbors of node i , and $\frac{1}{2}P_i(p_i - 1)$ is the total number of possible links between the neighbors of node i . If the value of C is high, the corresponding nodes in the collaboration network form a cohesive circle with a relatively unimportant position, whereas if the value of C is low, the corresponding nodes in the collaboration network form a circle with a relatively important position.

Betweenness centrality

The concept of betweenness centrality was first introduced by Freeman, and is defined as “the number of shortest paths (between all pairs of nodes) that pass-through a given node”. The betweenness centrality of a node is given by formula (8) (Freeman 1980).

$$C_B(ni) = \sum_{j < k} g_{jk}(ni) / g_{jk}, \quad (8)$$

where $g_{jk}(ni)$ is the number of paths through node i from node j to node k and g_{jk} is the total number of the paths from node j to node k .

Betweenness centrality is an indicator of the potential that a node (institution) will play the role of a broker or gatekeeper. Such a node frequently controls the resource flows in the collaboration network. A node with high betweenness centrality is located in a central position.

Eigenvector centrality

Eigenvector centrality is a measure of the importance of a node in the network. This index assigns relative scores to all nodes in the network according to the following concept: connections to high-scoring nodes contribute more than those to the low-scoring nodes. We use formula (9) to calculate eigenvector centrality (Bonacich and Lloyd 2001). The institutions with high eigenvector centrality in our collaboration networks always possess abundant collaboration resources and belong to an elite scientific collaboration circle.

$$C_e(i) = \lambda^{-1} \sum_{j=1}^n a_{ij}e_j, \tag{9}$$

where n is the total number of nodes, and A represents the adjacency matrix of the network. If node i and node j have collaboration links with each other, $a_{ij} = 1$. In contrast, if node i and node j have no collaboration link, $a_{ij} = 0$. $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A , and the eigenvector that corresponds to λ_i is $a = (e_1, e_2, \dots, e_n)$.

Nodes with high eigenvector centrality are connected to nodes that have high weighted degree.

Results and discussion

Appearance of important institutions

We collected 32,592 publications from the period 2005–2015 from WOS (www.webofknowledge.com), of which 10,525 were multi-institutional. The percentage of multi-institutional publications was approximately 40% during 2005–2012, and increased rapidly to approximately 57% in 2015, as Fig. 2 shows, which suggests an increase in scientific collaboration in the field of materials science since 2013.

As an increasing number of multi-institutional publications appeared, scientific collaboration among institutions gradually became mainstream. According to Fig. 3, the scale of the collaboration network expanded year after year. The number of nodes increased from 4367 to 10,631, and some nodes with a very high degree or weighted degree appeared, as shown in Fig. 4. This implies that in the field of materials science, more institutions gradually realized the significance of collaboration and actively participated in interinstitutional scientific collaboration. In addition, some “powerful” institutions allocated their resources to participating in interinstitutional scientific collaboration.

Although we know that the collaboration scale has increased over time, its internal variation is still poorly understood. Here, we exploited the node weighted degree distribution, which reflects the distribution of collaboration levels, to determine the inherent characteristics of the collaborations. As the fat tail characteristic appears in the node degree distributions, we examined the power-law distribution of the node degree distribution (Barabási and Albert 1999). The formation of a power-law distribution is given by

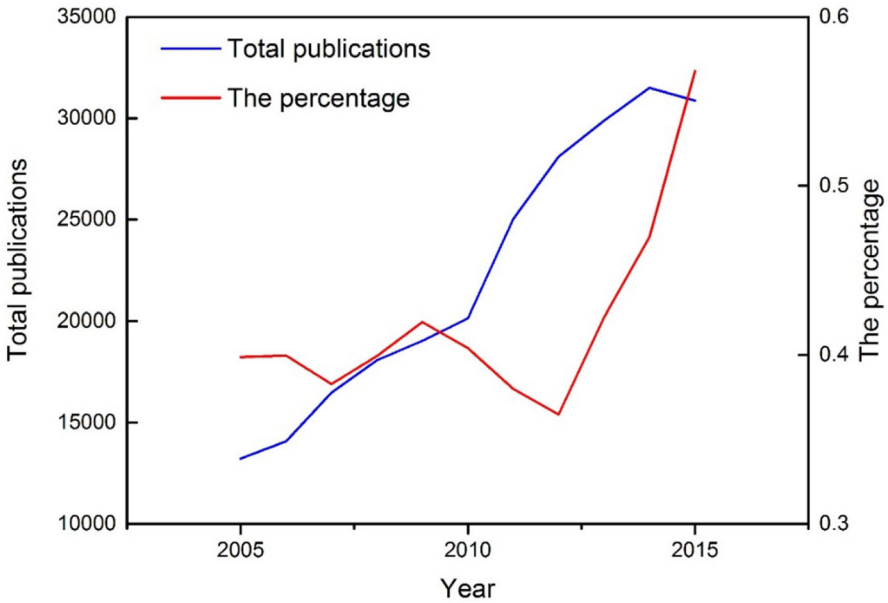


Fig. 2 Total number of publications and the percentage of multi-institutional publications

