Mining Core Teams from Directed Co-authorship Networks: A Case Study in Library and Information Science (Postprint)

Authors: Gao Nan, Zhou Qingshan

Date: 2023-04-01T16:02:57+00:00

Abstract

[Purpose/Significance] To facilitate institutional talent acquisition, this study constructs a standardized process for identifying domain core teams, conducting exploratory research from four aspects: determination of core-region authors, discrimination of team guidance/mentorship relationships, team publication preferences, and team research directions. [Method/Process] First, preliminary data standardization is performed; subsequently, author co-authorship undirected binary matrices are combined with core-periphery structure analysis to partition core-region authors, who are then appropriately adjusted based on multi-dimensional indicators; finally, domain core research teams are identified through the directed co-authorship network between first authors and other authors. [Results/Conclusion] By clarifying the fundamental process for discriminating team guidance/mentorship relationships, general patterns are distilled, including: In co-authorship teams with an "outward radiating star" topology, nodes in central positions exhibit a high probability of being mentors/supervisors, whereas in co-authorship teams with an "inward cohesive star" topology, authors at the network center possess stronger research capabilities;

In directed co-authorship networks between first authors and other authors, if two nodes simultaneously share first-author and corresponding-author relationships, have identical affiliations, and exhibit a certain gap in academic age, then a guidance relationship definitively exists between these nodes, with a high probability of a mentorship relationship.

Full Text

Core Team Mining Research Based on Directed Co-authorship Networks: A Case Study of the Library and Information Science Field

Gao Nan[1,2], **Zhou Qingshan**[2] [1]Institute of Scientific and Technical Information of China, Beijing 100038 [2]Department of Information Management, Peking University, Beijing 100871

Abstract

[**Purpose/Significance**] To facilitate institutional talent recruitment and establish a standardized process for identifying domain core teams, this study conducts mining research on domain core teams from four aspects: core area author identification, team guidance/mentorship relationship discrimination, team publication tendency, and team research direction. [**Method/Process**] The process involves: first, conducting preliminary data standardization; then, combining an undirected binary co-authorship matrix with core-periphery structure analysis to identify core area authors, with appropriate adjustments made based on multi-dimensional indicators; finally, identifying domain core research teams based on a directed co-authorship network between first authors and other authors. [**Result/Conclusion**] In clarifying the basic process for judging team guidance/mentorship relationships, generalizable patterns are distilled: (1) In co-authoring teams with an "outward radiating star" topology, the central node has a high probability of being a mentor/supervisor, while in "cohesive star" topology co-authoring teams, authors at the network center demonstrate strong research capabilities; (2) In a directed co-authorship network between first authors and other authors, if two nodes simultaneously possess the relationship of first author and corresponding author, share the same affiliated institution, and have a certain gap in academic age, then there must be a guidance relationship between these two nodes, with a high probability of a mentorship relationship.

Keywords: social network analysis; directed co-authorship network; co-authorship; core-periphery structure

---

1 Introduction

Core teams constitute the fundamental units for conducting scientific and technological research, formed by groups of researchers with complementary advantages and close connections who collaborate and share responsibilities for common research goals or tasks [1]. In the era of big science, with increasingly deep interdisciplinary integration and continuously rising complexity and comprehensiveness of research content, many scientific studies can no longer be completed by individuals alone but require core teams composed of members with complementary advantages and diverse knowledge structures. Therefore, core teams have become the backbone driving academic innovation research. Mining closely collaborating core teams helps promote institutional research management and holds significant importance for talent evaluation system construction and high-tech talent recruitment.

Core team identification methods can be broadly categorized into four types: traditional core team identification methods, association rule-based methods, hierarchical clustering-based methods, and social network analysis (SNA)-based methods [2]. Among these, SNA is the most widely applied approach. Grounded in mathematics and graph theory, SNA employs metrics such as centrality and density, along with methods including cohesive subgroup analysis, core-periphery structure analysis, and network topology analysis, to measure and visualize relationships among bibliometric entities [3]. This method typically constructs scientific collaboration networks based on author co-authorship relationships and citation relationships among documents, adjusts cooperation thresholds to obtain more closely collaborating groups, and uses network node characteristic indicators (such as betweenness centrality and closeness centrality) to assist in identifying important authors in the network, thereby enhancing the credibility of team identification results through combined manual interpretation.

