Using Relative Tie Strength to Identify Core Teams of Scientific Research

Meishu Zhang [(✉)], Yu Jia, Nianxin Wang, Shilun Ge
Jiangsu University of Science and Technology, Zhenjiang, China
`zhangms@just.edu.cn`

Abstract—In China, it has long become imperative for the management of education, science and technology to build up high-level teams of scientific and technological innovation. Scientifically and accurately identifying core scientific research teams is an important prerequisite for cultivating and building such teams. The absolute threshold method (e.g., c-level clique at, n- clique, k-core) is the prevailing means of identifying core teams and their core members. Multiple factors such as "the preference-dependent effect", "the apostle effect" and "the star effect", and the cooperative relationship between the researchers could be considered. This study, based on the co-authorship network, found that not choosing the absolute threshold properly could easily lead to poor identification of core members of some teams. Even worse, when the absolute threshold is too large, this "uniform" evaluation criterion of tie strength results in the elimination of some core teams in some disciplines. This paper uses the relative tie strength to identify core scientific research teams from a new perspective, which can effectively avoid the situation of some core team members being ignored because of the mandatory requirements of the absolute tie strength among members, and can also solve the challenge of threshold selections for identifying different teams.

Keywords—Scientific cooperation, Relative strength of cooperative relationship, Team leaders, Team core members, Scientific research team, Science of Science

1 Introduction

As the world moves from the age of "small science" to the age of "big science", collaboration has become the main characteristic of modern scientific work [1, 2]. Scientific research teams are the basic form of scientific collaboration [3]. They have leaders at the core and include researchers who collaborate to pursue scientific goals. They also have increasingly become important innovation performance units in various research organizations of science and technology. In China, it has long become imperative for the management of education, science and technology to build up high-level teams of scientific and technological innovation. Scientifically and accurately identifying core scientific research teams is a critical step for cultivating and building

such teams. Scientific research collaboration exists widely in colleges and universities, and a stable core scientific research team arises from long scientific research collaboration. It is of great importance for colleges and universities to consolidate their fragmented scientific research efforts and strengthen their scientific research core team construction in order to enhance their discipline cohesion and raise their academic levels.

The absolute threshold method (e.g., c-level clique, n- clique, k-core) is the prevailing means of identifying core teams and their core members. Based on the co-authorship network, this study brings forward a new method for identifying core scientific research teams from a new perspective of the relative tie strength. This method allows core members of different scientific research teams to be identified. It can provide a useful technique for comprehensive universities with many disciplines to scientifically and accurately identify core scientific research teams, and it can also solve the challenge of threshold selections for identifying different teams. Current study addresses two research questions. Those are:

- How can we identify core teams of different scientific research teams based on relative tie strength?
- What is the difference in the identification results from the methods based on the absolute threshold or from the relative tie strength?

2 Literature Review

Social network analysis (SNA) has become the mainstream method of identifying scientific research teams [4-6]. The identification of scientific research teams mainly includes two aspects: selection of team leaders and identification of team members. For the selection of team leaders, some scholars choose the degree centrality as an index, by which team leaders are selected on the basis of their "core" status in the team [7, 8], while others adopt the betweenness centrality as an index, by which team leaders are identified according to their advantage in access to resources and their ability of "controlling" members [9, 10]. For the identification of team members, scholars mostly focus on the identification of a cohesive subgroup in a team, which is apparent in the following instances. Based on the similarity of team members' research activities, Calero (2016) used the k-core research method to identify cohesive research groups in the field of nanotechnology, whereby researchers who have more than k adjacent nodes in a group are clustered into a subgroup [11]. Based on the reachability of the connected nodes, Liu et al. (2015) separately analyzed the secondary research groups in the fields of genetic engineering and materials through the 2-clique method, whereby researchers whose geodesic distance between the two nodes is less than 2 are all aggregated into a subgroup [12]. Furthermore, based on the relationship strength among members, Reffay and Chanier (2003) differentiated tight-knit groups in specific environments and specific subject areas through the c-level clique method [13]. Finally, based on the collaboration pattern among members, Newman (2001, 2004) employed a component-analysis method to look into the network structure of scientific collaboration in the fields of bio-medicine, physics and mathematics,

and found that the collaboration pattern among researchers exemplifies the characteristics of the small world structure, which led him to conclude that the distribution of individual collaborators in all areas is uneven, with most people having only a few collaborators, while a few have as many as tens, hundreds, or even thousands of collaborators [14, 15]. These studies demonstrated that the research of team identification based on scientific collaboration networks is remarkably fruitful. To select team leaders, scholars generally introduce the centrality measurement index; and to identify team members, they mainly use the research method of cohesive subgroups, some absolute threshold determined by the scholar in a subjective manner on the basis of his or her research needs.

However, there exist "preferential attachment effect" [16], the "apostle effect" [17], and the "star effect" [18] in social network. Barabási and Albert (1999) suggested that networks grow continuously by the addition of new actors, which attach preferentially to well-connected ones [19]. Petersen (2015) borrowed "apostle" to describe their findings, where a distinguished cooperator selected according to his or her noteworthy attributes from among a large pool of candidates [17]. John Scott (2017) pointed that a central actor was one "at the center" of a number of connections and with many direct contacts with other actors [18]. Meanwhile, these phenomena exist in scientific research activities. As Petersen pointed out, there exist not only weak ties and strong ties, but also super ties among researchers in scientific research collaboration [17]. Hence, cooperative relationship among scientific researchers is not even. When choosing partners, individuals are more likely to cooperate with members who are at the center or well-connected than with other ones of team networks, that is, with "star" members of the team. A core scientific research team is a congregation of researchers who have strong relationships with the team leaders. Given that teams vary in means of collaboration, the introduction of the absolute threshold as an evaluating indicator for team identification can lead to a situation where some core members of a team are neglected due to the "one-size-fits-all" absolute threshold. This in turn can lead to an inadequate identification of a scientific research team.

