Collaboration Patterns of AI Research Teams: A Comparative Study (Postprint)

Authors: Wang Yuefen, Yang Xue, Yu Houqiang, Cao Jiajun

Date: 2023-04-01T16:16:02+00:00

Abstract

[目的/意义] To explore the collaboration patterns of research teams in the artificial intelligence field and compare the differences and impacts among teams with different collaboration patterns. [方法/过程] Taking identified leading AI research teams as research subjects, core scholars are identified within the teams based on scholars' collaboration frequencies and social network metrics, thereby classifying the collaboration patterns of AI research teams and conducting exemplary analyses of teams with different patterns. On this basis, comparative analyses are performed across leading teams with different collaboration patterns from multiple dimensions, including network structural characteristics, research performance, and geographical distribution. [结果/结论] The collaboration patterns of AI research teams are classified into four types: single-core pattern, dual-core pattern, multi-core pattern, and balanced pattern, among which research teams with single-core and dual-core patterns demonstrate relatively superior performance across all examined dimensions.

Full Text

The Collaboration Patterns and Comparative Analysis of Research Teams in the Artificial Intelligence Field

Wang Yuefen[1], **Yang Xue**[1], **Yu Houqiang**[1], **Cao Jiajun**[1,2] [1]School of Economics and Management, Nanjing University of Science & Technology, Nanjing 210094 [2]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798

Abstract: *[Purpose/Significance]* This paper investigates the collaboration patterns of research teams in the artificial intelligence (AI) field and compares the differences and impacts among teams with different collaboration patterns. *[Method/Process]* Taking identified leading AI research teams as the study object, we identify core scholars within teams based on scholars' collaboration numbers and social network metrics, thereby classifying the collaboration patterns of AI research teams and providing illustrative analyses for each pattern. Building upon this, we conduct comparative analyses of leading teams with different collaboration patterns across multiple dimensions, including network structural characteristics, research performance, and geographical distribution. *[Result/Conclusion]* The collaboration patterns of AI research teams are classified into four types: single-core pattern, dual-core pattern, multi-core pattern, and equilibrium pattern. Among these, research teams with single-core and dual-core patterns demonstrate superior performance across all studied dimensions.

Keywords: artificial intelligence; scientific research collaboration; collaboration pattern; core scholars; data analysis

Classification Number: G253.1
DOI: 10.13266/j.issn.0252-3116.2020.20.002

Research collaboration patterns refer to the general modes of behavior among scientific research entities, characterized by simplicity, repetitiveness, structure, and stability, reflecting the collaborative relationships and underlying patterns among participants in the scientific research process. Different collaboration patterns not only create distinct behavioral characteristics within research groups but also generate varying research performance outcomes.

The global proliferation of artificial intelligence has not only enabled the emergence of large-scale research teams but also continuously fostered new research findings through ongoing collaboration, driving the extensive and in-depth development of AI. What collaboration patterns exist in AI research? What features define these patterns? How do the structural characteristics, research performance, and geographical distribution of research teams differ across various collaboration patterns? Investigating these questions holds significant value for understanding the collaborative landscape and trends within research teams and for providing deeper interpretation and analysis of AI field development.

Existing research on collaboration pattern classification primarily follows two approaches: first, classifying co-authorship relationships based on specific author attributes; and second, classifying patterns based on collaboration network structural features and group connectivity characteristics. Author attributes include institution, geography (country, city), age, gender, title, etc., which enable classification of research paper co-authorship patterns. For instance, Li Keli [2] examined research collaboration between 14 research universities in the middle Yangtze River urban agglomeration and institutions within and outside the region from an institutional affiliation perspective, revealing that collaboration patterns mainly include university-university, university-enterprise, government-university, and university-research institute collaborations, with university-university and university-research institute collaborations dominating, and showing that inter-institutional collaboration tends to concentrate on advantageous resources. Guo Yingtao et al. [3] analyzed author collaboration patterns in library and information science from a geographical perspective, categorizing patterns as intra-city, inter-city, and international collaboration, noting that intra-city collaboration is the mainstream pattern both domestically and internationally. Some scholars have adopted comprehensive institutional and geographical perspectives, such as Y. Wang et al. [4], who studied co-authored papers from Chinese scientific databases and classified Chinese author collaboration into four patterns: collaboration within the same institution, collaboration across different institutions within the same region, inter-regional collaboration, and international collaboration. Qiu Junping et al. [5] examined author collaboration from an age perspective in information science, analyzing age structures in two-author and three-author collaborations and the overall network, finding that young-young collaboration is the most important pattern and advocating for intergenerational collaboration. B. Bozeman et al. [6] found that in collaboration strategies, female researchers prefer same-gender collaboration compared to male researchers with the same title. From a title perspective, existing research shows that the mainstream collaboration pattern in domestic research is teacher-student collaboration [7], while internationally it is teacher-teacher collaboration [8]. Cui He et al. [7] analyzed collaboration patterns among different title combinations in Chinese higher education research, finding that professor-PhD student collaboration yields the highest publication output.