SNA has been extensively applied in research on social entity relationships. Qiu Junping et al. [4] identified core authors in the field of informetrics based on Price's Law, used n-clique indices to filter collaborative groups, and combined centrality metrics to assess author importance. Li Gang et al. [5] discovered core team members, research themes, and inter-team relationships based on co-authorship networks, co-word networks, and author keyword coupling networks, using cohesive subgroup analysis and snowball sampling methods, and employed cooperation network centrality indicators to identify academic leaders in core teams [6]. Yu Yongsheng et al. [7] proposed an iterative betweenness centrality ranking method for team leader identification, combined with 2-clique methods and snowball sampling to identify other core team members. Zhao Yuxiang et al. [8] employed SNA to empirically analyze actor collaboration characteristics in open data competitions from three dimensions: centrality, clustering, and small-world effects. Pang Hongqian et al. [9] selected the WISE Lab of Dalian University of Technology as a research team sample and analyzed research team collaboration closeness through constructing multi-valued matrices and cooperation network graphs combined with SNA methods. Shen Gengyu et al. [10] measured author cooperation relationships by calculating inter-author similarity through vector space models, then analyzed author cooperation networks using cohesive subgroup analysis in SNA. G. Gonzalez-Alcaide et al. [11] analyzed author co-authorship of Chagas disease literature in the Medline database from 1940-2009, identifying 148 research teams composed of 1,750 authors using SNA. X. Liu et al. [12] constructed author co-authorship networks through binary undirected network models, defined AuthorRanks as influence metrics for individual authors in the network, and conducted empirical analysis in the digital library field. T. Krichel et al. [13] constructed both binary and weighted co-authorship networks to identify author collaboration teams in the RePEc economics database. X. Wang [14] and J. Kang [15] respectively applied SAO (Subject-Action-Object) and LDA (latent Dirichlet allocation) models combined with SNA for identifying potential collaboration partners.

However, these studies have focused on undirected network analysis, overlooking the effectiveness of directed networks in revealing implicit information in author co-authorship networks, and have excessively relied on network centrality indicators while neglecting analysis of important information such as authors' affiliated institutions and author order. Therefore, this study employs directed co-authorship networks to replace traditional undirected co-authorship networks. Additionally, to highlight the importance of first authors and corresponding authors, we construct directed co-authorship networks between first authors and other authors, and between first authors and corresponding authors for domain core team identification and intra-team guidance/mentorship relationship judgment. Furthermore, combined with two-mode networks and keyword co-occurrence networks, we analyze core teams' publication tendencies and research directions.

2 Data Sources and Preprocessing

2.1 Institution Normalization and Author Deduplication

Data were retrieved from Web of Science on November 2, 2020, with the subject set to Library & Information Science (LIS), time span of 2016-2020, and index databases including SCI-EXPANDED, SSCI, CPCI-S, and CPCI-SSH, yielding 12,098 documents.

The data processing consists of four components: data cleaning, core area author identification, guidance/mentorship relationship judgment, and team research tendency analysis. The overall framework is shown in Figure 1 [Figure 1: see original paper].

(1) Institution Normalization. Due to inconsistent expression of institution names (full names, abbreviations, colloquial names, etc.), the same institution may have multiple name variants. Historical evolution of institutions also leads to coexistence of old and new names. Additionally, hierarchical relationships exist among institutions, necessitating institution name standardization. We employed Web of Science's institution expansion field for preliminary institution name standardization, combined with the "Institution Name Standardization Database" self-built by the Institute of Scientific and Technical Information of China for secondary standardization.

(2) Author Deduplication. Similar to institution names, author names also have multiple expression forms (full names, abbreviations, etc.). We used the AF field (author full name) provided by Web of Science as the standardized expression of author names. The name disambiguation problem among different authors is a major issue affecting accurate calculation of author output. In the downloaded dataset, we found that the RI (ResearcherID) and OI (ORCID identifier) fields provided by Web of Science had substantial data missing. Moreover, since ORCID can be registered for free by individuals who can provide different email addresses and apply for multiple ORCID IDs, its accuracy is limited. Therefore, we employed tagging combined with author co-authorship relationships for author name disambiguation. The specific process includes: (1) Using "author name + primary institution + country" as a unique identifier for preliminary statistics, yielding 31,121 authors; (2) Merging authors with inconsistent primary institutions due to affiliation with multiple institutions. As shown in Table 1 , comparing the institution names affiliated with M. Chen in two papers, we determined these papers belong to the same author and performed author merging, assigning unique identifiers using "UT number + author order."