To solve the issue of poor identification of core members for a scientific research team, this study presents a new method, which uses the relative tie strength of collaboration as an evaluation criterion for team identification to identify the core members of the scientific research team. This method can effectively avoid the situation where some core members of the team are left out due to the absolute threshold requirements, and thus eliminate the possibility that these core teams are not recognized due to the falsely equal treatment of the absolute threshold.

3 Methodology

The main steps of our study are illustrated in Figure 1. An important concern in social network studies is how to define the network boundaries, or what actors deserve to be included in the network. To observe relatively important researchers and their cooperation, the Price's law was used to select the prominent author group in this study. Then, based on co-authorship analysis of the group, a network was established

to show the cooperation among the prominent authors in the group. Centralities of social network analysis were used as measurement indexes to perform the selection of team leaders and team core members, who were identified on the basis of a relative tie strength analysis. To better assess the results of this identification, we employed the technique of cohesive subgroup analysis to identify core members of teams and compare and discussed the results of the identification obtained by the two different methods. In summary, the research methods we employed mainly include the Price's law, co-author analysis, relative tie strength and social network analysis (SNA). These methods and their related principles are discussed below.
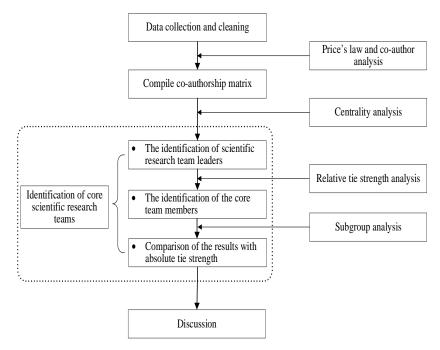

Fig. 1. Research methods and steps

3.1   The Price's law

The number of papers written by an author reflects his or her research ability to some degree. Price noted the great role played by the elite researchers in the development of science. In his book entitled "Little science, Big science" (1963), Price wrote that "the total number of scientists goes up as the square, more or less, of the number of good ones", and proposed that "the nature of the crude inverse-square law of productivity is that if the total production of those who write n papers, it emerges that the large number of low producers account for about as much of the total as the small number of large producers" [20]. Clearly, Price's law, or Price's square root law, refers to that half of all scientific papers are written by a group of prolific authors, and the size of this prolific group is the top square root of the total number of

authors. Hence, if T(a) is the number of total authors who write all scientific papers, the number of prolific authors is n(a), then "$n(\alpha) = \sqrt{T(\alpha)}$". Furthermore, according to the law of Lotka's inverse-square of productivity, an exact formulation of Price's square root law is that if the number of papers written by the low producers in the group of prolific authors is m, then "$m = 0.749 * \sqrt{n(\max)}$", where n(max) is the number of outputs of the largest producers in the group of total authors [21, 22].

In this study, the group of high-productivity authors whose papers account for half of the total is called the "prominent author group". Prominent authors refer to ones who have made key contributions to the academic development of colleges and universities. As mentioned earlier, to observe relatively important researchers and their cooperation, we selected the prominent author group by the Price's law.

3.2 The co-authorship analysis

Bibliometrics is the quantitative study of physical published units as they are reflected in bibliographies [23]. The most common techniques of bibliometric analysis are mainly bibliographic coupling [24], co-citation [25], co-word [26], and co-author [27]. As shown in table 1, the co-word analysis is used to study the conceptual structure and the main concepts treated by a research field. Its analysis unit is usually a keyword or concept but not an author [28]. Bibliographic coupling and co-citation are of citation analysis, which is employed as a measure of similarity between some authors or documents. Co-author analysis as another common bibliometric analysis is usually used to study collaborative relationships. Collaboration in scientific research is reflected by the corresponding co-authorship of published results, and can be analyzed with the help of bibliometric techniques. In line with the focus of this study, we chose co-author analysis as the research methodology of the paper.

Table 1. Most common bibliometric techniques among the units of analysis

Bibliometric technique	Unit of analysis used	Kind of relation
Bibliographic Coupling	• Author	• Common references in authors' oeuvres
	• Document	• Common references in documents
	• Journal	• Common references in journals' oeuvres
Co-citation	• Author	• Co-cited authors
	• Reference	• Co-cited documents
	• Journal	• Co-cited journals
Co-word	• Keyword	• Co-occurrence of terms in a document
	• Abstract	
	• Document's body	
Co-author	• Author	• Co-occurrence of authors in the author list of a document
	• Country from affiliation	• Co-occurrence of countries in the address list of a document
	• Institution from affiliation	• Co-occurrence of institutions in the address list of a document

Authorship is a main bibliometric descriptor of a scientific paper [29]. And author co-occurrence is dominated in aggregation because almost every type of network is composed of authors' academic relationships such as citation or collaboration among authors [30]. Co-authorship (CA) is a typical type of author co-occurrence. We call researchers who appear as authors in one or more same papers as co-authors. The co-author analysis can be described as co-occurrence analysis of authors in the author list of a paper. Newman [14] proposed that co-authorship of a paper can be considered as documenting a collaboration between two or more authors. Co-writing and co-publication of authors are the most visible manifestation of cooperation. The fact that individuals appear as co-authors in a paper represents an easily quantifiable outcome of teamwork. We refer to such a group of co-authors as "scientific research team".

3.3   The relative tie strength analysis

Tie strength is an important concern that has attracted most research attention in social network analysis. It is a quantifiable property of the link between two nodes. Granovetter (1973) indicated that tie strength is a "combination of the amount of time, the emotional intensity, the intimacy (mutual confiding) and reciprocal services which characterize the tie" [31]. There exist correlations between network structure and tie strength, in which strong ties are associated with dense communities, whereas weak links act as bridges between these. As one of collaboration networks, co-authorship network exhibits academic communities. By Granovetter's definition, it can be assumed that strong ties tend to exist in academic communities, while weak ties tend to connect these communities [31, 32]. Numerous measurement methods of tie strength have been proposed, and indications of the "closeness" of a relationship have been used as the most common tactic. Other measures of tie strength such as frequency and duration of contact have also been adopted [33]. This study uses the frequency of author's co-occurrence as a measure of team member's tie strength.

