Postprint: Evolutionary Study of High-Productivity Research Teams in Artificial Intelligence

Authors: Zou Bentao, Wang Yuefen, Yu Houqiang

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] [JP+1] Team collaboration has become a crucial organizational form for knowledge innovation in contemporary times. Investigating the dynamic evolution patterns of research teams from a dynamic network perspective holds significant importance for facilitating the discovery, formation, and management of research teams. [Method/Process] Taking the field of artificial intelligence as a case study, this research employs the Louvain community detection algorithm to identify research teams within the artificial intelligence domain. By calculating the extreme value distributions of four topological indicators in team collaboration networks—specifically, the number of nodes, number of edges, network density, and average clustering coefficient—we examine the characteristics and patterns of high-productivity team evolution from both micro and macro perspectives, aiming to reveal the intrinsic drivers underlying research team evolution. [Results/Conclusion] From a micro perspective, the extreme value distributions of co-authorship network topological indicators unveil the dynamic attributes of high-productivity team evolution in the artificial intelligence field; from a macro perspective, high-productivity teams demonstrate evolutionary commonalities in network density and average clustering coefficient, with most teams catalyzing the generation of new collaborative relationships throughout their evolution. When examining team evolutionary pathways, the phenomenon of "small group" collaboration in each period proves prominent within high-productivity teams in the artificial intelligence domain, and cooperation among these "small groups" directly influences the overall trajectory of the team.

Study on the Evolution of Prolific Research Teams in the Field of Artificial Intelligence

Zou Bentao[1], **Wang Yuefen**[1,2], **Yu Houqiang**[1,2]
[1]School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094
[2]Jiangsu Social Public Security Technology Cooperative Innovation Center, Nanjing 210094

Abstract: [Purpose/Significance] Team collaboration has become an important organizational form for knowledge innovation today. Exploring the dynamic evolution patterns of research teams from a dynamic network perspective is of great significance for promoting the discovery, formation, and management of scientific research teams. [Method/Process] Taking the field of artificial intelligence as an example, this paper employs the Louvain community discovery algorithm to identify research teams in the AI domain. By calculating the extreme value distributions of four topological indicators—node count, edge count, network density, and average clustering coefficient—in team collaboration networks, it examines the evolutionary characteristics and patterns of high-yield teams from both micro and macro perspectives, aiming to reveal the intrinsic motivations behind team evolution. [Result/Conclusion] From a micro perspective, the extreme value distributions of co-authorship network topological indicators reveal the dynamic attributes of high-yield team evolution in the AI field. From a macro perspective, high-yield teams exhibit evolutionary commonalities in network density and average clustering coefficient, with most teams fostering an increasing number of new collaborative relationships during evolution. Combined with team evolution paths, the phenomenon of "small group" collaboration is prominent across periods in AI high-yield teams, and cooperation between these small groups directly affects the overall team' s trajectory.

Keywords: artificial intelligence; research team; extreme value distribution; dynamic evolution; topological structure

The increasing interdisciplinary nature and complexity of scientific research problems are driving research teams to bring together more experts from different fields to conduct collaborative research [1]. Notably, this team-based collaboration has been an intrinsic attribute of scientific activities since the birth of science, and research over recent decades indicates that it remains one of the important characteristics of scientific research today [2-3]. Moreover, research teams make significant contributions to promoting scientific discoveries and output [4]. For instance, stable teams are more likely to bring their members more JCR (Journal Citation Reports) journal publications [5] and have opportunities to undertake more research projects and apply for more patents [6]. By exploring and analyzing research teams from fields such as computer science, sociology, physics, and biology from different perspectives [8], researchers have even formed a specific research field called "Science of Team Science," which focuses on the organization, communication, and research processes of scientific teams [9].