Based on studies by Zhou Ping and L. Leydesdorff et al. [16], on the basis of steps (1) and (2), we merged authors with the same name and shortest collaboration distance of 2. As shown in Figure 2 [Figure 2: see original paper], in papers published by K. Bystrom, two papers both collaborated with the same person E. E. Isah, therefore we determined that K. Bystrom in these two papers is the same person. After merging homonymous authors, we obtained the K. Bystrom co-authorship network shown in Figure 2.

After author name disambiguation, we obtained 27,221 authors, discovering and processing 3,900 homonymous authors, with significant effect.

2.2 Related Indicator Calculations

(1) Network Node Characteristic Indicators: Degree centrality reflects a node's ability to develop relationships with other nodes; Closeness centrality—higher values indicate the node is closer to the network core, reflecting the node's control over network resources [17]; Betweenness centrality indicates the extent to which a node lies on paths between other nodes, acting as an "intermediary"; Eigenvector centrality is based on the idea that connections to high-scoring nodes contribute more influence than connections to low-scoring nodes [18].

(2) Price's Core Author Formula is as follows:

$$P \geq K(p) = 0.749\sqrt{\mathrm{Max}(p)} \quad \text{(Formula 1)}$$

In Formula 1, P represents publication count, and K(p) represents the minimum publication count for inclusion as a core author candidate. When an author's publication count is not less than the minimum value K(p), the author can be included as a core author candidate [19].

(3) Hm Index Calculation Formula is as follows:

$$H_m = h + \frac{N_c - h^2}{2h+1} \quad \text{(Formula 2)}$$

In Formula 2, Nc represents the author's total citation count. Since $\frac{N_c - h^2}{2h+1} \leq 2$, we have $h \leq h_m \leq 2h$. When a scholar's $N_c \to +\infty$, then $h_m \to h$. Therefore, the greater the scholar's academic influence (i.e., the higher the total citation frequency), the smaller the difference between h_m and h [20].

(4) Academic Age. Refers to the duration from the publication year of an author's first paper to the present, representing how long the author has been conducting scientific research. In a collaboration team, scholars with older academic ages generally provide some academic guidance to scholars with younger academic ages [21].

3 Guidance/Mentorship Relationship Judgment

3.1 Core Area Author Identification

Since this study aims to mine domain core teams and authors within the same team necessarily have co-authorship relationships, single-author publications are excluded from the scope of this study. Based on collaboration relationships among 27,221 authors, matrix construction and visualization would involve excessive data volume with poor results. Therefore, weak collaboration relationships where authors only collaborated once are eliminated, and network construction is based on remaining collaboration relationships. The specific steps are as follows:

(1) Construct an undirected binary co-authorship matrix based on author collaboration relationships, retaining only author pairs with collaboration frequency of 2 or more (3,971 pairs). Considering the distribution of collaboration frequencies, relationships with exactly 2 collaborations are recorded as "0," and those exceeding 2 as "1."

(2) Input the undirected binary co-authorship matrix and apply core-periphery structure analysis, yielding 398 authors in the core region and 2,911 authors in the periphery region.

(3) Appropriately adjust the core author group based on Price's core author formula, H-index, Hm index, publication count, and citation count. The specific adjustment process is as follows: Core region author removal: Based on author quantity distribution across different indicators, we separately identified 79 authors with 2 total publications in the past 5 years, 19 authors with 5 or fewer total citations, 24 authors with H-index of 1, and 26 authors with Hm-H index difference exceeding 0.2. After merging and deduplication, 99 authors were removed. Periphery region author addition: According to Price's core author formula, core authors need 7 or more publications. Therefore, we separately identified 53 authors with 7+ publications and 25+ citations, 18 authors with H-index of 7+, and 29 authors with Hm-H index difference of 0.01 or less. After merging and deduplication, 59 authors were added to the core region. After adjustment, the final core region includes 359 authors.

3.2 First Author-Other Author Directed Co-authorship Network

In a team, authors with higher first-author publication counts typically represent greater research activity and capability. Moreover, first authors and corresponding authors contribute more to papers than other authors. Therefore, to highlight the contribution of first authors (hereinafter "first authors"), we establish a directed co-authorship network between first authors and other authors.