Usually, the higher the co-occurrence frequency of authors in papers is, the stronger the ties among the authors is. Drawing on Petersen's definition of weak, strong, and super tie in scientific research cooperation [17], we used the method of co-author measurement to calculate the average tie strength of cooperation among authors $<K_i>$ as in formula (1).

$$< K_i >= S_i^{-1} \sum_{j=1}^{S_i} K_{ij} \tag{1}$$

Here, S_i represents the co-authorship radius of author I, K_{ij} is the cumulative number of co-authored papers by i and j during a certain period.

If $K_{ij} \geq <K_i>$, there is a strong tie between authors i and j, and if $K_{ij} << K_i>$, there is a weak tie between the two authors.

Furthermore, if $K_{ij} \geq K_i^c$, and if the co-authorship between i and j lasts for 6 years or more during the study period, it is said that there is a super tie of co-authorship between i and j. Here,

$$K_i^c = (< K_i > -1) \ln S_i \tag{2}$$

It is important to note that the definition of tie strength is not symmetrical. In other words, if author j is a strong or super co-author to author i, i is not necessarily a strong or super co-author to author j.

3.4 Social network analysis

A social network is a network composed of social actors and the ties among them, with the actors as nodes or vertices and their ties as lines or edges connecting pairs of nodes. Both actors and ties can be diverse. Social network analysis (SNA) is a socio-logical endeavor aimed at analyzing the relations among social actors and the patterns of these relations so as to explore the underlying social structure [34]. Centrality and cohesive subgroup analysis are major concerns in SNA.

Centrality is determined through a quantitative analysis of the power of an individual or an organization in a network. This includes the measurement of centrality of point and graph. Centrality of point measures the size of power or the degree of status an individual has in the network, while centrality of graph, or centralization, measures the centralizing trend of the whole network. Degree centrality, betweenness centrality, closeness centrality and eigenvector centrality are the fundamental components of centrality. The importance of an author depends on his or her influence in the collaboration network and can be measured by the degree centrality and betweenness centrality. Indeed, degree and betweenness centralities were used to identify team leaders in our study.

Degree centrality is in some sense a measure of the popularity or prestige of an actor as a node, which can be seen through the number of other nodes to which it is connected directly [35, 36]. Generally, the fact that an actor has relatively high "degree" implies that he or she has a popular or prestigious position in the network. Furthermore, an actor with high degree centrality can be identified as the informal leader of the group [37].

Betweenness centrality is simplest and most widely used. It measures the extent to which a node lies on the shortest paths between pairs of other nodes [35, 38]. Betweenness can be calculated as the fraction of shortest paths between node pairs that pass through the node of interest. Obviously, a node with high betweenness plays the role of a broker or gatekeeper to connect other nodes in the network. In other words, betweenness centrality can be determined by measuring the extent to which an actor has control over information flowing between other actors [36, 38, 39]. The higher betweenness centrality an actor has, that is, the more the individuals who have to rely on this actor to connect with each other, the more advantage this actor has to obtain information and the greater "power" to control others [37]. The high betweenness centrality is often used as an indicator of the influence of a node in a group or organization.

Cohesive subgroup analysis are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties [34]. They always represent an important social construct for sociologists who study individuals and organizations [40]. Some scholars use the words "clustering" and "component" to describe the subgroups, and others call them "gangs" or "circles". The common feature of these definitions is

that there is maximum close relationship or degree of aggregation among members in a subgroup. Various techniques for identifying cohesive subgroups are developed and employed, such as c-lique [41], n-clique [42], k-core [43] etc.

A clique is a maximal complete subgroup. It consists of a subgroup of three or more nodes in a network, all of which are adjacent to each other, and there are no other nodes that are also adjacent to all of the nodes of the clique. An n-clique is a maximal subgroup, in which the largest geodesic distance between any two nodes is less than or equal to n. A k-core is a subgroup in which each node is adjacent to at least a minimum number, k, of the other nodes in the subgraph.

In this study, a core scientific research team refers to a secondary research group with its leaders at the core and close collaboration among its members. So, the technique of cohesive subgroup analysis was employed to complete the identification of core members of team.

4 Data Collection and Processing

4.1 Data collection

The goal of this study was to identify core scientific research teams of an institution. The sample data used in this paper were extracted from the source journals of SCI (Science Citation Index), EI (Engineering Index), CSSCI (Chinese Social Sciences Citation Index) and core journals in the CAJD (China Academic Journal Network Publishing Database) of the CNKI (China National Knowledge Infrastructure). These publication data were related to Jiangsu University of Science and Technology (JUST), an important engineering-oriented university in China, with a well-established talent training system of bachelor, master and doctoral degree levels.

To construct our dataset, publications from JUST during the period from 2005 to 2017 were searched. A total of 7,943 papers were found. We exported and recorded relevant data, such as the name of the journal, the year of its publication, and its title and authors. To ensure the reliability of the downloaded data, we checked and validated all the sample data in these papers except for 9 papers, which were not considered as academic publications, including the type corrections, epitomal materials, reports, meeting abstracts, speeches, news items, etc. Having removed these non-academic articles, we had 7,934 papers for our study. These data were analyzed with MATLAB and Excel. There were 1,070 single-author papers and 6,864 co-authored papers, which accounted for 86.51%. The results also showed that 7,882 authors were involved. The average productivity was only about 1.007 during this period. The paper publications of JUST during the specified period are shown in Figures 2 and 3. And the distribution of the number of papers that an author published is shown in Figures 4.
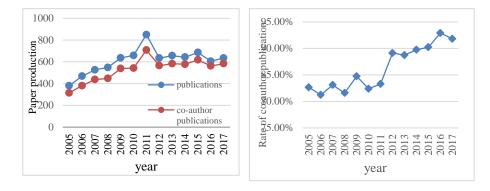
Fig. 2. Paper production and co-author publications
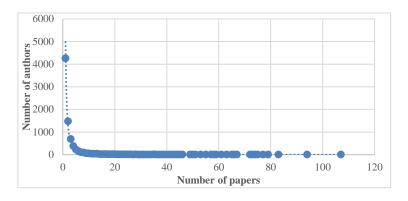

Fig. 3. The rate of co-authored papers

Fig. 4. Distribution for the number of papers that an author published

From Figures 2 and 3, we observed that scientific research was flourishing in JUST and that its main characteristic was cooperation with a rising trend. The rate of cooperative publications increased from 82.68% in 2005 to 91.81% in 2017. Figure 4 shows that the number of papers per author appears an obvious long-tail distribution with a small fraction of authors who had a very large number of papers. It also shows that authors with large number of papers were primarily in the tail of the distribution, indicating that they were the more productive authors. This is a phenomenon of Matthew effect, in which the rich tend to get richer. It is closely related to the concept of preferential attachment in network science, where some authors with the more connected nodes are destined to acquire many more collaborators than others during their scientific career [44]. To identify relatively important researchers and their cooperation, we used the Price's law to select the prominent author group.