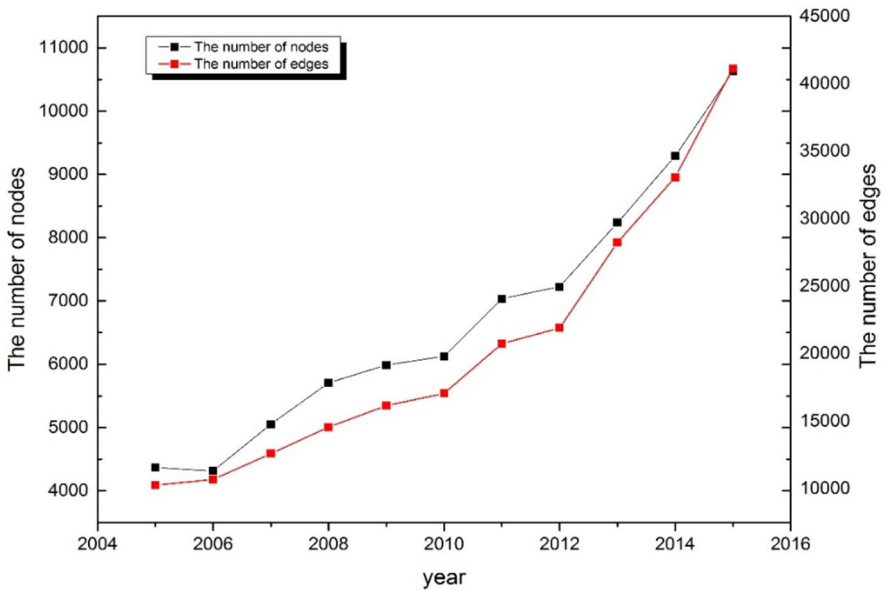


Fig. 3 The numbers of nodes and edges of a global interinstitutional scientific collaboration network, 2005–2015

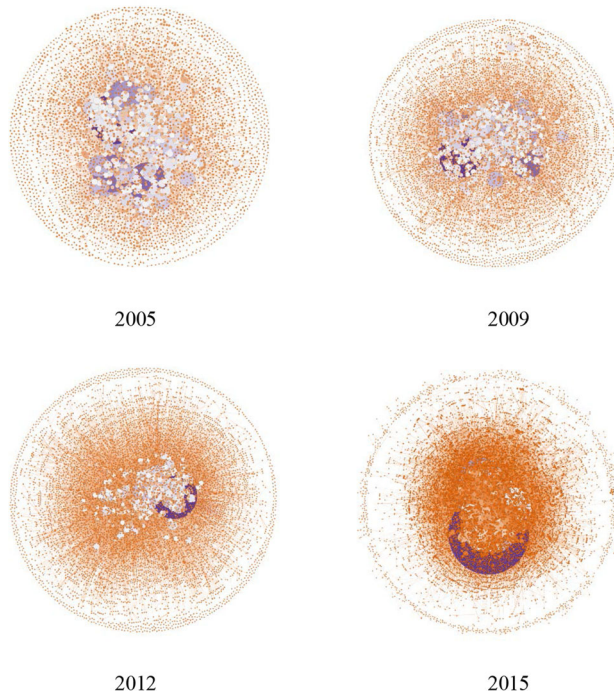


Fig. 4 Visualization of a global interinstitutional scientific collaboration network in 2005, 2009, 2012 and 2015

$p(x) \sim x^{-\lambda}$. $P(x)$ means the distribution of nodes degree, x means the node degree and λ is the scaling parameter. As shown in Fig. 5 and Table 1, we calculated the fitting function of the node weighted degree distribution in log–log coordinates from 2005 to 2015. The measure of goodness of fit (R^2) was always larger than 0.85. It indicated that the nodes degree distribution followed the power-law distribution. Various complex networks with a power-law distribution are highly heterogeneous and perform scale-free characteristics. Thus, these interinstitutional collaboration networks possess scale-free characteristics. Thus, few important institutions in the network have many collaboration links, whereas most institutions have few collaboration links; this suggests a high level of focused collaboration in select institutions. In addition to heterogeneity, the scale-free characteristics of social networks indicate an expanding tendency, which supports the increasing scale of collaboration. Another important scale-free characteristic is the preferential attachment mechanism: in global interinstitutional scientific collaboration in materials science, new institutions prefer to engage in interinstitutional scientific collaboration with the important institutions. In other words, the important institutions have a catalytic effect on the expanding collaboration scale and play an important role in promoting global interinstitutional scientific collaboration (Table 1).