First, we filtered papers from the dataset involving 359 core authors and excluding single-author publications, totaling 1,199 papers involving 741 first authors. Second, we constructed co-authorship pairs among these 741 authors based on their collaboration relationships. Only collaboration frequencies between first authors and other authors were calculated; relationships between non-first authors were excluded, making the position of first authors more prominent and simplifying inter-author relationships. Third, separate node and edge files were created. The node file consists of author unique identifiers, author names, and first-author publication counts. Edge type is set as directed, from source nodes (other authors) to target nodes (first authors), with edge weights revealed by the number of co-authored papers between authors. The overall network contains 2,408 nodes and 3,162 edges. Among them, there are 4 relatively large connected components with node counts of 795, 578, 376, and 176 respectively.

To present domain core teams more clearly, we increased the co-authorship threshold to 3, retaining only strong co-authorship relationships with 3+ collaborations between first authors and other authors, and teams with 3+ authors. After removing isolated nodes, we obtained Figure 4 [Figure 4: see original paper]. This network contains 25 domain core teams with 3+ authors, including 21 star-type teams (e.g., Rousseau Ronald team), 2 bus-type teams (e.g., Glanzel Wolfgang team), and 2 K-core type teams (e.g., Ding Ying team).

3.3 First Author-Corresponding Author Directed Co-authorship Network

Among all authors of a paper, the first author and corresponding author contribute the most. When publishing in international journals, the "U-shaped" authorship format is commonly used, where the first and last positions are most important, occupied by the first author and corresponding author respectively, with remaining authors listed in order of contribution [22]. Generally, corresponding authors are project leaders or mentors of first authors. To emphasize their rights and obligations in research work, corresponding authors are placed at the end with special identification. Furthermore, if the corresponding author's academic age is older than the first author's, the corresponding author typically provides academic guidance to the first author. Therefore, to further highlight the important positions of first authors and corresponding authors in paper publication and network structure, we construct a directed co-authorship network between first authors and corresponding authors.

The construction process is similar to the first author-other author directed co-authorship network, but only calculates co-authorship frequency between first authors and corresponding authors; relationships between other authors are excluded, retaining only strong co-authorship relationships between first and corresponding authors and further simplifying the co-authorship network to facilitate exploration of guidance and even mentorship relationships between authors. Separate node and edge files were created. The node file consists of author unique identifiers, author names, and first-author publication counts. Edge type is set as directed, from corresponding authors (source nodes) to first authors (target nodes), with edge weights revealed by the number of co-authored papers between authors.

The first author-corresponding author directed co-authorship network contains 614 nodes and 473 edges. The largest connected component involves 33 nodes; there are 11 connected components with 10+ nodes. The network overall exhibits weak connectivity characteristics, with 423 edges representing single collaborations, accounting for 89% of total edges.

To construct a judgment path for domain core team guidance/mentorship relationships, we use two teams with larger scale and representative network topology as examples: the Glanzel Wolfgang team (bus type) and the Ding Ying team (K-core type). As shown in Figure 5 [Figure 5: see original paper], the Ding Ying team's undirected binary co-authorship network is a typical K-core structure with very tight connections among authors. However, precisely because there are too many edges in the network, it is impossible to judge guidance or mentorship relationships among authors based on this graph.

We constructed the Ding Ying team's first author-other author directed co-authorship network and first author-corresponding author directed co-authorship network. Since only Bu Yi has both first-author and corresponding-author relationships with Ding Ying, the "first author-corresponding author directed co-authorship network" of Ding Ying consists of only one edge. Therefore, we merged it into Ding Ying's "first author-other author directed co-authorship network," using dashed directed edges for annotation. As shown in Figure 6 [Figure 6: see original paper], node size represents authors' first-author publication counts, solid arrows point from other authors to first authors, and dashed arrows point from corresponding authors to first authors. Compared with the undirected co-authorship network, the directed co-authorship network significantly simplifies relationships among authors in the Ding Ying team and highlights the position of first authors, facilitating guidance/mentorship relationship judgment.