4.2   Data processing

We focused on cooperation among authors for the data process with MATLAB and Excel. After cleaning the data, including checking authors who had the same names and removing punctuation symbols or extra spaces, we converted the selected data into a suitable format by some bibliometric tools to facilitate further analysis. Total co-authorship matrix was compiled with Visual Basic for Applications.

According to Price's law, "$n(\alpha)=\sqrt{7,882}=88.78\approx89$", i.e., the most prolific producers were about 89 authors (including non-first authors of papers). Each of these authors has produced 28 papers or more in our study period. Table 2 shows the top 15 most prolific authors. The author with the highest productivity had 107 articles. Furthermore, according to the formulation of prolific author productivity, "$m=0.749*\sqrt{107}\approx7.75$", that is, the lowest number of articles of an author in the prolific author group was about 7.75. In this study, authors who had 7 or more articles were considered prominent authors; there were 708 such authors in total. Among them, 12 were single authors who wrote papers independently. As we were only interested in co-authors, all of the single authors were excluded. As a result, 696 authors who had co-authored papers were included in the prominent author group and numbered according to their productivity. The co-author information of these 696 authors from the authorship data was extracted, and a 696 × 696 co-authorship matrix was compiled. Then, by importing the authorship matrix into UCINET and NetDraw, we visualized co-authorship network of the 696 scientific authors (Figure 5). In the graph, nodes are authors, and lines are co-authorships. If two authors have co-authored one or more papers, they are connected by a line. When drawing the graph, the node size was set in degree by analysis of the centrality measure using the NetDraw tool. Hence, the size of a node in the co-author network represents the number of co-authors he or she has. The larger the node is, the more co-authors he or she has. The size of a line in the network represents the frequency of co-authorship between a pair of authors. The thicker the line is, the stronger co-authorship the pair of authors has. There are single nodes in the diagram because other authors who co-authored with these prominent authors in our study period were not in the prominent author group, and they were not retained in our analysis samples. Therefore, these prominent authors appear as single nodes in the network.

Table 2. Top 15 most productive authors

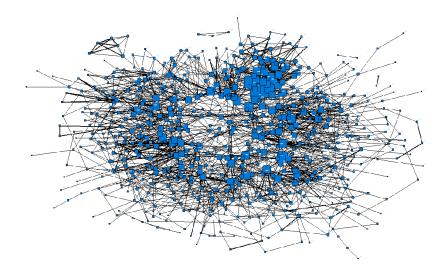| Seq.	Author	No. of papers	Seq.	Author	No. of papers	Seq.	Author	No. of papers
1	ZZY	107	6	WJM	75	11	CFY	67
2	LWT	94	7	GYM	75	12	GXJ	66
3	WJ	83	8	TYB	74	13	ZJ	65
4	TWX	79	9	TYB	73	14	WZL	65
5	GSL	77	10	ZB	72	15	GS	63

Fig. 5. Co-authorship network of 696 scientific authors. Nodes in the network represent authors, and a line between two of them indicates they co-authored one or more papers during the period of our study. When drawing the graph, we set node size in degree by analysis of the centrality measure using the NetDraw tool. The size of a node in the network represents the number of co-authors he or she has, and the size of a line represents the frequency of co-authorship between a pair of authors.

The crucial features of a social network [45] are:

- Short distance (shown in the "small-world effect")
- Skewed degree distributions (a result of "preferential attachment")
- High clustering.

Network analysis in UCINET revealed that the density of the co-authorship network is 0.0317 and that the size of the giant component is 654, accounting for 94% of the total number of nodes of the network. This indicates that the network is very sparse, but information flow among the prominent authors is smooth. A further analysis showed that the average distance of the network is 4.58, that the maximum distance is 13, and that the external–internal (E-I) index is −0.969. It is clear that the co-authorship network consisting of 696 prominent authors has the small-world effect and high clustering.

5   Results

Teams can be divided into core teams and extended teams [46]. A core team consists of its leader and his or her main co-cooperators. An extended team is formed by adding new members to the core team. Hence, a research team is mainly composed of team leaders, team core members, and non-core members. Studies show that although larger teams are more likely to succeed, the workload is highly focused across the

team, with most work actually performed by a few of the team members or even just a single person [47]. This means that most members of the team are "hitchhikers" and are in fact "pseudo-members" of the team. Therefore, this study mainly focused on the identification of team leaders and the set of core members who undertake and complete most work, thereby identifying core scientific research teams.

5.1   The selection of team leaders

Team leaders are researchers who have a high academic level in their research fields. They can obtain superior resources from outside of the team and have a core position in the team. They lead team members to conduct academic research and achieve scientific goals. As Freeman et.al. (1979) stated, centrality implies a theory of how centrality might effectively influence team processes, and perceived leadership depends on centrality [36]. Based on the implication of centrality of SNA, our study used degree centrality and betweenness centrality as measurement indexes for the identification of team leaders. With the help of Ucinet, we showed the centrality features of the scientific cooperation network of 696 authors in Table 3. It shows that the network centralization is low, with its degree centrality being only 0.55% and its betweenness centrality being only 5.92%. This indicates that many authors do not collaborate at all, but information transmission throughout the network remains fluent and does not rely on individual nodes.