Early research primarily adopted a static network perspective, treating teams in a certain time period or at a certain moment as constant, static structural entities [10-11]. However, in reality, such interaction networks are constantly changing over time [12-13]. Team members and their relationships are dynamically evolving. For example, after new members join a team, they will evaluate whether the team can achieve their individual-level goals, while the team will assess whether the member can enhance the team' s overall value. The degree of matching between these two factors will determine whether the member can maintain a stable relationship with the team [14]. When the research object is limited to research teams, their evolution will be influenced by more factors. For example, many research teams are formed temporarily to apply for or complete specific projects [15], and scientific collaboration is also affected by factors such as the geographical distance between scholars and their scientific visibility, prominence, and recognition [16]. Therefore, researchers no longer only analyze the static network properties of teams but go further to analyze team evolution and attempt to reveal the factors affecting evolution [17].

Existing research mainly focuses on: Identifying events to characterize and discover evolutionary states [21], such as Y. Wang et al. [22] defining four events—birth, split, merge, and death—to explore the dynamic evolution characteristics of communities from the perspectives of network scale and network members; G. Palla et al. [24] identifying events such as born, grow, split, merge, contract, and disappear, analyzing their correlation with community age and scale from a stationarity perspective. These methods can better handle unstable changing communities. Characterizing and reflecting the overall evolution through topological structure changes [19], such as A.L. Barabási et al. [20] calculating the evolution of co-authorship network topological indicators within sliding time windows, pointing out that during evolution, authors tend to choose scholars with more co-authors for collaboration; Guo Shiyue et al. [27] analyzing the evolution of teams in terms of topological indicators such as scale, network density, and average distance. Combining semantic information such as themes to analyze evolution through topic correlation, with F. Osborne et al. [25] pointing out that attention should be paid to the evolution of relationships between authors and research topics over time; L. Backstrom et al. [28] proposing to explore whether team evolution is "topic-driven" (Topics follow people) or "people-driven" (People follow topics). Treating research communities as teams for evolution analysis through conceptual extension [16, 26] to uncover changes in network structural properties [18], evolutionary characteristics [12], and the internal driving forces of dynamic changes [7].

Currently, methods applied to team evolution analysis mainly fall into two categories (foreign studies often use the concept of "community" to refer to various entities including "teams"): The first is to first divide the dataset into multiple time slices, identify communities in each slice, and then compare the similarity between communities in different slices to obtain community evolution paths. The second is evolutionary community discovery methods, which mine the evolution of the next stage based on community identification results from the previous time slice combined with relevant attributes of nodes and edges [21]. No research has confirmed which method is more effective. In terms of applicability, the first method can better handle unstable changing communities.

In summary, in-depth exploration of the evolutionary characteristics and patterns of research teams and the development of effective and feasible analytical methods is not only of important academic value but also of practical significance for supporting policy formulation. In existing research, analysis from a network topology perspective is relatively singular, mainly focusing on all identified teams and analyzing the overall co-authorship network' s topological indicator changes over time from a holistic perspective. Although this approach can reveal the evolution of some teams, different teams make different contributions to field development, making the "partial teams" unable to reflect targeted development of different teams. Meanwhile, increasing research shows that research teams have high instability during evolution [22-23]. Therefore, this paper takes high-yield research teams in the artificial intelligence field as the research object, combines team evolution paths from micro and macro dual perspectives, and analyzes the evolution of high-yield team topological structures by calculating the extreme value distributions of node count, edge count, network density, and average clustering coefficient topological indicators in team collaboration networks, aiming to explain the formation and internal evolution reasons of high-yield teams to a certain extent and promote the development of academic research and practical applications.

2 Research Design and Team Selection

Based on the research purpose and the above review, starting from data analysis, this study follows the research design shown in Figure 1 [Figure 1: see original paper].

First, in the data collection and cleaning stage, this paper selects the Web of Science Core Collection as the data source, with the search formula WC = "Computer Science, Artificial Intelligence", time limited to 2009-2018, retrieving a total of 421,148 records in January 2019. The author name disambiguation problem seriously affects the construction of the co-authorship network, and the quality of the co-authorship network directly determines the accuracy of team identification results. Therefore, we first complete the name disambiguation work by combining co-author and author institution information, then import the cleaned data into a MySQL database.