As shown in Table 2 , in the Ding Ying team, the two authors with highest degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality are Ding Ying and Bu Yi, followed by Lu Chao, Xu Jian, Min Chao, and Huang Yong with relatively close metrics. These five authors are important nodes in the Ding Ying team's directed co-authorship network, collaborating frequently with others.

Clearly, if authors have a mentorship relationship, they belong to the same institution for a certain period. As shown in Table 3 , among authors in the Ding Ying team, those whose affiliated institutions include Indiana University (USA) are Ding Ying, Lu Chao, Huang Yong, Bu Yi, and Min Chao.

Based on normalized network centrality and authors' affiliated institutions, Ding Ying, Lu Chao, Huang Yong, Bu Yi, and Min Chao are important authors for mentorship relationship judgment. From academic age perspective: Ding Ying's academic age is 22 years, Lu Chao's is 5 years, Huang Yong's is 3 years, Bu Yi's is 4 years, and Min Chao's is 5 years. In a co-authorship relationship, scholars with older academic ages generally provide academic guidance to those with younger academic ages. Therefore, combining network node normalized centrality, authors' affiliated institutions, and academic age, Ding Ying provides guidance to Lu Chao, Huang Yong, Bu Yi, and Min Chao.

Integrating first author-corresponding author relationships, network node normalized centrality, authors' affiliated institutions, and academic age, Ding Ying and Bu Yi have a guidance relationship, with high probability of a mentorship relationship—that is, Ding Ying is Bu Yi's mentor (see Figure 7 [Figure 7: see original paper]). This judgment is verified by the personal academic homepages of Ding Ying (https://info.sice.indiana.edu/~dingying/me.html) and Bu Yi (https://buyi08.wixsite.com/yi-bu).

Following similar procedures, we judged the guidance/mentorship relationships in the Glanzel Wolfgang team. As shown in Figure 8 [Figure 8: see original paper], Glanzel Wolfgang has a high probability of being the mentor of Chi Peishan and Zhang Lin, which is verified by the three scholars' personal academic homepages.

From the mentorship relationship judgment processes of the Ding Ying and Glanzel Wolfgang teams, we can distill generalizable patterns. First, in co-authoring teams with star topology, central nodes with older academic ages and "radiating" arrow directions have a high probability of being mentors. This is because such nodes possess abundant academic network resources and often play assisting roles in research for first authors. As shown in Figure 9 [Figure 9: see original paper], in the "first author-corresponding author directed co-authorship network," we filtered 6 relatively large connected components with star topology. Nodes marked with triangles have high probability of being mentor/supervisor nodes, with directed edges presenting outward radiation patterns. The connected graph in the lower left corner also shows star topology, but the central node's directed edges are inward and aggregated. Such nodes belong to members with high first-author publication counts and strong research capabilities in the team, but we cannot determine whether they are mentor/supervisor nodes.

Second, in the "first author-other author directed co-authorship network," if two nodes simultaneously possess first author-corresponding author relationships, share the same affiliated institution, and have a certain gap in academic age, then there must be a guidance relationship between these two nodes, with high probability of a mentorship relationship. If nodes in the same co-authorship network have high normalized centrality, share the same affiliated institution, and have a certain gap in academic age, we can determine that there is a high probability of guidance relationships between network nodes, representing guidance from scholars with older academic ages to those with younger academic ages.

4 Team Research Tendency Analysis

4.1 Publication Tendency Analysis

From Section 3.1, we identified 25 domain core teams with 3+ authors. To reveal team publication tendencies, we constructed a team-journal two-mode network. First, we merged all publications by a team's authors as the team's publication count. Second, we established team-journal co-occurrence pairs, totaling 236 pairs. Third, we created separate node and edge files. The node file includes two node types: Type 1 consists of team unique identifiers, team names, and team publication counts; Type 2 consists of journal full names, journal abbreviations, and journal publication counts (journal publication count refers to the total number of publications by the 25 core teams in a journal over the past 5 years). Edge type is set as directed, from teams (source nodes) to journals (target nodes), with edge weights revealed by the number of team publications in each journal.

For clearer presentation, we used component analysis to segment the network, adjusting the edge threshold to 3. The largest connected sub-network is shown in Figure 10 [Figure 10: see original paper]. This sub-network consists of 20 nodes, including 5 journal nodes and 15 team nodes. Edge thickness represents the number of team publications in the journal, with a minimum of 3 publications. Five journals gather the largest number of core research teams in the informetrics field: SCIENTOMETRICS, JINFORMETR, and JASSOCIINFSCITECH.