Advantages of being important in ISCN

In this article, important institutions are defined as the top-20 nodes in terms of degree, as such nodes have both the largest-scale collaboration behavior and play a dominating role in

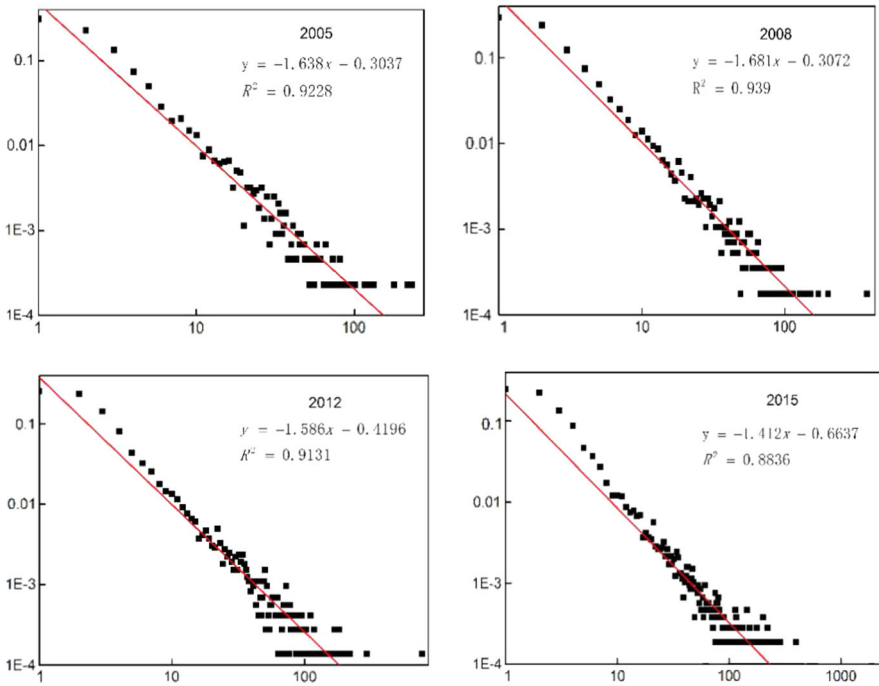


Fig. 5 Node weighted degree distributions in log–log coordinates

Table 1 The indices of the fitting function

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
λ	1.683	1.718	1.657	1.681	1.641	1.623	1.557	1.586	1.495	1.474	1.412
R^2	0.9228	0.9211	0.9022	0.939	0.9059	0.922	0.9075	0.9131	0.8797	0.8776	0.8836

structural and functional properties. However, is being a hub in an ISCN a contributing factor toward success in research? To identify the benefits of being a hub in an ISCN, we investigated the research performance of such institutions in terms of papers accepted each year into the WOS core collection and *h* index as shown in Figs. 6 and 7. The WOS core collection is the world’s leading scholarly literature database; the *h* index of a researcher is the number of papers co-authored by the researcher with at least *h* citations each (Hirsch 2005). A variety of studies have shown that the *h* index, by and large, agrees with other objective and subjective measures of scientific quality in a variety of different disciplines (Bornmann and Daniel 2005; Hirsch 2005; Raan 2006). The *h* index, annual publications and growth rate of citations of important institutions are shown in Table 2. First, we found that the average number of papers accepted by the WOS core collection over 11 years is linearly correlated with the number of times N_{core} that the institutions were presented as hubs, except for the Chinese Academy of Sciences, which has published papers that were accepted by the WOS core collection beyond what is expected from linear behavior. Furthermore, we referred to the *h* index of core institutions’ publications during 2005–2015

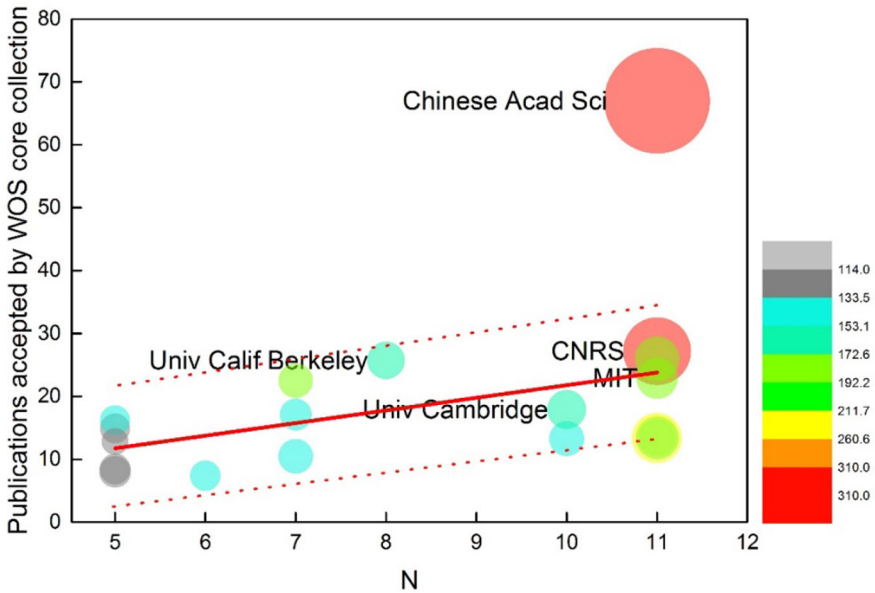


Fig. 6 Publications accepted by the WOS core collection and N values of core institutions

to capture the research depth. We observed that a few institutions outperformed the rest of the institutions, which was a marked deviation from the linear behavior that was observed at the collaboration level. Their outstanding collaboration profiles seem to enable hubs to develop the depth of research that leads to the generation of high-quality papers.