Then, we proceed to the research team identification and selection stage. After constructing the overall author co-authorship network, we select the Louvain community detection algorithm to identify research teams in the AI field. This step identifies 23,423 research teams, involving 656,668 authors. Publication volume has always been used as a traditional bibliometric indicator, but research on the dynamic evolution of high-publishing teams is still relatively rare. This paper defines the top 100 teams in terms of publication volume as prolific research teams and uses these teams as the research object for subsequent analysis.

Finally, we conduct topological measurement and evolution analysis. After obtaining the teams to be analyzed, we match the article sets published by team members, deduplicate the articles, and divide them into 10 time intervals according to publication year (2009-2018). In each interval, we obtain all authors corresponding to the articles and then construct the co-authorship network of a specific team within that interval. These networks are connected in series over time to form the team' s evolution path. For each co-authorship network within a time interval, we calculate four topological indicators: node count, edge count, network density, and network average clustering coefficient. The measurement of team topological structure is to understand the structural characteristics of research teams from a network analysis perspective, and analyzing the topological structure of co-authorship networks in each time slice will reveal the changes in network structure during team evolution. At the same time, it can also uncover the commonalities and particularities exhibited by prolific research teams during evolution to deepen understanding of team dynamic evolution. Based on this, this paper combines micro and macro perspectives to analyze the dynamic evolution of research teams. From a micro perspective, this paper explores the year distribution of extreme values of selected topological indicators. Among them, the extreme value distributions of point and edge counts directly reflect fluctuations in members and member relationships (co-authorship relationships), while network density and average clustering coefficient reflect the closeness of cooperation among members during team evolution, and their extreme value distributions can reflect fluctuations within high-yield teams to a certain extent. From a macro perspective, this paper analyzes the overall evolution of topological indicators over the ten-year span. Among them, changes in node count reflect changes in team size, changes in edge count reflect changes in team members' co-authorship relationships, evolution of network density represents the overall closeness of team cooperation, and changes in average clustering coefficient more specifically reflect the evolution of "small group" collaboration phenomena.

3 Microscopic Perspective Analysis of Prolific Research Team Topological Structure Evolution

3.1 Extreme Value Distribution of Node and Edge Counts

The year distribution of topological indicator extremes can intuitively reveal the fluctuation of corresponding indicators, and the underlying reasons for such fluctuations are the concrete manifestations of team evolution. Figures 2 [Figure 2: see original paper] and 3 [Figure 3: see original paper] show the distribution of node and edge count extremes during the evolution of selected teams, respectively. The two figures reflect a high degree of consistency: for most AI field teams, the minimum values of node and edge counts are distributed in the first five years and the final year, while maximum values are distributed in the middle and later stages of the research interval (2012-2017). This distribution indicates that in the initial years, teams had relatively few nodes and edges, but as time progressed, teams exhibited different evolutionary behaviors—some teams reached their maximum in the middle stage, indicating that node and edge counts would decrease in later stages; other teams reached their maximum in the later stage, meaning node and edge counts either continued to increase or fluctuated with increases and decreases.

Table 1 provides more specific numerical details. In 2010, 33 prolific teams reached minimum values for both node and edge counts, with 29 teams simultaneously reaching minimum values for both indicators in that year. It can be seen that after filtering, the identified prolific teams show highly synchronized extreme value distributions for nodes and edges. Combined with Figures 2 and 3, the minimum values are mainly distributed in 2009-2012 and 2018. Furthermore, Table 2 lists example teams from these years (showing the top 2 teams by publication volume each year). Teams with node minimum values in 2009 rank as high as 5th, while those in 2018 rank as high as 17th. It can be found that whether from the perspective of average ranking, highest ranking, or average publication volume, these teams do not show significant differences.

From the perspective of evolution cycles, different teams are at different "life stages," causing different distributions of topological indicators among teams. Among them, teams reaching minimum node and edge values in 2018 are experiencing a reduction in member count and weakening co-authorship intensity, which at the team level corresponds to team "decline." Figure 4 [Figure 4: see original paper] shows the evolution process of representative teams. The evolution characteristics of these teams are that small-group collaboration phenomena are quite obvious, with tightly connected "small groups" visible across multiple evolution windows. Ultimately, the balancing effect between "small groups" fails to converge the overall team, leading to the "decline" of the entire team in later stages. Correspondingly, another evolution pattern is represented by team #207, led by Jiao Licheng from Xidian University. The leader appears in all 10 time windows, and co-authorship networks centered on him can be seen in the early, middle, and late stages of evolution. Although small-group collaboration phenomena also exist, there are stable connections between "small groups" and the overall team, ultimately leading "small groups" and the overall team toward integration.