Table 3. Centrality characteristics of co-authorship network of 696 scientific authors

Type of Centrality	Mean of Centrality	Standard Deviation of Centrality	Maximum Value of Centrality	Centralization
Degree centrality (absolute value)	22.052	26.539	244	0.55%
Betweenness centrality (relative value)	0.456	0.793	6.371	5.92%

Authors with high point centrality are in the center of the scientific co-authorship network, and therefore, they are more likely to be team leaders. In this study, the authors were selected as potential team leaders based on following criteria: their betweenness point centrality is 4 times higher than the mean value, and their degree point centrality is also 4 times higher than the mean value—that is, the relative betweenness point centrality is above 1.824, and absolute degree point centrality is more than 88.208. Through manual screening of the measurement data, 5 of 696 authors were found to meet these criteria and were identified as the team leaders. The centrality analysis of these 5 authors is shown in Table 4.

Table 4. Research team leaders (from large to small in betweenness centrality)

No.	Team Leader	Betweenness Centrality	Degree Centrality
1	GSL	6.371	121
2	WJ	4.929	142
3	GXJ	4.322	244

4	TYB	3.844	125
5	LWT	3.054	104

5.2 The identification of the core team members

The node at the center of the network is called the "peak point," and "clustering" is a set of nodes that are directly linked to the peak point [18]. The purpose of the identification of core teams in scientific research, based on the relative strength of the co-author relationship, is to find collections of core members who have strong ties with each team leader. There are two steps to achieve this. The first is to find the circle of scientific research co-authorship of the five team leaders by using the clique technique of SNA and the "snowball method" proposed by Alba [45]. If two-thirds of members in a smaller circle overlap member of another larger circle, the smaller circle is merged into the larger circle. The second step is to calculate the number of co-authored papers (K_{ij}) between each member and the team leaders of each circle, and then to calculate the average tie strength of co-authorship, $< K_i >$, according to Formula (1). If $K_{ij} < < K_i >$, the member j has a strong tie with team leader i and therefore can be identified as a core member of the team. In this way, five core scientific research teams are finally obtained (as shown in Table 5, with authors listed by their initials). These core teams vary in size from 5 to 19 authors. The average tie strength of co-authorship with the five leaders in each team varies from 4 to 7 papers. Additionally, each core team has at least one core member who has a super tie with the team leader, according to the definition of "super tie" used in this study.

Further analysis of the research topic of the original data (as shown in Table 5) indicated that the five teams are in different disciplines. Team 1 and team 2 have the same research field of management, with team 1 focusing on management information systems, and team 2 on knowledge management. The other three teams have different subject areas, with team 3 being in the field of life science, team 4 in the field of environmental science, and team 5 in the field of electrical engineering.

Table 5. Core members and sizes of core scientific research teams

No.	Team Leaders	Size of Circle	$< K_i >$	Team Core Members	Members with Super Tie	Number of Core Members	Size of Core Team	Subject
1	GSL	29	4.3	WNX, RN, WZY, SX, ZH, YJ, PYH, ML, MH, YJJ, LFA, WPB	WNX, RN	12	13	Management information systems
2	WJ	21	7.0	SQF, SYX, ZYH, WJG, YJJ, ZQ, LP, WXJ	SQF, SYX	8	9	Knowledge management
3	GXJ	42	5.9	SXJ, TGX, XAY, ZQL, QHY, LMW, TSM, GK, QZY, GT, ZYH,	SXJ, TGX	18	19	Life science

			WSP, HCX, LT, ZGZ, SPJ, WPC, GQH				
4	TYB	22	5.9	CFY, LXW, WXG, LLL, ML, ZRJ, LXQ	CFY	7	8	Environ-mental sci-ence
5	LWT	21	5.2	ZZY, ZB, ZXB, DXQ	ZZY	4	5	Electrical engineering

5.3 A comparison with another method based on the absolute threshold

In this study, we proposed a new method for identification of core team members based on the relative tie strength; we demonstrated advantages of our method as compared with the traditional method by the absolute threshold as the evaluation index of core team members.

The identification of core team members based on the absolute threshold mainly involves the cohesive subgroup research techniques of SNA, such as c-level clique, n-clique, and k-core. These techniques have different focuses. c-level clique focuses on the tie strength between a pair of members in a subgroup and requires that the tie strength among all of the pairs of members be less than the value of c. n-clique, based on the geodesic distance between subgroup members, focuses on the length of the shortest path between any pairs of nodes. It requires that the largest geodesic distance between any subgroup members be no greater than n. The k-core method focuses on the number of adjacent nodes that all of the subgroup members have. A k-core is a measure of the activity similarity among nodes, by which all of the members who have more than k adjacencies are clustered in a subgroup. Considering the focus of this study, we chose to use the c-level clique and snowball method to identify the core members of the five team leaders in 3-level clique, 4-level clique, 5-level clique, 6-level clique, and 7-level clique. Results of this identification are shown in Table 6, where the identification of each of five teams whose core team members are identified based on the relative co-authorship strength is also presented for comparison.

The c-level clique method based on the absolute threshold seeks to identify the core members of the team, using the absolute co-authorship strength among authors as an index for the identification of core members. It requires that the co-authorship strength among all of the members be a value of c or greater. In Table 6, the larger the threshold value of c is, the smaller the number of core members of the team is. When the threshold value of c is changed from 3 to 7, the number of core members identified in team 5, which is the smallest size; while the number of core members identified in the other four teams has undergone different degrees of reduction. Table 6 also illustrates that by increasing the threshold value of c gradually, the cuts of core members in some teams are slow, while other teams experience more severe cuts of core members.

When the value of 3 is used as the absolute threshold of identification, there are 10 core members identified in team 1, 10 core members in team 2, 21 core members in team 3, 6 core members in team 4, and 3 members in team 5. Interestingly, the thresh-

old value of 3 is lower than the values of the average tie strength of the co-authorship relation with the team leader in each of the five teams. In this case, 2 core members (114, 149) in team 2 and 3 core members (83, 355, 236) in team 3 failed to be identified by relative tie strength. However, in each of the other three teams, the number of core members is higher than that identified by absolute threshold. Of note, the identification results based on the absolute threshold are all contained in the results based on relative tie strength.