Collaboration preferences of important institutions

For an ordinary institution, constructing collaboration links with important institutions such as MIT is definitely a good way to develop research and expand academic reputation, as important institutions are always bellwethers of research direction and have already accumulated abundant resources and achievement. However, fierce competition among important institutions with high-level impact makes their relationships much more complex, although active interinstitutional scientific collaboration among important institutions has brought numerous benefits such as more publications and higher-quality research. In this part, we uncovered different collaboration preferences of important institutions in ISCNS. We used four indices to describe their collaboration preferences: eigenvector centrality, betweenness centrality, clustering coefficient and assortativity (Figs. 8, 9).

Effect of cooperators’ collaboration positivity on important institutions

The higher the eigenvector centrality is, the higher the collaboration level of one’s cooperators is. In other words, eigenvector centrality reflects the cooperators’ collaboration positivity. First, to examine the impact of cooperators’ collaboration positivity on citation advantages, we fitted data on core institutions’ annual citations linearly and obtained the slope k , which is the annual citation growth rate. Then, we examined the relationship between the eigenvector centrality of core institutions and these institutions’ slope k .

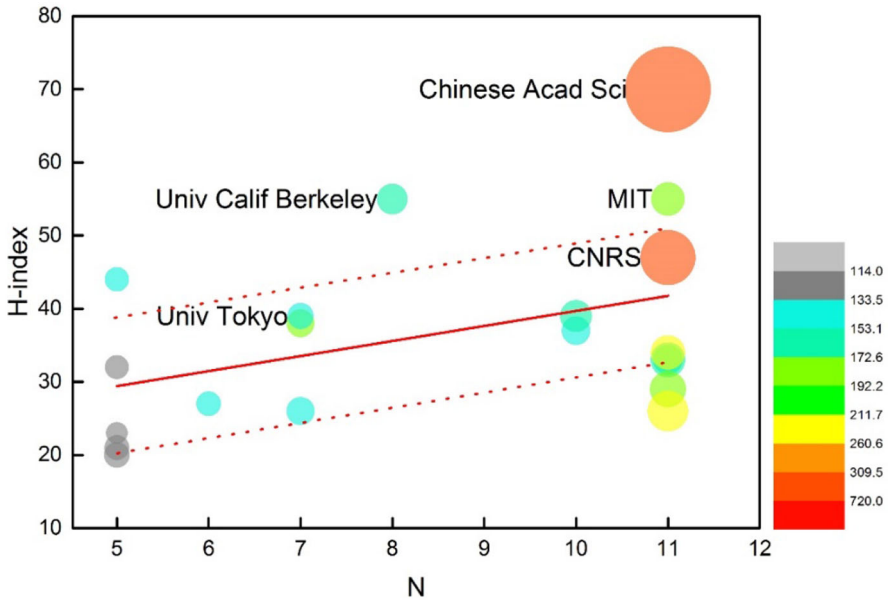


Fig. 7 h -index and N values of core institutions

According to Fig. 8, the annual citation growth rate over these 11 years is linearly correlated with eigenvector centrality, except for MIT, which accumulated outstanding scientific achievements and enjoyed a high reputation in the field of materials science. In general, collaborating with active cooperators is a good collaboration strategy for important institutions, as their cooperators' collaboration positivity brings them obvious citation advantages.

Two different collaboration preferences between important institutions

An institution is said to be an intermediary if it occupies an advantageous position in the network for developing opportunities through its connections. We use the betweenness centrality and clustering coefficient to quantify an institution's mediating effect. In a collaboration network, the betweenness centrality of an institution reflects the superiority of the position that it occupies in the network. The collaboration network would crash in a short time without the institutions with the highest betweenness centrality. In addition, institutions with high betweenness centrality always possess abundant resources and control the crucial points of the information flow. The clustering coefficient ranges from 0 to 1. It attains its smallest value when an institution is the center of a star, and there is no connection between any of its partners, and it attains its largest value when an institution is a general member in its circle, as its partners are fully connected to each other.

According to Fig. 9, institutions such as the Chinese Academy of Sciences and the Russian Academy of Science with low clustering coefficients (less than 0.04) always have high betweenness centrality, which means that these institutions are the center of their circles and act as intermediaries in their collaboration networks. Although institutions such as MIT, University of Cambridge, and University of California, Berkeley also have

Table 2 The *h* index, annual publications and growth of citations of important institutions