3.2 Extreme Value Distribution of Network Density and Average Clustering Coefficient

Contrary to the distribution of node and edge extremes, the minimum values of network density during team evolution are concentrated in the middle and later stages of the research interval (2014-2017), while maximum values are mainly distributed in the first five years and the final year. Combined with the distributions in Figures 2 and 3, some teams had fewer members in the initial years but more internal co-authorships, thus showing maximum network density concentrated in the early years, as shown in Figure 6 [Figure 6: see original paper]. Figure 7 [Figure 7: see original paper] shows the distribution of teams' average clustering coefficients, with maximum and minimum values distributed more evenly than the previous three indicators but mainly in the first five years. Most minimum values of average clustering coefficient fall between 0.4-0.85, while maximum values mainly fall between 0.8-1.0, reflecting that teams generally have high clustering degrees. Meanwhile, the uniform distribution of average clustering coefficient also reflects the existence of many small-group collaboration phenomena.

From a social network theory perspective, network density reflects the closeness of collaborative relationships within research teams. For teams in close collaborative relationships, later cooperative relationships tend to stabilize (as shown in the left side of Figure 6 [Figure 6: see original paper]). At this time, adding one member to the team may only establish a few cooperative relationships in practice, but the potential cooperative relationships increase tenfold, making the team' s network density more likely to reach minimum values in later stages. For the network average clustering coefficient, the "triad" small-group collaboration phenomenon is the most intuitive reflection. When cooperative relationships develop to a certain extent, there are more "closed triads," causing some teams' network average clustering coefficient to reach maximum values in later stages. The high publication volume of prolific teams means that the team itself has some relatively stable cooperative relationships during development, so some teams also reach maximum network average clustering coefficient in the middle and early stages (as shown in the right side of Figure 7 [Figure 7: see original paper]).

Similarly, using teams shown in Figures 4 and 5 as examples: Team #736 shown in Figure 4 reached its maximum network density in 2018 and minimum in 2017. In 2018, the team' s node count decreased sharply, leading to increased network density; while in 2017, the network node count was at a medium level and edge count at a relatively low level, reflecting lower cooperation among team members. For network average clustering coefficient, it reached minimum and maximum values in 2013 and 2018, respectively. Observing its evolution path diagram explains this extreme distribution—the proportion of "closed triads" in the team was relatively low in 2013, while in 2018, almost all points were in "closed triads." Team #207 shown in Figure 5 reached minimum and maximum network density values in 2010 and 2014, respectively. From the evolution diagram, the team showed a radial pattern centered on Jiao Licheng in 2010, but the radiating points were not closely connected, resulting in minimum network density. In 2014, the network presented a radial network centered on Jiao Licheng, Gong Maoguo, and Ma Wenping, with more edges connecting points at the radiation edges, bringing the maximum network density. For network average clustering coefficient, its minimum and maximum values were achieved in 2012 and 2011, respectively, at 0.8197 and 0.9222. Due to the consistently close internal cooperation within the team, the proportion of "closed triads" remained stable at a high level, making the clustering degree of the co-authorship network high during evolution.

4 Macroscopic Perspective Analysis of Prolific Research Team Topological Structure Evolution

4.1 Overall Evolution of Node and Edge Counts

Based on each team' s evolution across the four indicators, we attempt to mine the evolutionary commonalities of these prolific teams on certain indicators to help explain team dynamic evolution. Specifically, by measuring the topological indicators of teams in each time window, we can plot the changes of each indicator over time. Figure 8 [Figure 8: see original paper] shows the evolution of edge counts for all prolific teams over time. From the evolution shown, various teams do not exhibit obvious common characteristics, but it can be known that the edge counts of most teams show an increasing trend, reflecting that new cooperative relationships gradually increase during evolution. The evolution of node counts is similar, but due to space limitations, we will not discuss node count evolution further here.