On the other hand, classification based on collaboration network structural features and group connectivity involves calculating network metrics such as author collaboration network density, node centrality, and inter-node distances to characterize collaboration patterns. For example, Cao Ling et al. [9] identified highly productive authors and analyzed their collaborator networks, using metrics such as node count, network density, and publication numbers to classify individual network collaboration patterns into research team-type collaboration and major project collaboration patterns. Qiu Junping et al. [10] classified author collaboration patterns in knowledge transfer into four network patterns—single-mode subnet, dual-core subnet, development-mode subnet, and complete subnet—based on node count, distribution, and network density. Subsequently, Han Fangfang et al. [11] and Wang Xiaorong et al. [12] applied similar classifications to digital library and diabetes research fields. Liu Bei et al. [13] classified author collaboration patterns in domestic information science into single-point, dual-core, streamline, core, bridge, and network-frame types. Dong Lingxuan et al. [14] referenced Qiu Junping et al. [10] and Liu Bei et al. [13] to classify iConference paper author collaboration networks into four patterns: single-mode, complete-mode, bridge-mode, and multi-group collaboration patterns.

In summary, existing research has produced numerous results revealing characteristics and patterns of scientific collaboration. However, most studies start from individual authors, involve relatively small-scale disciplines, and have limited empirical data, leaving conclusions primarily at the level of pattern identification and simple interpretation. Therefore, exploring collaboration patterns and their impacts warrants deeper investigation.

2 Research Approach and Data Processing

2.1 Research Approach

Figure 1 [Figure 1: see original paper] illustrates the overall process for classifying and comparing collaboration patterns of AI research teams.

The data source is the Web of Science (WoS) Core Collection, with the logical search query WC="Computer Science, Artificial Intelligence," selecting records from the AI category within the Computer Science discipline classification. The search period was set to 2009-2018, with retrieval conducted in April 2019, yielding 421,148 records (including papers, books, etc.).

After obtaining raw data, we first cleaned institutional names to address author name ambiguity issues affecting team identification quality. We then performed author name disambiguation by combining co-authors and author affiliations—judging whether authors with the same name representation belonged to the same author based on institutional name similarity. From a social network analysis perspective, we employed the Louvain community detection algorithm to identify research teams in the AI field. This process identified 23,423 research teams involving 186,997 authors. Based on this, we selected six indicators—total publications, total citations, h-index, weighted degree centrality, betweenness centrality, and closeness centrality—to rank teams by the sum of each team member's indicator values (publications and citations calculated using fractional counting). We identified the top 15 teams for each indicator as leading teams and took their union, resulting in 68 teams as our study objects.

By constructing associations among countries/regions, institutions, and articles based on raw data, we further identified scholars' affiliated institutions and countries/regions according to the correspondence between scholars and articles. To analyze the national distribution of research teams, we merged scholars' regions into corresponding country representations (e.g., "Taiwan" was grouped under China; "England," "Scotland," "Northern Ireland," and "Wales" were all grouped under the United Kingdom; "Reno" was grouped under the United States). We ignored some records where specific country names could not be traced.