Important institutions	<i>H</i> index	Annual publications	Growth rate of citations
CHINESE ACAD SCI	72	67.45	372.91
MIT	56	23.00	413.65
CNRS	48	27.45	226.45
RUSSIAN ACADEMY OF SCIENCES	38	34.64	64.25
NASA	39	27.82	83.86
UNIVERSITY OF LONDON	48	26.64	108.93
UNIVERSITY OF CALIFORNIA BERKELEY	56	24.18	140.84
CNR	37	22.27	81.71
HARVARD UNIVERSITY	53	21.00	163.23
UNIVERSITY OF TEXAS SYSTEM	41	19.45	158.01
CALIFORNIA INSTITUTE OF TECHNOLOGY	44	19.36	75.95
UNIVERSITY COLLEGE LONDON	38	16.55	66.53
LBNA	52	16.36	103.11
TSINGHUA UNIVERSITY	43	17.09	82.64
NORTHWESTERN UNIVERSITY	49	16.73	120.57
INDIA INST TECHNOL	22	8.00	28.16
CSIC	35	13.36	59.71
TOHOKU UNIV	28	14.00	39.87
UNIV CAMBRIDGE	43	18.18	73.99
OSAKA UNIV	26	13.00	35.88

relatively high betweenness centrality, they do not have dominant positions among their cooperators with relatively high clustering coefficients (larger than 0.04).

We can easily divide these 20 important institutions into two groups. The first group contains the Indian Inst Technol, the Chinese Acad Sci, Russian Acad Sci, CNRS and CSIC, whose clustering coefficients are less than 0.04. We found that institutions in this group are all nationally supported institutions. Furthermore, most of their cooperators are from their own country, and most of their foreign partners are world-class institutions. The other group contains Tohoku Univ, CNR, Tech Univ Dresden, Univ Cambridge, Osaka Univ, MIT, Natl Inst Mat Sci, Georgia Inst Technol, Univ Tokyo, Oak Ridge Natl Lab, Seoul Natl Univ, Univ Calif Berkeley, Los Alamos Natl Lab, Northwestern Univ and Tsinghua Univ, whose clustering coefficients are larger than 0.04. Here, we chose the Chinese Acad Sci and MIT as examples of each group, due to their outstanding research performance and obvious difference in collaboration preference.

To make the difference clearer, we calculated the assortativity of the Chinese Acad Sci, MIT, and the whole collaboration network, as shown in Fig. 10. Due to the extremely large scale of the whole collaboration network, its assortativity was close to 0, which indicates that collaboration patterns at the institutional level do not present obvious heterogeneity or homogeneity. Surprisingly, this result is different from that obtained in research on collaboration at the author level (Fafchamps et al. 2010). They underlined that collaboration is most likely between authors of similar levels of ability, which implies positive assortativity.