4.2 Overall Evolution of Network Density and Average Clustering Coefficient

Notably, there are two obvious evolution types in the changes of average clustering coefficient and network density during team evolution, with Figure 9 [Figure 9: see original paper] showing examples of five teams. The two evolution types are specifically manifested as: Type I teams maintain a medium-to-high level of average clustering coefficient with stability, with teams numbered 1064, 205, and 203 in Figure 8 all belonging to this type; Type II teams have a low average clustering coefficient for 1-2 years, consistent with the first type for the remaining time, such as teams 342 and 1363 in Figure 9. Additionally, network density and average clustering coefficient basically show opposite trends during evolution: the density of most teams remains stable at a low level, with very few teams fluctuating to medium-high levels for 1-2 years. Moreover, the changes of these two indicators can maintain a high degree of synchronization—when network density fluctuates, the average clustering coefficient will fluctuate in the opposite direction within 1-2 years, and vice versa. Furthermore, the timing of fluctuations in these two indicators is basically distributed in the first five years, corresponding to the distribution of extreme values of team node and edge counts in the previous section.

Figures 10 [Figure 10: see original paper] and 11 [Figure 11: see original paper] select teams 205 and 342 as examples to illustrate these two types of evolution processes. Team 205 shows obvious "small group" collaboration phenomena across all 10 time windows, but cooperation between these "small groups" was not close in the first five years. After 2014, we can see that “small groups” began to develop close cooperative relationships. Moreover, 2014 corresponds exactly to the year when network density fluctuated for team 205 in Figure 9. The evolution process of team 342 reveals the growth process of this team: node and edge counts both grew slowly in the first five years, after which relatively stable internal cooperative relationships were formed, and these relationships were continuously consolidated, further developing “small group”phenomena until the last two years when various groups established more cooperative relationships.

5 Discussion and Conclusion

Through the above analysis, we first find that microscopic perspective analysis very intuitively reflects the dynamic attributes of the evolution of prolific research teams in the AI field. The extreme value distributions of node count, edge count, network density, and network average clustering coefficient in teams vary. Specifically, the minimum values of node and edge counts are distributed in the first five years and the final year (2009-2014 and 2018), the maximum values of node count are concentrated in 2015-2017, while the maximum values of edge count are evenly distributed in 2013-2017; the minimum values of network density are concentrated in 2014-2017, with maximum values distributed in 2009-2014 and 2018; the minimum values of average clustering coefficient are concentrated in 2009-2014, while maximum values are distributed across all years. Comparison shows that the minimum values of node count, edge count, and average clustering coefficient of prolific teams in this field are basically distributed in the same time period (2009-2014), which happens to be the distribution interval of maximum network density values.

Second, the macroscopic perspective analysis of the overall evolution of team topological indicators corroborates the microscopic analysis: the evolution of node and edge counts of AI field prolific research teams shows an increasing trend, with some teams decomposing in later stages, corresponding to a subsequent decrease in node and edge counts; the evolution of team network density and average clustering coefficient presents two patterns—one where both indicators remain at relatively stable levels, and another where when one indicator fluctuates, the other changes in the opposite direction in the following two years. Combining both analytical perspectives and team evolution paths, most AI field prolific research teams can continuously foster more new cooperative relationships—the average clustering coefficient indicator reveals that this cooperation shows very obvious “small group” collaboration phenomena, and cooperative relationships between “small groups” also tend to stabilize. Although these phenomena and their evolution require further in-depth explanation by domain experts combined with the characteristics of the AI field, preliminary results suggest that the domain characteristics of AI lead to two aspects of evolutionary behavior in prolific teams: first, the field is based on mathematics and computer theory and technology, solving interdisciplinary comprehensive problems, which greatly promotes the emergence of cooperative behavior. When a team can grow into a prolific team, the "small groups" within the team are either closely connected with the team leader, or "small groups" gradually move toward union; second, related technologies have considerable "circulation" between subfields. For example, feature extraction technology is applied in both computer vision tasks and natural language processing tasks, making "small groups" studying computer vision highly likely to establish connections with "small groups" studying natural language processing.