2.2 Data Source and Processing

[Content merged with 2.1 for logical flow]

3 Classification of Leading Team Collaboration Patterns

3.1 Classification Method and Process

Node degree centrality, representing the number of other nodes directly connected to a node, is a common metric for measuring node centrality in networks. In a social network, an actor with many direct connections occupies a central position and possesses greater "power" within the network. Therefore, higher node degree centrality indicates greater importance. In author co-authorship networks, node degree centrality represents the number of collaborators a scholar has and can measure scholar influence to identify core scholars. Weighted degree centrality incorporates edge weights, with a node's weighted degree centrality being the sum of edge weights connected to that node. In our constructed author co-authorship network, edge weights reflect collaboration intensity between two authors, calculated from their total number of collaborations.

Drawing from Fan Ruxia et al. [15], who identified highly collaborative authors, we used scholars' degree centrality, weighted degree centrality, and publication counts within their team networks as classification criteria. Through multiple experiments, we determined more reasonable thresholds and defined team collaboration patterns by identifying core scholars from a network perspective. Let Di represent the degree centrality of node i (number of collaborators), WDi represent the weighted degree centrality of node i (collaboration frequency), pubi represent the publication count of author i, and Z represent the total team size.

The specific operational procedure is as follows: 1. Sort nodes in descending order based on Di. If Di = Dj, compare WDi and WDj. If WDi = WDj, further compare pubi and pubj. 2. Identify core scholars in the team: First, identify the team's unique core based on the ratio (α) between the collaboration number of the author with the highest degree (Di) and total team size (Z). If the first node i satisfies Di/Z $\geq \alpha$ ($0 < \alpha < 1$), then node i is the team's unique core. Otherwise, proceed as follows: If nodes i, j, k satisfy Di > Dj > Dk (if Dj = Dk, compare WDi and WDj), and the number of core nodes/Z ≤ 0.2, such that (Di + Dj + Dk)/Z ≥ 0.8 (based on the Pareto principle), then nodes i, j, k are core scholars. Note that when summing Di, Dj, Dk, duplicate scholars among nodes i, j, k must be excluded (if nodes i, j, k all collaborate with node a, a is counted only once). 3. Classify collaboration patterns: Divide patterns according to the number of core scholars in the team, as shown in Table 1 .

3.2 Threshold Selection for α

The threshold setting for α (the ratio of an author's collaboration number to total team size, Di/Z) decisively affects classification results. We tested α values of 0.5, 0.6, 0.7, and 0.8 (with α satisfying $0.5 \leq \alpha \leq 0.8$ to ensure core node status while avoiding conflict with thresholds for dual-core and multi-core patterns). Different classification results were obtained for different collaboration patterns, as shown in Table 2 .

Table 2 reveals that as threshold α increases, the number of single-core pattern teams continuously decreases, while dual-core, multi-core, and equilibrium pattern teams increase. When $\alpha = 0.5$ or 0.6, leading teams can be divided into three patterns: single-core, multi-core, and equilibrium patterns, with single-core pattern dominating. However, when α increases to 0.7 or 0.8, dual-core pattern teams emerge, and multi-core and equilibrium pattern teams also increase.

To determine the optimal threshold, we first compared classification results between $\alpha = 0.5/0.6$ and $\alpha = 0.7/0.8$. Examining team network structures revealed that when $\alpha = 0.5$ or 0.6, some single-core pattern teams clearly exhibited dual-core structures, while at $\alpha = 0.7$ or 0.8, these could be distinctly classified as dual-core patterns. This indicates that using 0.5 or 0.6 as threshold α is unreasonable. Further comparison between $\alpha = 0.7$ and $\alpha = 0.8$ showed that even at $\alpha = 0.7$, some single-core pattern teams were clearly dual-core structures. Therefore, to distinguish single-core pattern from other patterns as accurately as possible, we set the threshold α for single-core pattern classification at 0.8 after multiple experiments. Notably, some leading teams contain only one scholar without any co-authorship relationships with others; these fall outside our collaboration pattern research scope. Thus, our final study object comprises 67 leading teams.

3.3 Illustrative Analysis of Leading Team Collaboration Patterns

3.3.1 Single-Core Pattern The single-core pattern features one core scholar in the team. Based on our threshold selection, this core scholar collaborates with over 80% of team members, manifested in the network as over 80% of nodes connecting to this core node, creating a radial structure from the center. In this pattern, the core scholar occupies an absolutely dominant position, and collaborating with the core scholar is equivalent to connecting with the entire team. Understanding the core scholar's research themes effectively grasps the whole team's research direction.

