

## Call the shots or follow suit? Research on the Evolution of Artificial Intelligence Scientific Collaboration Networks from a Leadership-Participation Perspective

Chaocheng He <sup>1</sup>  
Jiaxin Du <sup>2</sup>  
Jun Zhang <sup>3,\*</sup>

\*Corresponding author

<sup>1</sup> School of Information Management, Wuhan University, Wuhan, China, he\_chaocheng@whu.edu.cn

<sup>2</sup> School of Information Management, Wuhan University, Wuhan, China, jiaxin\_du@whu.edu.cn

<sup>3</sup> School of Information Management, Wuhan University, Wuhan, China, is\_junzhang@whu.edu.cn

### ABSTRACT

Scientific research cooperation has become a key driver of technological innovation, knowledge dissemination, and social progress, enhancing the integration of different disciplines and improving the quality and quantity of research outcomes globally. Artificial intelligence (AI), characterized by globalization, interdisciplinarity, and deep industry-academia-research integration, advances technological progress. However, traditional models overlook the leadership-participation dynamic in research collaboration. This paper uses the Exponential Random Graph Model (ERGM) to analyze the attributes and structures of network nodes, revealing that reciprocity, transfer, preferential attachment, and homophily mechanisms are crucial for the formation and evolution of research collaboration networks, providing insights for policymakers and research teams.

**Keywords:** Research collaborations, artificial intelligence, network evolution, network polarization.

### INTRODUCTION

In recent decades, artificial intelligence (AI) has achieved rapid development, with its influence and applications continuously expanding. Since the late 20th century, the number of AI-related research papers has shown exponential growth, particularly in the past five years, reflecting a significant increase in both the breadth and depth of research in this field worldwide. Platforms like Scopus and Web of Science indicate that AI research papers have grown at an annual rate exceeding 20%, far outpacing other scientific fields, highlighting the global enthusiasm and booming period for AI research (Clark & Llorens, 2012; González-Alcaide et al., 2017; He et al., 2023; Jiang et al., 2018).

In the era of globalization, informatization, and the knowledge economy, scientific research has become increasingly complex and specialized, demanding deeper and broader knowledge (Carayol et al., 2019); (Schummer, 2004). Consequently, individual scientists or single researchers find it challenging to achieve breakthroughs alone. In response, scientific cooperation has become a vital force in promoting technological innovation, knowledge dissemination, and social progress. Collaboration not only integrates knowledge from different disciplines but also significantly enhances the quantity and quality of research outcomes globally, becoming the mainstream mode of scientific research. Collaborative efforts pool researchers with diverse backgrounds, skills, and resources, fostering integrative innovation and advancing science and technology (He et al., 2021). At the individual level, the diversity, stability, and intensity of collaborations directly correlate with research output, enhancing the quality and quantity of scientific results. At the institutional level, especially among universities, research cooperation has been widely shown to positively correlate with research output. Closer partnerships typically lead to more academic papers, patent applications, and research projects due to resource complementarity, knowledge sharing, and technological exchange, which help overcome the limitations of single institutions and accelerate innovation. At the national level, international academic cooperation, through funding, shared goals, interaction, dependencies, and knowledge flow, aids in addressing complex research problems.

However, research cooperation involves more than task aggregation; it encompasses complex knowledge exchange, resource allocation, and division of responsibilities, where the formation and role of research leadership are crucial. Research leadership reflects the relational structure, knowledge flow patterns, and resource distribution strategies between leading and participating institutions, significantly impacting the success of cooperation, the influence of research outcomes, and the evolution of cooperation networks. The leadership-participation perspective offers a valuable analytical framework for understanding the roles and statuses of different entities in AI research collaboration networks. Institutions, renowned experts, or countries like the United States and China play leading roles in these networks, with their research outputs garnering significant international attention (Jiang et al., 2012). Citation analysis shows that collaborative AI papers are more frequently cited, indicating higher influence and credibility. AI's applications span various fields, including healthcare, finance, and engineering. Interdisciplinary collaboration has led to a rapid increase in cross-disciplinary papers, facilitating knowledge, technology, and resource exchange, thereby accelerating AI's application and innovation in different domains. This cross-field cooperation drives AI's integration into various sectors, offering new possibilities and innovative solutions.