In summary, the main conclusions of this paper can be refined as follows: (1) Microscopic analysis reveals that the extreme value distribution of co-authorship network topological indicators reveals the dynamic attributes of prolific team evolution in the AI field. The extreme value distributions of node and edge counts in co-authorship networks of prolific teams are highly correlated, and this correlation distribution corroborates with the extreme value distributions of network density and network average clustering coefficient. The extreme value distributions of various indicators reveal that even the members and internal relationships of prolific teams are in considerable flux. (2) Macroscopic analysis shows that AI field prolific research teams exhibit certain commonalities in the evolution of network density and network average clustering coefficient, and most prolific research teams can continuously foster more new cooperative relationships—new cooperative relationships (new cooperation between existing members or cooperation brought by new members), which means that most teams can maintain some stable topological structures in the dynamic evolution of member count and member relationships. (3) Whether analyzed from microscopic or macroscopic perspectives, in-depth observation and analysis of team evolution paths can reveal significant "small group" collaboration phenomena in the evolution of AI field prolific research teams. Moreover, cooperative relationships between "small groups" directly affect the future direction of the overall team. When cooperation between "small groups" becomes closer, both the overall team' s network density and average clustering coefficient will increase; when "small groups" neglect cooperation, the entire team may even move toward a "decline" stage, with both member count and cooperative relationships weakening. (4) In addition to showing certain common patterns, AI field prolific research teams still exhibit considerable differences in evolutionary behavior among different teams, such as the specificity of node and edge count increase/decrease patterns, with different teams showing different change patterns.

Regarding future research prospects, we plan to proceed from two aspects: first, the evolutionary commonalities identified in this study need further investigation. For example, the increasing trend of edge and node counts needs to clarify whether the increase in edge count comes from cooperation between "new members" and "old members" or between "old members" themselves. Additionally, we can try to introduce new analytical perspectives or increase the number of prolific teams to further explore evolutionary commonalities. Second, the selected research objects are specifically targeted prolific research teams, and we can further analyze the performance of research teams selected by other indicators (such as highly-cited teams) to explore the generalizability of related operations and conclusions.

References

[1] VERBREE M, HORLINGS S E, GROENEWEGEN P, et al. Organizational factors influencing scholarly performance: a multivariate study of biomedical research groups[J]. Scientometrics, 2015, 102(1): 25-49.

[2] PERIANES-RODRÍGUEZ A, OLMEDA-GÓMEZ C, MOYA-ANEGÓN F. Detecting, identifying and visualizing research groups in co-authorship networks[J]. Scientometrics, 2010, 82(2): 307-319.

[3] VAN RAAN A F J. Scaling rules in the science system: influence of field-specific citation characteristics on the impact of research groups[J]. Journal of the American Society for Information Science and Technology, 2008, 64(7): 1852-1863.

[4] YU Q, SHAO H, DUAN Z. Research groups of oncology co-authorship network in China[J]. Scientometrics, 2011, 89(2): 553-567.

[5] REY-ROCHA J, GARZÓN-GARCÍA B, MARTÍN-SEMPERE M J. Scientists' performance and consolidation of research teams in biology and biomedicine at the Spanish Council for Scientific Research[J]. Scientometrics, 2006, 69(2): 183-212.

[6] MARTÍN-SEMPERE M J, GARZÓN-GARCÍA B, REY-ROCHA J. Team consolidation, social integration and scientists' research performance: an empirical study in the Biology and Biomedicine field[J]. Scientometrics, 2008, 76(3): 457-482.

[7] MORRIS S A, GOLDSTEIN M L. Manifestation of research teams in journal literature: a growth model of papers, authors, collaboration, co-authorship, weak ties, and Lotka' s law[J]. Journal of the American Society for Information Science and Technology, 2007, 64(7): 1852-1863.

[8] AZAOUZI M, RHOUMADI, ROMDHANE L B. Community detection in large-scale social networks: state-of-the-art and future directions[J]. Social network analysis & mining, 2019, 9(1): 23.