Similarly, through team-journal two-mode network connected sub-networks, we can also find that the Venkatesh Viswanath team publishes more in information management and information systems journals, including MISQUART, JASCOCIINF SYST, INFORM SYSTRES, and JMANAGE INFORM SYST; the Pinto Maria team's publications concentrate in journals such as JLIBRINFSCI; and the Yao Yao team focuses on geographic information systems journals like INTJGEOGRINFSCI. This team's research field is relatively special compared to others. Through this analysis, we can see that most teams' research themes concentrate on popular domain directions, while some teams have more niche research directions.

4.2 Research Theme Analysis

To analyze research directions of core research teams in international information science, we need to reveal each team's publication themes. First, we merged all publications by a team's authors as the team's publication set. Second, we extracted author keywords from the team publication set, standardized the keywords, and calculated keyword frequencies to reflect team research themes through high-frequency keyword combinations, as shown in Table 4 .

When counting team publication numbers, we discovered an outlier. The Martin-Martin Alberto team had 116 publications over 5 years, far exceeding other teams. In this team, M. Thelwall had 62 first-author publications from 2016-2020, but most were single-author publications with low collaboration intensity with others. Such authors have strong individual research capabilities but few co-authored publications with other authors. Similar authors include L. Leydesdorff (20 first-author publications) and L. Bornmann (41 first-author publications).

To classify core teams in international information science by research theme, we conducted thematic clustering analysis on all 12,098 papers in the field from the past 5 years. First, we extracted author keywords from the field's document set and standardized the keywords. Second, we constructed a keyword co-occurrence matrix based on inter-document keyword co-occurrence relationships. Third, adjusting the co-occurrence strength threshold to 44, 80 keywords participated in clustering, using the modularity algorithm for thematic clustering, as shown in Figure 11 [Figure 11: see original paper].

Based on Figure 11, we classified core teams in international information science, obtaining Table 5 . In addition to informetrics research directions, the remaining five research directions each have 2-3 domain core teams, indicating relatively lower competition.

This study reveals author directed relationships in visual form, highlighting important authors in teams. On one hand, this can provide auxiliary judgment for research evaluation based on co-authorship; on the other hand, it can facilitate associated author retrieval in information systems and can be extended to author citation analysis. Additionally, the research results help scholars in other domains or new scholars entering the LIS field to quickly understand major international research teams and research directions, providing reference for strengthening collaboration in this field. However, this study has limitations. For instance, we used absolute values to reveal collaboration strength between authors, whereas in reality, collaborating with the same number of papers with ordinary domain scholars versus domain authorities necessarily produces different impacts. In future work, we will consider assigning weights to author academic influence and using relative rather than absolute values to reveal author collaboration relationships and strengths.