Research collaboration networks, as complex social networks, have structural characteristics (e.g., network density, centrality, clustering coefficient) that significantly influence research efficiency and innovation. Research shows these networks exhibit high concentration and polarization, with significant differences in research leadership among institutions, increasing reciprocity and decentralization. Geographic distance negatively impacts the diffusion of research leadership, while cognitive similarity and social proximity promote its formation. The evolution of these networks is influenced by research entities' attributes (e.g., research capacity, income level) and external environments (e.g., institutional and cultural factors). Identifying and analyzing key nodes (e.g., institutions with high research leadership) in research collaboration networks is crucial for understanding network structure, knowledge flow, and research impact. Methods like exponential random graph models and social network analysis help identify and assess these key nodes' influence, revealing core structures, knowledge pathways, and the roles and statuses within the network (Klavans & Boyack, 2010; Zaccaro et al., 2001).

In summary, research cooperation is a core feature of modern scientific activities, with research leadership, proximity factors, network structure, and evolution forming current research focuses. Exploring these aspects helps reveal the intrinsic rules of research collaboration, optimize resource allocation, and improve cooperation efficiency, contributing to technological innovation, knowledge dissemination, and social progress. Future research should deepen theoretical construction and empirical testing of research leadership, enhance multidimensional proximity analysis frameworks, and explore dynamic evolution mechanisms of collaboration networks, advancing both theoretical and practical applications of research cooperation.

### LITERATURE REVIEW

Scientific cooperation integrates diverse knowledge and skills to tackle complex challenges, acting as a catalyst for groundbreaking innovations. It enhances innovation efficiency through refined collaboration systems. Interdisciplinary collaboration, now a key trend in both natural and social sciences, expands knowledge boundaries and accelerates discoveries. The role of core leaders, such as corresponding or first authors, is crucial in guiding projects from conception to completion, ensuring close interactions with team members. This leadership framework supports the main structure of research collaboration, with corresponding authors often identified as the leaders in natural sciences and a balanced consideration of first and corresponding authors in social sciences.

Research has found that scientific collaboration networks are highly centralized and polarized, with a few leading institutions (corresponding authors' institutions) holding significant research dominance, while participating institutions (non-corresponding authors' institutions) have lower but rapidly growing research status. These networks show dynamic evolution and power restructuring, featuring strong reciprocal relationships and transitive effects. Geographical distance hinders the spread of research dominance, while cognitive similarity and social ties facilitate its formation (Newman, 2001; Sekara et al., 2018). Social Network Analysis (SNA) reveals that these networks often exhibit small-world and scale-free properties, with core scientists playing pivotal roles in innovation and leadership, supported by close-knit community structures (Kaltenberg et al., 2023).

Research collaboration in scientific networks exhibits inherent dynamics driving the spontaneous formation and evolution of cooperative relationships (Dagnino et al., 2016). These dynamics can be summarized in four main aspects: Firstly, the complementary need for knowledge and skills motivates individuals and teams to share specialized domains, facilities, and data, thereby enhancing project success rates and forming the foundation of collaborative network construction and evolution. Secondly, based on reciprocity, parties exchange knowledge, technology, and resources, achieving mutual benefits and allowing those with strong research leadership to gain advantageous positions within the collaboration. Thirdly, stable research collaborations rely on trust among members and adherence to shared norms, skills that individuals or teams with robust research leadership excel in fostering. This ensures the stability and efficiency of collaborative networks. Lastly, social capital, including reputation and networks, plays a crucial role in research communities, enabling efficient information acquisition, resource mobilization, and coordination for entities with strong research leadership. Studies indicate that reciprocity and transmission effects positively influence the establishment of research leadership, fostering cooperative relationships and further collaborations between institutions (Chinchilla-Rodríguez et al., 2019).

Research collaboration's exogenous attributes and research leadership are critical factors shaping the development and efficiency of scientific cooperation systems (Stimson et al., 2009). These factors interact to define the structure, function, and evolution of research collaboration networks. Exogenous attributes, such as policy environments, economic and social development needs, technological infrastructure, and legal frameworks, influence collaboration patterns and network structures significantly. For instance, government policies supporting specific research domains or encouraging interdisciplinary and inter-institutional cooperation play a pivotal role in network dynamics. Strong research leadership drives innovation and network development by adapting to and leveraging these external factors. Understanding this interaction is key to optimizing research environments and enhancing scientific innovation efficiency (Sutanto et al., 2021).