Figure 2 [Figure 2: see original paper] shows a typical single-core pattern team. This team has 49 nodes, with J. Cao_1 (Cao Jinde) having the highest degree centrality of 45 collaborators, representing over 80% of total team members, making him the unique core node. Professor Cao Jinde from Southeast University has formed a leading AI research team by collaborating with scholars from numerous institutions including Nanjing University of Information Science and Technology and Jiangsu Normal University. Analysis of the team's research output reveals three main research themes: global synchronization, global exponential stability, and multi-agent systems.

3.3.2 Dual-Core Pattern The dual-core pattern features two core scholars whose collaborator union accounts for over 80% of total team members. Networks of dual-core pattern teams often exhibit bridge structures, where the two core scholars serve as connection points between two sub-networks. Their collaboration forms a bridge between sub-networks, which together maintain the team's structure. Dual-core pattern teams may play important roles in promoting interdisciplinary and inter-institutional development in AI.

Figure 3 [Figure 3: see original paper] shows a typical dual-core pattern team. This team has 34 nodes, with D. Zhang having the highest degree centrality of 21 collaborators, which does not exceed 80% of total team members. Following our classification method, after sorting scholars by degree centrality and excluding duplicate collaborators, the second-ranked scholar L. Zhang has 20 collaborators. Their combined collaborator count is 30 (if a scholar collaborates with both D. Zhang and L. Zhang, they are counted only once), exceeding 80% of total team members, with core scholar count less than 20% of team size. Therefore, D. Zhang and L. Zhang are core nodes, making this a dual-core pattern team. Both scholars are from Hong Kong Polytechnic University, each forming their own research sub-networks with collaborators who have both intersections and differences, collectively forming this leading research team. The team's research focuses on face recognition and convolutional neural networks, specifically involving face images, sparse representation, and image classification.

3.3.3 Multi-Core Pattern Multi-core pattern teams can be further divided into triple-core, quadruple-core, etc., based on core member count. These patterns connect multiple multi-person collaboration sub-networks through several core nodes to form leading research teams. Collaboration between sub-networks in such teams facilitates deeper cooperation in AI and can integrate advantages from various disciplines and research methods to expand research scope and directions [14].

Figure 4 [Figure 4: see original paper] shows a typical multi-core pattern team with four core scholars. This team has 37 nodes, with M. Dastani having the highest degree centrality of 21 collaborators, not exceeding 80% of total team members. After sorting by degree centrality and weighted degree centrality and excluding duplicates, the core scholars are M. Dastani, J. Meyer_1, J. Meyer, and J. Dix_1. The first three core scholars are from Utrecht University in the Netherlands, while J. Dix_1 is from Clausthal University of Technology in Germany. These four scholars and their respective collaborators collectively form this leading team. The research primarily focuses on multi-agent systems, incorporating cloud computing methods and exploring robot motion coordination with emphasis on effectiveness and applicability.

3.3.4 Equilibrium Pattern The equilibrium pattern features no core scholars, with relatively balanced and dispersed collaboration among all team members. The network appears more scattered than multi-core patterns, with fewer connections between sub-teams. Figure 5 [Figure 5: see original paper] shows a typical equilibrium pattern team.

This team has 23 nodes, with A. Gumus having the highest degree centrality of 11 collaborators, not exceeding 80% of total team members. Following our classification method, the top 20% of scholars (first 4 scholars) have a combined collaborator count of 15, not exceeding 80% of total team members. Therefore, no core scholars exist, making this an equilibrium pattern team. The team comprises multiple sub-teams with minimal interconnections, resembling a grape cluster structure without prominent core scholars, though bridge scholars exist between sub-teams who could further develop collaborations to form larger, more closely connected research teams.

4 Comparative Analysis of Leading Teams with Different Collaboration Patterns

4.1 Network Structure Characteristics Analysis

Team network structure characteristics are represented by social network metrics such as density, clustering coefficient, and average shortest path length, reflecting the degree of interconnection and clustering among nodes.

(1) **Density**: The ratio of actually existing edges to possible edges in a network, measuring the intensity of node interconnections.