When the value of 5 is chosen as the absolute threshold of identification, team 1 has only 2 core members, team 2 has 7 core members, team 3 has 17 core members, team 4 has 5 core members, and team 5 has 3 core members. We also note that the value of 5 is the minimum value of the average tie strength of the co-authorship relation of the five team leaders. In this case, not only is the number of core members based on the relative tie strength higher than that identified by the absolute threshold method in each of the five teams, but all of the core members identified by the latter are also contained in the results identified by the former.

In addition, Table 5 shows that the average tie strength of the co-authorship relation for the leader of team 1 is 4.3 papers. In other words, the value of 5 is a cutoff value for the leader of team 1 to seek core team members from his research circles. By using this cutoff value as the index of the identification by the absolute threshold method, only member 64 (WNX) and member 465 (YJ) remain. From Table 6, we know that member 68 (RN) disappears when the value of C is increased from 4 to 5, and yet this member has co-authored 11 papers with the team leader, and their co-authorship relation has lasted for at least 13 years, covering our sample study period. According to the definition of a super tie and Formula (2), member 68 (RN) is a super-tie co-author of the team leader and should not be ignored, yet does end up being ignored when the absolute threshold method is used.

When the absolute threshold value of c goes up to 7, team 1 has no core members; in other words, team 1 cannot be identified. Meanwhile, team 2 has 4 core members, team 3 has 14 core members, team 4 has 4 core members, and team 5 has 3 core members. The results also showed that the value of 7 is the maximum value of the average tie strength of the co-authorship relation of the five team leaders. In this case, it can be easily found that, apart from the failure of identification of team 1, all core members of each of the other four teams identified by absolute threshold are clearly included in each result based on the relative tie strength.

In addition, Table 5 shows that the value of 7 is a cutoff value for the leader of team 2 to look for her core team members from her research circles. With this absolute threshold value, only four members (51, 87, 308, and 102) are obtained. These identification results only account for half of the results based on the relative tie strength.

Table 6. Identification results of five core scientific research teams

Team No.	Subject	Relative tie strength	Absolute threshold				
			3-level clique	4-level clique	5-level clique	6-level clique	7-level clique
1	Management Information Systems						NULL
2	Knowledge Management						
3	Life Science						
4	Environmental Science						
5	Electrical Engineering						

6 Discussion

6.1 Findings

First, there is an obvious uneven distribution of co-authorship in the network of teamwork. As a result of the "preferential attachment effect", "apostle effect", and "star effect", team members are more likely to co-author with the team leader than with other members of the team. From the teamwork network of Table 6, it shows that team members do not have co-authorships with all of the other members of the team. In team 1, for example, apart from cooperating with the team leader, member 101 had a co-authorship only with member 114. Meanwhile, member 68 had co-authorships with most members of the team, but not with members 101, 114, and 276. Furthermore, the bibliometric analysis of co-authored papers showed that member 68's co-authorship strength with the team leader was 11 papers, 5 papers each with members 71 and 309, and 1 to 4 papers with the other members. This shows that there is a

strong tie between team members and team leaders, but that such a tie is not always strong or existent between pairs of members.

Moreover, the traditional absolute threshold method can lead to the inadequate identification of core members in a scientific research team. As this method of identification does not take into consideration of the differences in the tie strengths among members within each team, a poorly selected absolute threshold value of c can result in a situation where some core members are ignored due to the mandatory requirements of the absolute tie strength.

Second, scientific research teams vary in their means of collaboration in the co-authorship network. Table 6 shows that each of the five teams has its own unique methods of co-authorship. From the perspective of co-authorship pattern, in teams 2, 3, 4 and 5, there are connections among core team members, whether they are direct or indirect, and all core members of each team form the maximal complete subgroup of their core team with a multi-node co-authorship pattern. In team 1, however, there exist different co-authorship patterns among core members, including single-node, dual-core, triangular, and multi-node co-authorship modes, which form a "star" co-authorship mode in which all of the members center on the leader of the whole team. In terms of the tie strength, the analysis shows that the relative tie strength of team members differs from team to team. As far as the tie strength with team leaders is concerned, the mean value in the five teams ranges from 4 to 7 papers. We also find that the closeness and modes of co-authorship among team members are different in teams in different disciplines, and can even differ in different teams within the same discipline.

Therefore, the practice of directly using an absolute threshold of tie strength as an evaluation index for the identification of core members of all teams—not just a specific team—can lead to a situation where some core members of a team and even some core teams in certain disciplines are ignored, as the differences in team co-authorship method patterns are neglected.

6.2 Limitations

While looking into the problem from a new perspective and introducing the relative strength of cooperative relationship in the identification of core scientific research team, allowing for eliminating the shortcomings of the traditional scientific-research-network-based absolute threshold method for the identification of core team members, we also find some limitations of our study.

First, in choosing sample data, we followed the practice of most scholars by focusing our study of scientific research cooperation relationships on the act of co-authoring papers. Scientific research cooperation is reflected not only in paper co-authoring, but also in activities like scientific book co-writing and patent co-inventing. Therefore, focusing only on co-authorship may lead to limited results for the core scientific research team identification.

Second, in recent years, with the rapid development of discipline internationalization, the publication of papers in international academic journals has gradually increased. However, we are unable to include such papers into our sample data analysis.

Although such papers are much less than papers in domestic journals, these data in our analysis could affect the outcome of the identification of core members of scientific research teams.

Third, in constructing and analyzing the network, we still follow the practice of most scholars by building a static network through a period of the database. In real academic life, researchers frequently face the decision to start or end their research cooperation. This means that over time, core team members change, and their numbers evolve. The stability of scientific research cooperation is reflected not only in the strength of the tie, but also in the length of time the tie lasts. Future studies for scientific research team will include further investigations of changes in collaboration.