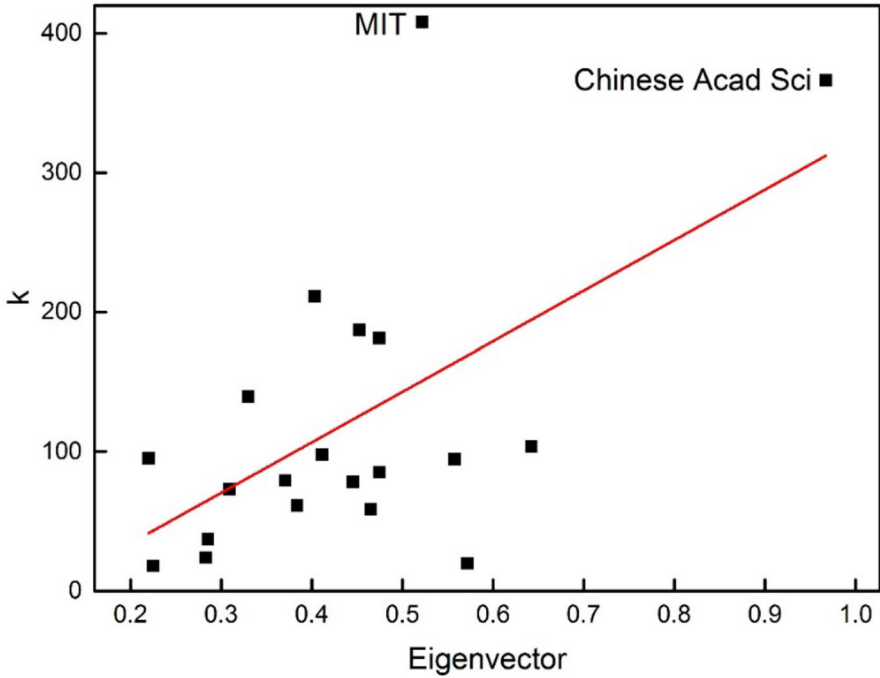


Fig. 8 Eigenvector and slope k of core institutions

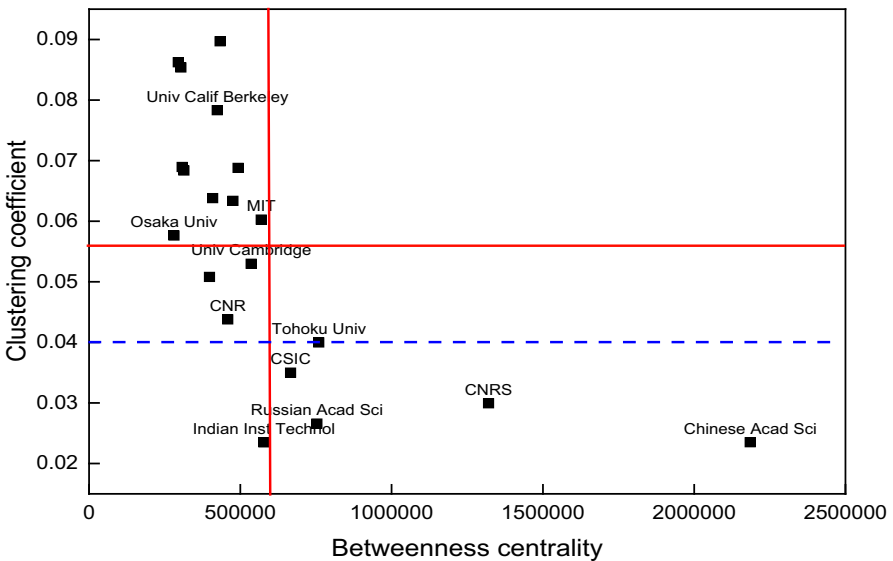


Fig. 9 Clustering coefficient and betweenness centrality of core institutions

Although the weighted degree assortativity of the Chinese Acad Sci and MIT are much lower than the average level, the Chinese Acad Sci’s weighted degree assortativity (which is close to -1) is much lower than that of MIT, which means the Chinese Acad Sci prefers

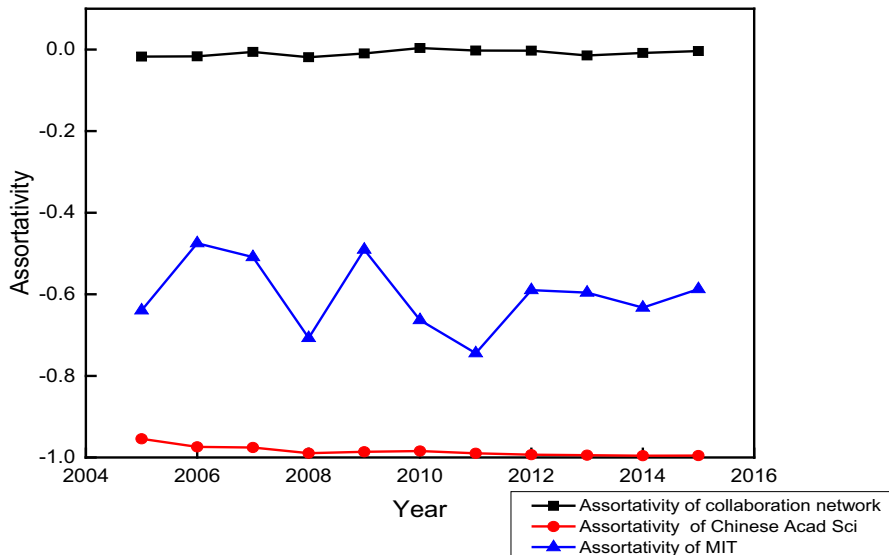


Fig. 10 Assortativity of the Chinese Acad Sci, MIT and the whole collaboration network, 2005–2015

collaborating with inactive institutions to a greater degree than MIT. In addition, we calculated the percentage of domestic collaboration and found that the Chinese Acad Sci has more domestic collaboration links than MIT. As a national institution in China, the Chinese Acad Sci, with large betweenness centrality and small clustering coefficient, is the center of its collaboration circle and plays the role of an intermediary between domestic partners and foreign world-level partners. As the Chinese Acad Sci actively collaborates with domestic institutions, it has had an apparent catalytic effect on connecting domestic institutions to global interinstitutional scientific collaboration and, in return, it has obtained obvious citation advantages from its active collaboration. Moreover, the Chinese Acad Sci has aggregated domestic research resources to keep pace with other worldwide top-level institutions. However, MIT seems to have less interest in promoting the performance of native institutions in the field of materials science, compared to the Chinese Acad Sci. Although MIT has the largest number of citations almost every year, its collaboration positivity is much lower than that of the Chinese Acad Sci. Although the Chinese Acad Sci has many more cooperators, the average collaboration positivity of MIT’s partners is much larger than that of the Chinese Acad Sci’s cooperators. This suggests that MIT is more selective when choosing partners. The difference in collaboration preferences between the two groups that we mentioned before has been made much clearer through this example. The institutions in the same group (whose clustering coefficients are less than 0.04) as the Chinese Acad Sci stressed their catalytic effect by acting as intermediary agents between domestic partners and foreign top-level institutions, while the other group (whose clustering coefficients are larger than 0.04) paid more attention to their own development by collaborating with high-impact cooperators.

The different collaboration preferences between these two groups can be explained very well by previous research. Kronegger identified the main factors driving collaboration at two levels: research-policy-related external factors and internal factors affecting the motivations of scientists regarding compatibility, cultural proximity and academic

excellence (Kronegger et al. 2015). Institutions in the first group, which contains the Chinese Acad Sci, are more likely to be affected by external factors, as they are influenced by their national research policy to pay more attention to supporting the domestic development of materials science research. In contrast, institutions in the other group are more likely to be affected by internal factors; however, the factors that affect their collaboration preferences, such as compatibility, cultural proximity and academic excellence, require future research.