(2) **Clustering Coefficient**: The ratio of actually existing edges to possible edges among node i's neighbors, with network clustering coefficient being the average of all nodes' clustering coefficients. In co-authorship networks, this represents the probability that a scholar's collaborators also collaborate with each other, measuring network clustering [16].

(3) **Average Path Length**: The average of distances (shortest path lengths) between any two nodes in the network.

Based on classified collaboration patterns, we calculated these metrics for leading teams and computed average values for each pattern to analyze differences in network structure characteristics, as shown in Table 3 .

Table 3 shows that regardless of pattern, leading teams have relatively low network density and clustering coefficient values, indicating sparse connections among nodes and limited research exchange activities. However, comparing across patterns reveals that single-core and dual-core pattern leading teams have higher network density and clustering coefficient values and shorter average path lengths than multi-core and equilibrium patterns, indicating higher clustering and closer scholar connections. This suggests that patterns with one or two core scholars better facilitate academic exchange and knowledge sharing.

4.2 Research Performance Analysis

Classifying and defining leading team collaboration patterns aims primarily to compare performance across patterns to identify which patterns most benefit team productivity. Team research performance assessment involves multiple aspects, including bibliometric indicators, patent metrics, and economic-financial indicators [17]. We selected five bibliometric indicators related to research papers: total team publications (sum of team members' publications), per-capita publications, total team citations (sum of team members' citations), per-capita citations, and average citations per article. The first three measure overall team performance, while the latter two measure individual performance. We calculated average values for each pattern to analyze pattern impacts on productivity, as shown in Table 4 .

Table 4 shows that single-core, dual-core, and multi-core patterns far exceed equilibrium patterns across all performance indicators, with single-core and dual-core patterns showing particularly outstanding performance. This indicates that patterns with one or two core scholars yield higher research output and improve team productivity efficiency.

4.3 Geographic Distribution Analysis

Geographic distribution is a crucial factor in team analysis, reflecting team reach, influence, and internal/external exchange levels. We compared geographic distributions of leading teams across patterns by tracking scholars' countries.

4.3.1 All Scholars' Country Distribution We tracked countries of all scholars in leading teams, listing the top 10 countries by frequency and their continents for each pattern, as shown in Table 5 .

China appears most frequently across all patterns, indicating that Chinese scholars in AI are not only numerous but also hold important positions in teams. Analysis by continent shows that single-core pattern teams have the broadest continental distribution, covering Asia, Europe, Africa, North America, and Oceania, while dual-core, multi-core, and equilibrium pattern teams concentrate mainly on collaborations between Asia and Europe.

4.3.2 Core Scholars' Country Distribution We further analyzed core scholars' country and continent distribution across patterns, as shown in Table 6 (equilibrium pattern excluded as it has no core scholars).

Similar to all scholars' distribution, Chinese scholars constitute the highest proportion across the three patterns. Comparing patterns reveals that core scholars in single-core pattern teams distribute more widely across five continents (Asia, Europe, Africa, North America, Oceania), while core scholars in dual-core and multi-core pattern teams mainly distribute across Asia, Europe, and North America.

5 Conclusion and Future Directions

This study takes identified leading AI teams as research objects, constructs a data analysis-based method and process for classifying research collaboration patterns, and excavates differences among patterns from three dimensions—network characteristics, research performance, and geographic distribution—to provide decision-making support for improving team efficiency, expanding influence, strengthening team construction, and promoting scholar collaboration. It also helps researchers grasp AI collaboration trends and promotes AI field development. Main conclusions are:

(1) Based on the combination of research teams and individual scholars, leading team collaboration patterns are classified into four types: single-core, dual-core, multi-core, and equilibrium patterns, each presenting distinct structural states where core scholars play significant roles.

(2) From a network characteristics perspective, single-core and dual-core pattern leading teams show higher network density and clustering coefficient values and shorter average path lengths than multi-core and equilibrium patterns, indicating higher clustering and closer scholar connections.

(3) From a research performance perspective, single-core and dual-core patterns outperform multi-core and equilibrium patterns in both overall team performance and individual scholar performance, demonstrating higher research output and productivity efficiency.

(4) From a geographic distribution perspective, single-core pattern teams have broader collaboration scope, while dual-core, multi-core, and equilibrium pattern teams concentrate more on collaborations among Asia, Europe, and North America. Core scholars in single-core pattern teams distribute across five continents, while those in dual-core and multi-core patterns mainly distribute across Asia, Europe, and North America.