References

[1] Chen Chunhua, Yang Yingshan. Research on scientific research management based on team operation mode [J]. Science & Technology Progress and Policy, 2002, 19(4): 79-81. [2] Guo Yiting, Deng Shengli. Review of user relationship network research [J]. Digital Library Forum, 2014(8): 59-65. [3] Oliveira M, Gama J. An overview of social network analysis [J]. Wiley interdisciplinary reviews data mining & knowledge discovery, 2012, 2(2): 99-115. [4] Qiu Junping, Wu Chao. Research on co-authorship relationships among domestic informetrics authors based on social network analysis [J]. Library and Information Service, 2011(6): 12-17. [5] Li Gang, Li Chunya, Li Xiang. Research on scientific research team discovery based on social network analysis [J]. Library and Information Service, 2014, 58(7): 63-70, 82. [6] Li Gang, Liu Xianhong. Research on academic leader identification in scientific research teams based on cooperation network centrality indicators [J]. Science and Technology Management Research, 2016(8): 127-132. [7] Yu Yongsheng, Dong Cheng, Han Hongqi, et al. Research on scientific research team identification method based on social network analysis—identifying scientific research team leaders based on iterative betweenness centrality ranking method [J]. Information Studies: Theory & Application, 2018, 41(7): 105-110. [8] Zhao Yuxiang, Tong Chong, Zhang Yan. Research on actor collaboration characteristics in digital humanities open data competitions [J]. Library Tribune, 2020, 40(11): 56-67. [9] Pang Hongqian, Fang Shu, Yang Bo, et al. Analysis of scientific research team collaboration closeness—a case study of WISE Lab of Dalian University of Technology [J]. Library and Information Service, 2011, 55(4): 28-32, 99. [10] Shen Gengyu, Huang Shuiqing, Wang Dongbo. Research on scientific research team discovery method using author cooperation co-occurrence as source data [J]. New Technology of Library and Information Service, 2013(1): 57-62. [11] Gonzalez-Alcaide G, Park J, Huamani C, et al. Scientific authorships and collaboration network analysis on Chagas disease: papers indexed in PubMed (1940-2009) [J]. Revista do instituto de medicina tropical de sao paulo, 2012, 54(4): 219-223. [12] Liu X, Bollen J, Nelson M, et al. Co-authorship networks in the digital library research community [J]. Information processing & management, 2005, 41(6): 1462-1480. [13] Krichel T, Bakkalbasi N. A Social network analysis of research collaboration in the economics community [J/OL]. Journal of information management & scientometrics to appear, 2006, 3(1). [2021-07-18]. http://core.ac.uk/download/pdf/11879617.pdf. [14] Wang X, Fu Y, Huang Y, et al. Identifying R&D partners using SAO analysis: a case study of dye-sensitised solar cells [J]. International journal of technology management, 2019, 81(1/2): 70-93. [15] Kang J, Lee J, Jang D, et al. A methodology of partner selection for sustainable industry-university cooperation based on LDA topic model [J/OL]. Sustainability, 2019, 11(12). DOI: 10.3390//su11123478. [16] Zhou Ping, Leydesdorff L, Wu Yishan. Visualization of citation environment of Chinese scientific journals [J]. Chinese Journal of Scientific and Technical Periodicals, 2005, 16(6): 773-780. [17] Hao Zhihan, Chen Yang, Wang Pusheng. Research collaboration network centrality and academic impact—a sample from

Science (2000-2018) [J]. Library Tribune, 2020, 44(4): 79-88. [18] Liu Yang. Research on author co-authorship relationships in library and information science based on social network analysis [J]. Journal of the Library Science Society of Sichuan, 2019(2): 97-100. [19] Yao Xue, Xu Chuanping, Li Jie, et al. Construction of core author user database for a scientific journal based on Price's law and Pareto principle and online submission system [J]. Acta Editologica, 2017, 29(1): 64-66. [20] Gong Shuye. Core author analysis of Information Science based on h-index and hm-index [J]. Information Science, 2013, 31(1): 82-85, 95. [21] Bu Y, Ding Y, Liang X, et al. Understanding persistent scientific collaboration [J]. Journal of the Association for Information Science & Technology, 2018, 69(3): 438-448. [22] Lei Xue, Wang Lixue, Zeng Jianxun. Construction and analysis of directed author co-authorship networks [J]. Library and Information Service, 2015, 59(5): 94-99.

Author Contributions:

Gao Nan: Research design, data processing and figure creation, full paper writing; Zhou Qingshan: Research discussion and revision, paper revision.

English Abstract

[**Purpose/Significance**] To facilitate institutional talent introduction and establish a standardized process for identifying domain core teams, this study conducts mining research on domain core teams from four aspects: core area author identification, team guidance/mentorship relationship discrimination, team publication tendency, and team research direction. [**Method/Process**] The process involves: first, conducting preliminary data standardization; then, combining an undirected binary co-authorship matrix with core-periphery structure analysis to identify core area authors, with appropriate adjustments based on multi-dimensional indicators; finally, identifying domain core research teams based on a directed co-authorship network between first authors and other authors. [**Result/Conclusion**] In clarifying the basic process for judging team guidance/mentorship relationships, generalizable patterns are distilled: (1) In a co-authoring team of the "outer star" topology, the node at the center has a higher probability of being a mentor, while authors of the "cohesive star" topology whose co-authoring team is located in the center of the network have strong scientific research capabilities; (2) In a directed weighted network between the first author and the other authors, if two nodes have the relationship between the first author and the corresponding author at the same time, the same affiliated institution, and a certain difference in academic age, there must be a guidance relationship, and a high probability of a mentoring relationship between these two nodes.

Keywords: social network analysis; directed weighted network; co-authoring; core-edge structure