Conclusion

In this paper, we constructed 11 interinstitutional scientific collaboration networks based on co-authorship over 11 years to gain insight into important institutions in the field of materials science. Although the scale of institutional scientific collaboration networks is becoming larger, the networks still have apparent heterogeneity and fit a power-law distribution with high goodness of fit. The institutions with high degree which we called important institutions over-attracted a large number of co-authorships. Furthermore, we used the number of publications and the h index to represent these institutions' research performance and found an apparent positive correlation between research performance and the number of times of being important. Becoming important in the collaboration network is truly a contributing factor to research success. In addition, among these institutions, two institutions—the Chinese Academy of Science and MIT are important institutions through all time periods and have outstanding research performance exceeding the linear behavior. The Chinese Academy of Science had the largest number of publications accepted by the Web of Science Core collection. The Chinese Academy plays an intermediary role between domestic institutions with low degrees and foreign top-level institutions with high betweenness centrality and a low clustering coefficient. MIT has the largest growth of citations and large eigenvector centrality which means its cooperators have a relatively high collaboration positivity. In contrast with the Chinese Academy of Science, MIT has a relatively high assortativity which means it prefers collaborating with the positive institutions with high degree. It may be the possible reason for MIT's larger citation growth rate.

There is still some work needed to improve our study. On the one hand, we merely use co-authoring records to identify scientific collaboration limited to the availability of data. On the other hand, we discovered the difference between two types of important institutions in interinstitutional scientific collaboration networks. In future research, we aim to further uncover the differences in collaboration preference among elite institutions and establish an econometric model to determine the reason for the difference.

Acknowledgements This research is supported by grants from the National Natural Science Foundation of China (Grant No. 41701121). The authors would like to express their gratitude to Prof. Haizhong An, Dr. Xiangyun Gao and Dr. Shupeí Huang who provided valuable suggestions, and AJE-American Journal Experts who provided professional suggestions about language usage, spelling, and grammar.

References

- Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. *TEMEP Discussion Papers*, 5, 594–607.
- Adams, J. (2012). Collaborations: The rise of research networks. *Nature*, 490, 335.

- Asadi, S., Hussin, A. R. C., & Dahlan, H. M. (2017). Organizational research in the field of Green IT: A systematic literature review from 2007 to 2016. *Telematics and Informatics*, *34*, 1191–1249.
- Athen, M., Mondragón, R. J., & Vito, L. (2015). Anatomy of funded research in science. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 14760.
- Avkiran, N. K. (2013). An empirical investigation of the influence of collaboration in Finance on article impact. *Scientometrics*, *95*, 911–925.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512.
- Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2001). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3), 590–614.
- Bhattacharyya, M., & Bandyopadhyay, S. (2015). Finding quasi core with simulated stacked neural networks. *Information Sciences*, *294*, 1–14.
- Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, *23*, 191–201.
- Bornmann, L., & Daniel, H. D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, *65*, 391–392.
- Breiger, R. L. (1974). The Duality of Persons and Groups. *Social Forces*, *53*, 181–190.
- Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis*. Cambridge: Cambridge University Press.
- Çavuşoğlu, A., & Türker, İ. (2014). Patterns of collaboration in four scientific disciplines of the Turkish collaboration network. *Physica A: Statistical Mechanics and its Applications*, *413*, 220–229.
- Chang, H. W., & Huang, M. H. (2013). *Prominent institutions in international collaboration network in astronomy and astrophysics*. New York: Springer.
- Choi, S., Yang, S. W., & Han, W. P. (2015). The triple helix and international collaboration in science. *Journal of the Association for Information Science and Technology*, *66*, 201–212.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, *51*, 1079–1187.
- Drożdż, S., Kulig, A., Kwapien, J., Niewiarowski, A., & Stanuszek, M. (2017). Hierarchical organization of H. Eugene Stanley scientific collaboration community in weighted network representation. *Journal of Informetrics*, *11*, 1114–1127.
- Ebadi, A., & Schiffauerova, A. (2015). How to become an important player in scientific collaboration networks? *Journal of Informetrics*, *9*, 809–825.
- Egghe, L., & Rousseau, R. (1990). Introduction to Informetrics. *Information Processing and Management*, *28*, 1–3.
- Fafchamps, M., Leij, M. J. V. D., & Goyal, S. (2010). Matching and network effects. *Journal of the European Economic Association*, *8*, 203–231.
- Freeman, L. C., Roeder, D., & Mulholland, R. R. (1980). Centrality in social networks: II. experimental result. *Social Networks*, *2*, 119–141.
- Fuchs, C. (2017). Sustainability and community networks. *Telematics and Informatics*, *34*, 628–639.
- Gazni, A., & Thelwall, M. (2016). The citation impact of collaboration between top institutions: A temporal analysis. *Research Evaluation*, *25*, 219–229.
- Ghosh, J., Kshitij, A., & Kadyan, S. (2015). Functional information characteristics of large-scale research collaboration: network measures and implications. *Scientometrics*, *102*, 1207–1239.
- Goh, K. I., Oh, E., Kahng, B., & Kim, D. (2003). Betweenness centrality correlation in social networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *67*, 017101.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 16569–16572.
- Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, *39*, 662–673.
- Karlovčec, M., & Mladenčić, D. (2015). Interdisciplinarity of scientific fields and its evolution based on graph of project collaboration and co-authoring. *Scientometrics*, *102*, 433–454.
- Kockelmans, J. J. (Ed.). (1979). *Interdisciplinarity and higher education*. State College: The Pennsylvania State Univ.
- Kronegger, L., Mali, F., Anu, X., Ferligoj, K., & Doreian, P. (2015). Classifying scientific disciplines in Slovenia: A study of the evolution of collaboration structures. *Journal of the Association for Information Science and Technology*, *66*, 321–339.
- Latora, V., Nicosia, V., & Panzarasa, P. (2013). Social cohesion, structural holes, and a tale of two measures. *Journal of Statistical Physics*, *151*, 745–764.
- Li, H. J., An, H. Z., Huang, J. C., Gao, X. Y., & Shi, Y. L. (2014). Correlation of the holding behaviour of the holding-based network of Chinese fund management companies based on the node topological characteristics. *Acta Physica Sinica*, *63*, 048901–048913.