The leadership-participation relationship in research collaboration plays a crucial role throughout its lifecycle (Lester & Kezar, 2017). While existing research has extensively explored and achieved a comprehensive understanding of the field of research collaboration, there remain opportunities to uncover and address research gaps.

Firstly, descriptions of research collaboration relationships largely remain qualitative, highlighting a notable gap in using quantitative methods to depict complex networks in research collaboration (Walther et al., 2017). The leadership-participation relationship in research collaboration is not merely a concept but holds rich connotations and extensions. Quantitative depiction through complex network analysis enriches both the physical and social dimensions of research collaboration, exploring its evolutionary mechanisms and influencing factors. Secondly, comprehensive studies have examined the mechanisms, attributes, and evolution of research collaboration networks within complex network research (Abbasi et al., 2011). However, the growth mechanism of edges in research collaboration networks remains relatively ambiguous. From the perspective of leadership-participation, identifying who dominates and who participates in this directional network differs fundamentally from undirected networks (Chen et al., 2019). Understanding which node attributes of leaders in research collaboration impact network growth is pivotal. Thus, investigating the intrinsic mechanisms and extrinsic attributes of research collaboration networks from both macro and micro perspectives is crucial for deeper insights (Powell et al., 2005).

## RESEARCH METHODS AND MODELS

The study on the evolution of AI research collaboration networks from a leadership-participation perspective aims to achieve several key objectives: analyzing the topological and spatial characteristics of research collaboration networks and their evolution, and further investigating their evolutionary mechanisms, including the influence mechanisms of network link strength and the growth mechanisms of network links (Balahurovska, 2023).

In delving into the characteristics and evolutionary mechanisms of research collaboration networks, we employ a comparative analysis method of network topological features. From a macro perspective, we describe in detail the key characteristics of the network, such as the number of nodes, edges, network density, average path length, average clustering coefficient, network reciprocity, and network entropy. These indicators provide a comprehensive understanding of the overall structure and dynamic changes of the network. On a micro level, we examine the attributes of network nodes, including degree, weighted out-degree, betweenness centrality, and closeness centrality, to more precisely depict the importance and influence of individual nodes within the network, thereby revealing the intrinsic rules of network evolution. Through this comprehensive, multi-level analysis approach, we can better understand the complexity and dynamics of research collaboration networks.

The order of coauthors represents their contributions to the work. In this article, we extend the harmonic scheme (Xia et al., 2014) to measure the coauthor link weight by considering the leadership-participation relationship. To be specific, we consider the coauthor leadership-participation relationship. Second, we assign equal credits to all the leading authors. We set the first author (s) and the corresponding author (s) as the leader (Zhou et al., 2017).

The edge growth mechanism of research collaboration networks is influenced by various factors, including homophily, preferential attachment, transitivity, and reciprocity. This study uses the Exponential Random Graph Model (ERGM) to simulate these mechanisms, testing and comparatively analyzing their roles in network evolution, thereby gaining deeper insights into the dynamic patterns of research collaboration.

The basic research approach of this study is grounded in the evolution of AI research collaboration networks from a leadership-participation perspective. Utilizing scientific paper data as the dataset, methods such as social network analysis and ERGM are employed to explore the edge growth mechanisms of research collaboration networks.