7 Conclusion

In this study, we have identified and evaluated core scientific research teams of a university during a certain period by using a new method based on the relative strength of cooperative relationship. As compared to a traditional method by using absolute threshold, we have demonstrated significant advantages of our method. Based on the data obtained from the four major sources of CNKI, targeting the scientific research cooperation in the form of co-authoring papers and using the method of social network analysis and bibliometric analysis, this study concludes as follows.

First, we find that under "the preferential attachment effect", "the apostle effect", and "the star effect", the cooperative relationships among scientific researchers are not evenly distributed. Moreover, teams in different disciplines and even different teams in the same discipline have different patterns of cooperation.

Second, the use of an absolute threshold as the index for the identification of core members of different teams can easily lead to inadequate identification of core members of some teams if the value is not properly chosen. In some cases, the absolute threshold is too large, and this "uniform" evaluation criterion of cooperative relationship strength results in the elimination of some core teams in some disciplines.

Third, considering the heterogeneity of the team, use of the relative strength of the cooperative relationship as the evaluation criterion for identifying scientific research teams and their core members can effectively avoid the risk of some core members being erased due to the mandatory requirements of absolute tie strength. This can also make up for the insufficient identification of scientific research core teams in some disciplines as a result of the falsely equal treatment of the absolute threshold.

Networks grow by the addition of connections and a newly cooperative research results will contribute new connections to an individual [48]. In the future, we will seek to use more comprehensive scientific research results as data for further testing the research method proposed in this paper. At the same time, we will follow the formation and development of a core scientific research team from the perspective of dynamic evolution, which may be helpful for colleges and universities, research institutes, and other scientific organizations establishing and managing their scientific research teams.

8 Acknowledgements

9 References

[1] O. Persson, W. Glänzel, and R. Danell, "Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies," Scientometrics, vol. 60, no. 3, pp. 421-432, 2004. https://doi.org/10.1023/b:scie.0000034384.35498.7d

[2] S. Milojević, "Modes of collaboration in modern science: Beyond power laws and preferential attachment," Journal of the Association for Information Science and Technology, vol. 61, no. 7, pp. 1410-1423, 2010. https://doi.org/10.1002/asi.21331

[3] X. Zhi, C. Liyu, and W. Sihui, "A study of the influence factors of university scientific research team cooperation," Science Research Management, vol. 36, no. 5, pp. 149-161, 2015.

[4] P. Van Den Besselaar, S. Hemlin, and I. Van Der Weijden, "Collaboration and competition in research," Higher Education Policy, vol. 25, no. 3, pp. 263-266, 2012. https://doi.org/10.1057/hep.2012.16

[5] Z. Pengcheng and P. Han, "On the relationship between scientific cooperative networks' characteristics and teams' knowledge creation," Science Research Management, vol. 32, no. 7, pp. 104-112, 2011.

[6] C. Wang, Z. Cheng, and Z. Huang, "Analysis on the co-authoring in the field of management in China: Based on social network analysis," International Journal of Emerging Technologies in Learning, vol. 12, no. 6, pp. 149-160, 2017.

[7] P. Balkundi, M. Kilduff, and D. A. Harrison, "Centrality and charisma: Comparing how leader networks and attributions affect team performance," Journal of Applied Psychology, vol. 96, no. 6, pp. 1209-1222, 2011. https://doi.org/10.1037/a0024890

[8] R. T. Sparrowe, R. C. Liden, S. J. Wayne, and M. L. Kraimer, "Social networks and the performance of individuals and groups," Academy of Management Journal, vol. 44, no. 2, pp. 316-325, 2001. https://doi.org/10.2307/3069458

[9] V. Venkataramani, A. W. Richter, and R. Clarke, "Creative benefits from well-connected leaders: Leader social network ties as facilitators of employee radical creativity," Journal of Applied Psychology, vol. 99, no. 5, pp. 966-975, 2014. https://doi.org/10.1037/a0037088

[10] J. C. Long, F. C. Cunningham, P. Carswell, and J. Braithwaite, "Who are the key players in a new translational research network?," BMC Health Services Research, vol. 13, no. 1, pp. 338-349, 2013. https://doi.org/10.1186/1472-6963-13-338

[11] C. Calero, R. Buter, C. Cabello Valdés, and E. Noyons, "How to identify research groups using publication analysis: An example in the field of nanotechnology," Scientometrics, vol. 66, no. 2, pp. 365-376, 2006. https://doi.org/10.1007/s11192-006-0026-z

[12] Y. Liu, S.-M. Tang, and X. Cai, "Structural analysis of co-authorship networks of R&D groups based on weight factor—a case study of a local technological university's material school," R&D Management, vol. 27 no. 2, pp. 129-137, 2015.

[13] C. Reffay and T. Chanier, "How social network analysis can help to measure cohesion in collaborative distance-learning," in Designing for Change in Networked Learning Environments: Springer, 2003, pp. 343-352. https://doi.org/10.1007/978-94-017-0195-2_42

[14] M. E. Newman, "Coauthorship networks and patterns of scientific collaboration," Proceedings of the National Academy of Sciences, vol. 101, no. suppl 1, pp. 5200-5205, 2004. https://doi.org/10.1073/pnas.0307545100

[15] M. E. Newman, "The structure of scientific collaboration networks," Proceedings of the National Academy of Sciences, vol. 98, no. 2, pp. 404-409, 2001.

[16] A. Abbasi, L. Hossain, and L. Leydesdorff, "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks," Journal of Informetrics, vol. 6, no. 3, pp. 403-412, 2012. https://doi.org/10.1016/j.joi.2012.01.002

[17] A. M. Petersen, "Quantifying the impact of weak, strong, and super ties in scientific careers," Proceedings of the National Academy of Sciences, vol. 112, no. 34, pp. E4671-E4680, 2015. https://doi.org/10.1073/pnas.1501444112

[18] J. Scott, Social network analysis. Sage, 2017.

[19] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, no. 5439, pp. 509-512, 1999. https://doi.org/10.1126/science.286.5439.509

[20] D. J. de Solla Price, Little science, big science. Columbia University Press New York, 1963.