- Li, J., & Li, Y. (2015). Patterns and evolution of coauthorship in China's humanities and social sciences. *Scientometrics*, *102*, 1997–2010.
- Ma, A., & Mondragón, R. J. (2015). Rich-cores in networks. *PLoS ONE*, *10*, e0119678.
- Mattsson, P., Laget, P., Nilsson, A., & Sundberg, C. J. (2008). Intra-EU vs. extra-EU scientific co-publication patterns in EU. *Scientometrics*, *75*, 555–574.
- McPherson, J. M. (1982). Hypernetwork sampling: duality and differentiation among voluntary organizations ☆. *Social Networks*, *3*, 225–249.
- Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, *36*, 363–377.
- Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, *69*, 213–238.
- Newman, M. E. J. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, *64*, 016131.
- Newman, M. E. (2003a). Mixing patterns in networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *67*, 026126.
- Newman, M. E. J. (2003b). The structure and function of complex networks. *SIAM Review*, *45*, 167–256.
- Pike, T. W. (2010). Collaboration networks and scientific impact among behavioral ecologists. *Behavioral Ecology*, *21*, 431–435.
- Porter, A. L., & Youtie, J. (2009). How interdisciplinary is nanotechnology? *Journal of Nanoparticle Research*, *11*, 1023–1041.
- Raan, A. F. J. V. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, *67*, 491–502.
- Said, Y. H., Wegman, E. J., Sharabati, W. K., & Rigsby, J. T. (2008). RETRACTED: Social networks of author-coauthor relationships. *Computational Statistics & Data Analysis*, *52*, 2177–2184.
- Shahadat, U., Liaquat, H., & Kim, R. (2013). Network effects on scientific collaborations. *PLoS ONE*, *8*, e57546.
- Sigelman, L. (2009). Are two (or three or four... or nine) heads better than one? Collaboration, multidisciplinary, and publishability. *PS Political Science and Politics*, *42*, 507–512.
- Sonnenwald, D. H. (2014). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*, 643–681.
- Taşkın, Z., & Aydinoglu, A. U. (2015). Collaborative interdisciplinary astrobiology research: a bibliometric study of the NASA Astrobiology Institute. *Scientometrics*, *103*, 1003–1022.
- Thijs, B., & Glänzel, W. (2010). A structural analysis of collaboration between European research institutes. *Research Evaluation*, *19*, 55–65.
- Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, *5*, 14–26.
- Wagner, C. S., Whetsell, T. A., & Leydesdorff, L. (2017). Growth of international collaboration in science: revisiting six specialties. *Scientometrics*, *110*, 1633–1652.
- Wallace, M. L., Larivière, V., & Gingras, Y. (2011). A small world of citations? The influence of collaboration networks on citation practices. *PLoS ONE*, *7*, e33339.
- White, J. C. (1992). Publication rates and trends in international collaborations for astronomers in developing countries, Eastern European countries, and the former Soviet Union. *Publications of the Astronomical Society of the Pacific*, *104*, 472.
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*, 1036–1039.
- 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*. Hoboken: Wiley.
- Zhou, J., Zeng, A., Fan, Y., & Di, Z. (2018). Identifying important scholars via directed scientific collaboration networks. *Scientometrics*, *114*, 1327–1343.

Affiliations

Yang Li^{1,2,3} · Huajiao Li^{1,2,3}  · Nairong Liu^{1,2,3} · Xueyong Liu^{1,2,3}

¹ School of Economics and Management, China University of Geosciences, Beijing 100083, China

² Key Laboratory of Carrying Capacity Assessment for Resource and Environment, Ministry of Land and Resources, Beijing 100083, China

³ Key Laboratory of Strategic Studies, Ministry of Land and Resources, Beijing 100812, China