[9] BÖRNER K, CONTRACTOR N, FALK-KRZESINSKI H J, et al. A multi-level systems perspective for the science of team science[J]. Science translational medicine, 2010, 2(49): 49cm24-49cm24.

[10] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12): 7821-7826.

[11] NEWMAN M E J. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(23): 8577-8582.

[12] ASUR S, PARTHASARATHY S, UCAR D. An event-based framework for characterizing the evolutionary behavior of interaction graphs[C]//ACM SIGKDD international conference on knowledge discovery & data mining. New York: ACM, 2007: 913-921.

[13] MASMOUDI A, MEZGHANI E, BELLAAJ H, et al. A web-based knowledge management system for scientific research team[C]//2011 IEEE

26th international conference on enabling technologies: infrastructure for collaborative enterprises. Poznan: IEEE, 2011: 94-101.
[14] OWENS D A, MANNIX E A, NEALE M A. Strategic formation of groups: issues in task performance and team member selection[J]. Research on managing groups and teams: composition, 1998, 1(1998): 149-165.
[15] JOHNSON N F, XU C, ZHAO Z, et al. Human group formation in online guilds and offline gangs driven by a common team dynamic[J]. Physical review E, 2009, 79(6): 1-11.
[16] CEBALLOS H G, GARZA S E, CANTU F J. Factors influencing the formation of intra-institutional formal research groups: group prediction from collaboration, organizational, and topical networks[J]. Scientometrics, 2018, 114(1): 181-216.
[17] MILOJEVIC S. Principles of scientific research team formation and evolution[J]. Proceedings of the National Academy of Sciences of the United States of America, 2014, 111(11): 3984-3989.
[18] TAKAFFOLI M, SANGI F, FAGNAN J, et al. MODEC—modeling and detecting evolutions of communities[C]//Fifth international AAAI conference on weblogs and social media. California: The AAAI Press, 2011: 626-629.
[19] WANG Yuefen, LI Dongqiong, YU Houqiang. Research on the evolution of scientific collaboration networks and the growth characteristics of highly influential scholars in life cycle stages[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(2): 121-131.
[20] BARABÁSI A L, JEONG H, NÉDA Z, et al. Evolution of the social network of scientific collaborations[J]. Physica A: Statistical mechanics and its applications, 2002, 311(3-4): 590-614.
[21] TAKAFFOLI M, FAGNAN J, SANGI F, et al. Tracking changes in dynamic information networks[C]//International conference on computational aspects of social networks. Piscataway: IEEE, 2011: 3-10.
[22] WANG Y, WU B, DU N. Community evolution of social network: feature, algorithm and model[EB/OL]. [2020-03-31]. http://arxiv.org/abs/0804.4356.
[23] HOPCROFT J, KHAN O, KULIS B, et al. Tracking evolving communities in large linked networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(SUPPL. 1): 5249-5253.
[24] PALLA G, BARABÁSI A L, VICSEK T. Quantifying social group evolution[J]. Nature, 2007, 446(7136): 664-667.
[25] OSBORNE F, SCAVO G, MOTTA E. A hybrid semantic approach to building dynamic maps of research communities[C]//International conference on knowledge engineering and knowledge management. Cham: Springer, 2014: 356-372.
[26] LIU Xianhong, LI Gang. Analysis of cooperation stability of research teams during continuous funding by the National Natural Science Foundation of China[J]. Bulletin of National Natural Science Foundation of China, 2016, 30(4): 372-378.
[27] GUO Shiyue, LI Rui. Research on the correlation between co-authorship network patterns and paper output quantity and quality—taking the co-authorship network of Chinese catalysis field journal papers as an example[J].

Journal of Intelligence, 2013, 32(3): 314-327.
[28] BACKSTROM L, HUTTENLOCHER D, KLEINBERG J. Group formation in large social networks: membership, growth, and evolution[C]//Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2006: 44-54.

Author Contributions

Zou Bentao: Proposed the research idea, responsible for literature review, data analysis, and paper writing;
Wang Yuefen: Proposed the paper idea, guided paper revision and finalization;
Yu Houqiang: Participated in discussing research ideas and proposed revision suggestions.

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