[21] P. T. Nicholls, "Price's square root law: empirical validity and relation to Lotka's Law," Information Processing & Management, vol. 24, no. 4, pp. 469-477, 1988. https://doi.org/10.1016/0306-4573(88)90049-0

[22] W. Glänzel and A. Schubert, "Price distribution. An exact formulation of Price's "square root law"," Scientometrics, vol. 7, no. 3-6, pp. 211-219, 1985. https://doi.org/10.1007/bf02017147

[23] W. Hood and C. Wilson, "The literature of bibliometrics, scientometrics, and informetrics," Scientometrics, vol. 52, no. 2, pp. 291-314, 2001.

[24] M. M. Kessler, "Bibliographic coupling between scientific papers," American Documentation, vol. 14, no. 1, pp. 10-25, 1963.

[25] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," Journal of the American Society for Information Science, vol. 24, no. 4, pp. 265-269, 1973. https://doi.org/10.1002/asi.4630240406

[26] M. Callon, J.-P. Courtial, W. A. Turner, and S. Bauin, "From translations to problematic networks: An introduction to co-word analysis," Information (International Social Science Council), vol. 22, no. 2, pp. 191-235, 1983. https://doi.org/10.1177/053901883022002003

[27] W. Glänzel, "National characteristics in international scientific co-authorship relations," Scientometrics, vol. 51, no. 1, pp. 69-115, 2001.

[28] M. Aria and C. Cuccurullo, "Bibliometrix: An R-tool for comprehensive science mapping analysis," Journal of Informetrics, vol. 11, no. 4, pp. 959-975, 2017. https://doi.org/10.1016/j.joi.2017.08.007

[29] W. Glänzel, "Coauthorship patterns and trends in the sciences (1980-1998): A bibliometric study with implications for database indexing and search strategies," Library Trends, vol. 50, no. 3, pp. 461-473, 2002.

[30] J.-P. Qiu, K. Dong, and H.-Q. Yu, "Comparative study on structure and correlation among author co-occurrence networks in bibliometrics," Scientometrics, vol. 101, no. 2, pp. 1345-1360, 2014. https://doi.org/10.1007/s11192-014-1315-6

[31] M. S. Granovetter, "The strength of weak ties," Social Networks, vol 78, no. 6, pp. 347-367, 1977. https://doi.org/10.1016/b978-0-12-442450-0.50025-0

[32] M. Girvan and M. E. Newman, "Community structure in social and biological networks," Proceedings of the National Academy of Sciences, vol. 99, no. 12, pp. 7821-7826, 2002. https://doi.org/10.1073/pnas.122653799

[33] P. V. Marsden and K. E. Campbell, "Measuring tie strength," Social Forces, vol. 63, no. 2, pp. 482-501, 1984. https://doi.org/10.1093/sf/63.2.482

[34] S. Wasserman and K. Faust, Social network analysis: Methods and applications. Cambridge university press, 1994.

[35] L. C. Freeman, "Centrality in social networks conceptual clarification," Social Networks, vol. 1, no. 3, pp. 215-239, 1978. https://doi.org/10.1016/0378-8733(78)90021-7

[36] L. C. Freeman, D. Roeder, and R. R. Mulholland, "Centrality in social networks: II. Experimental results," Social Networks, vol. 2, no. 2, pp. 119-141, 1979. https://doi.org/10.1016/0378-8733(79)90002-9

[37] D. Krackhardt, N. Nohria, and B. Eccles, "The strength of strong ties," Networks in the Knowledge Economy, v 82, 2003.

[38] M. E. J. Newman, "A measure of betweenness centrality based on random walks," Social Networks, vol. 27, no. 1, pp. 39-54, 2005. https://doi.org/10.1016/j.socnet.2004.11.009

[39] R. S. Burt, Structural holes: The social structure of competition. Harvard university press, 2009.

[40] K. A. Frank, "Identifying cohesive subgroups," Social Networks, vol. 17, no. 1, pp. 27-56, 1995.

[41] R. D. Luce and A. D. Perry, "A method of matrix analysis of group structure," Psychometrika, vol. 14, no. 2, pp. 95-116, 1949.

[42] R. D. Alba, "A graph-theoretic definition of a sociometric clique," Journal of Mathematical Sociology, vol. 3, no. 1, pp. 113-126, 1973. https://doi.org/10.1080/0022250x.1973.9989826

[43] S. B. Seidman and B. L. Foster, "A graph-theoretic generalization of the clique concept," Journal of Mathematical sociology, vol. 6, no. 1, pp. 139-154, 1978. https://doi.org/10.1080/0022250x.1978.9989883

[44] M. Perc, "The Matthew effect in empirical data," Journal of the Royal Society Interface, vol. 11, no. 98, p. 1-15, 2014.

[45] M. E. Newman, "The structure and function of networks," Computer Physics Communications, vol. 147, no. 1-2, pp. 40-45, 2002.

[46] S. Milojević, "Principles of scientific research team formation and evolution," Proceedings of the National Academy of Sciences, vol. 111, no. 11, pp. 3984-3989, 2014. https://doi.org/10.1073/pnas.1309723111

[47] M. Klug and J. P. Bagrow, "Understanding the group dynamics and success of teams," Royal Society Open Science, vol. 3, no. 4, pp. 1-11, 2016.

[48] D. d. S. Price, "A general theory of bibliometric and other cumulative advantage processes," Journal of the American Society for Information Science, vol. 27, no. 5, pp. 292-306, 1976. https://doi.org/10.1002/asi.4630270505

10   Authors

Meishu Zhang is a Ph.D. student of Management Information Systems in the School of Economics and Management, Jiangsu University of Science and Technology, Jiangsu, China.

Yu Jia is a Ph.D. student of Management Information Systems in the School of Economics and Management, Jiangsu University of Science and Technology, Jiangsu, China.

Nianxin Wang is an Associate Professor of Management Information Systems in the school of Economics and Management, Jiangsu University of Science and Technology, Jiangsu, China.

Shilun Ge is a Chair Professor of Management Information Systems in the School of Economics and Management, Jiangsu University of Science and Technology, Jiangsu, China. Uncategorized References