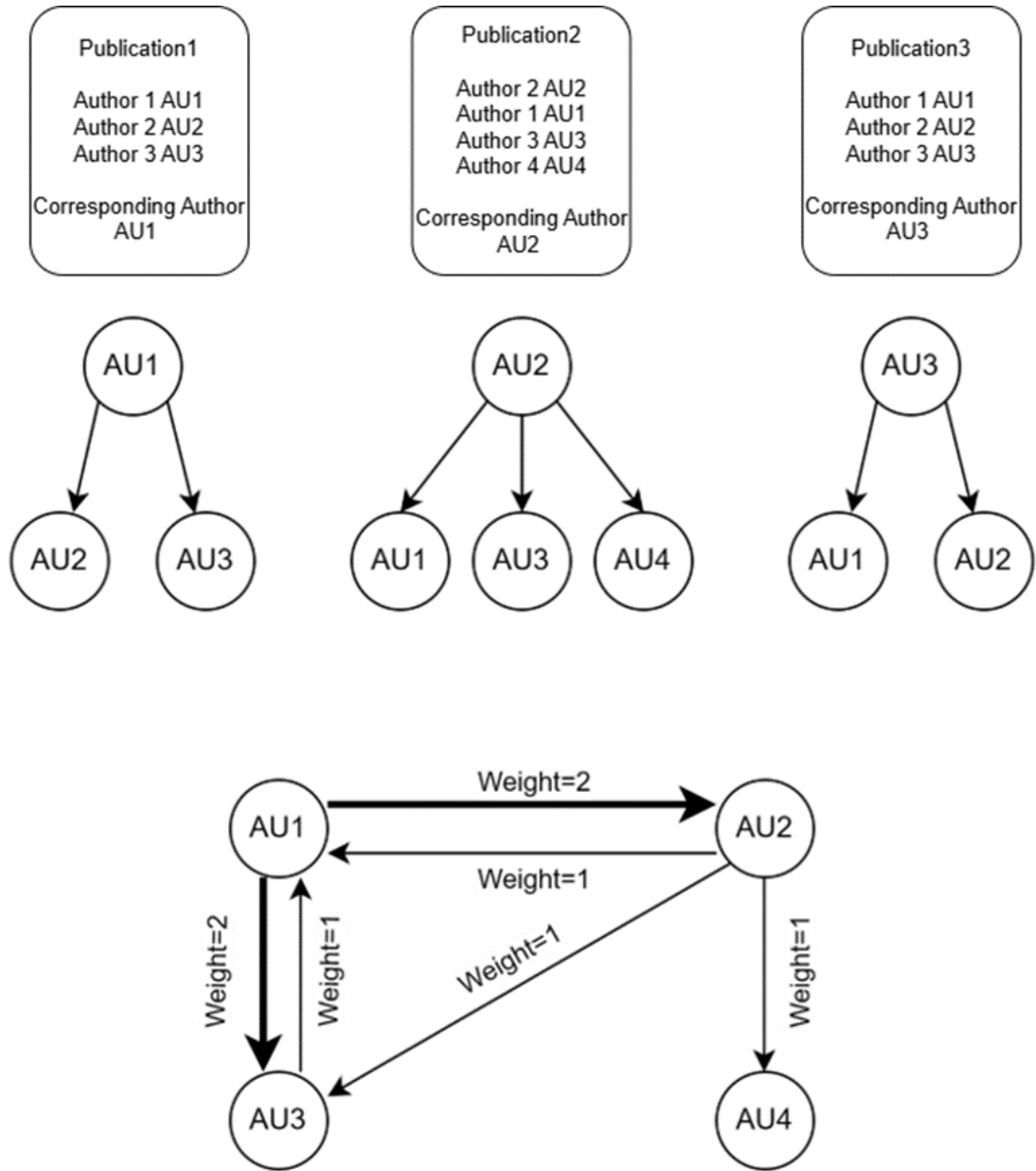


Figure 1: Scientific research cooperation network model

### RESEARCH DESIGN

This study's data is sourced from the Web of Science TM Core Collection, ensuring continuity and reliability of the research metrics. We collected data from the past five years with the search query: TS= (artificial intelligence) AND PY= (2020-2024), filtered for ARTICLE type, SCI-EXPANDED index, and Web of Science category: Computer Science Information Systems. As of March 2024, we retrieved 7,518 relevant articles with a total of 74,288 citations, averaging 9.88 citations per article, and an h-index of 97, indicating 97 papers were cited at least 97 times. Following(He et al., 2023), we removed entries with data quality issues or incomplete data, resulting in a research collaboration network model comprising 7,501 nodes.