This study has limitations, primarily focusing only on AI research teams. To improve method generalizability and research depth, future work could apply this method to other fields and further compare differences across patterns based on author attributes such as age, title, and mentor-mentee relationships.

References

[1] Wang Xinming, Ding Jingda. Review of research on author collaboration patterns in scientific papers [J]. Modern Information, 2018, 38(8): 172-177.

[2] Li Keli. Research collaboration innovation among universities in the middle Yangtze River urban agglomeration: Patterns, characteristics, and strategies—Based on social network analysis of 14 research universities [J]. Journal of Xiangtan University (Philosophy and Social Sciences Edition), 2017, 41(4): 23-27, 38.

[3] Guo Yingtao, Yang Siluo, Kuang Piao. Comparative study on collaboration patterns in library and information science between China and abroad [J]. Information Studies: Theory & Application, 2013, 36(10): 120-123, 128.

[4] Wang Y, Wu Y, Pan Y, et al. Scientific collaboration in China as reflected in co-authorship [J]. Scientometrics, 2005, 62(2): 183-198.

[5] Qiu Junping, Yu Houqiang. Analysis of scientific collaboration age structure of scholars in China's information science field [J]. Information Science, 2014, 32(8): 3-8.

[6] Bozeman B, Corley E. Scientists' collaboration strategies: Implications for scientific and technical human capital [J]. Research Policy, 2004, 33(4): 599-616.

[7] Cui He, Wang Dan. Research on individual, institutional, and regional collaboration in China's higher education research field—Based on co-authorship analysis of 18 Chinese core educational journals in 2016 [J]. China Higher Education Research, 2017(4): 48-55.

[8] Benckendorff P, Zehrer A. Career and collaboration patterns in tourism research [J]. Current Issues in Tourism, 2016, 22(16): 1-19.

[9] Cao Ling, Jin Yinyu, Cao Jinshuang. Empirical analysis of meteorological scholars' collaboration patterns in China based on WoS data [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2013, 5(6): 566-572.

[10] Qiu Junping, Li Wei. Empirical analysis of author collaboration patterns in knowledge transfer field [J]. Information Studies: Theory & Application, 2012, 35(3): 53-56, 46.

[11] Han Fangfang, Zhou Chang. Research on author collaboration patterns in digital library field [J]. Knowledge Management Forum, 2013(4): 38-43.

[12] Wang Xiaorong, Lü Yanhua, Zhang Yingjie, et al. Analysis of highly cited paper author collaboration patterns in diabetes field based on CiteSpace [J]. Chinese Journal of Medical Library and Information Science, 2014, 23(11): 68-74.

[13] Liu Bei, Yuan Yi, Boutin Eric. Application of social network analysis in paper collaboration networks [J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(3): 407-417. doi: 10.3969/j.issn.1000-0135.2008.03.013.

[14] Dong Lingxuan, Liu Youhua, Zhu Qinghua. Research on iConference paper author collaboration based on SNA [J]. Journal of Intelligence, 2013, 32(10): 82-88.

[15] Fan Ruxia, Zeng Jianxun, Gao Yarui. Research on dynamic academic influence pattern recognition of scholars based on collaboration networks [J]. Data Analysis and Knowledge Discovery, 2017, 1(4): 30-38.

[16] Wang Feifei, Chen Xiaoxuan, Yang Chenyuyan. Research team identification and output performance measurement in gene editing field from collaboration network perspective [J]. Journal of Zhongzhou University, 2018, 35(3): 75-80.

[17] Meng Can, Li Qiang, Liu Wenbin. Construction of research institution evaluation index system based on 3E theory [J]. Studies in Science of Science, 2007(5): 908-914, 852.

Author Contributions

Wang Yuefen: Conceptualization, supervision of manuscript revision and finalization.
Yang Xue: Participation in research design discussion, responsible for literature review, writing, and revision.
Yu Houqiang: Conceptualization, research design, supervision of writing and revision.
Cao Jiajun: Participation in logical discussion and manuscript revision.

Note: Figure translations are in progress. See original paper for figures.