When modeling the research collaboration network using Exponential Random Graph Models (ERGM), we treat the edge  $Y_{ij}$  between nodes  $i$  and  $j$  as a random variable. ERGM is a statistical model that explains the probability of edge formation through a set of network characteristic parameters, revealing the underlying rules of network structure generation and evolution. This probabilistic model integrates both the exogenous attributes of network nodes and the endogenous structural features of the network. Specifically, ERGM is generally expressed in the following form:

$$P(Y = y|X = x) = \frac{\exp\{\sum_{i=1}^1 \theta_i^T g_i(y) + \sum_{a=1}^A \theta_a^T g_a(y, x)\}}{\kappa} \quad (1)$$

The normalizing constant  $\kappa$  ensures that the model outputs a valid probability distribution, meaning that the sum of the probabilities of all possible network configurations equals 1.  $X$  represents the attribute matrix of nodes and edges.  $g_i(y)$  denotes

the statistics that influence network structure and internal relationships, while  $g_a(y, x)$  represents the statistics related to node attributes and the interactions between nodes. The values of  $g_i(y)$  are as follows:

$$g_i(y) = \begin{cases} 1, & \text{If the statistic } g_i(y) \text{ exists in the network} \\ 0, & \text{If the statistic } g_i(y) \text{ does not exist in the network} \end{cases} \quad (2)$$

The values of  $g_a(y, x)$  are similar to those of  $g_i(y)$ .  $I$  and  $A$  represent the number of network structure statistics and node attribute statistics, respectively.  $\theta_i^T$  and  $\theta_a^T$  are the parameter vectors for  $g_i(y)$  and  $g_a(y, x)$  statistics. Exponential Random Graph Models (ERGM) use Monte Carlo-Markov Chain Maximum Likelihood Estimation (MCMC-MLE) to accurately estimate the model parameters, while the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are employed to evaluate the model's fit. As iterations progress, the decreasing values of AIC and BIC indicate that the model increasingly aligns with the observed network structure, thus effectively enabling statistical inference and in-depth analysis of complex network structures.

In the analysis of the results,  $\theta_i^T$  and  $\theta_a^T$  are the focal points. According to the model's definition and implications,  $\theta_i^T$  and  $\theta_a^T$  explain the contribution of endogenous network structure features and exogenous node attribute features to the network's formation and evolution. Specifically, an increase of one unit in  $g_i(y)$  increases the log-odds of network formation by  $\theta_i^T$ , and an increase of one unit in  $g_a(y, x)$  increases the log-odds of network formation by  $\theta_a^T$ . Therefore, the magnitude and sign of  $\theta_i^T$  and  $\theta_a^T$  indicate the influence and direction of the corresponding factors on network formation and evolution.

### HYPOTHESIS DEVELOPMENT

Reciprocity mechanisms specifically apply to directed networks. If node  $i$  connects to node  $j$ , then node  $j$  is likely to connect back to node  $i$  in the future. Reciprocity, a fundamental rule of social interaction, involves the exchange of information or resources in directed networks. When an individual receives information, support, or resources, they are inclined to reciprocate similarly in the future. This principle is widely recognized as a core mechanism of social network interaction and can be summarized as "give and take": individuals feel compelled to return favors to maintain or strengthen relationships.

In directed networks, reciprocity is indicated by the high proportion of bidirectional edges (i.e.,  $A$  points to  $B$  and  $B$  points to  $A$ ). These bidirectional connections reflect the frequency of reciprocal interactions within the network, serving as a quantifiable measure of reciprocity. When individuals receive attention and support, they are more likely to reciprocate by sharing information, fostering a positive feedback loop.

In research collaboration networks, a similar dynamic can be observed. The leadership-participation relationships often involve one party leading in a specific area or project, with the other party following, learning, or collaborating. If individual  $i$  demonstrates leadership towards individual  $j$  (e.g., through technical guidance or knowledge transfer), over time,  $j$  might reciprocate by taking the lead in another context. This reciprocal leadership exemplifies the reciprocity mechanism within research collaboration networks, promoting information flow, knowledge sharing, and collaborative innovation. Based on established reciprocity theories and its potential impact on the dynamic shift of leadership-participation relationships in research collaborations, this study proposes the following hypothesis:

H1. The reciprocity mechanism plays a crucial role in the formation and evolution of edges in research collaboration networks.

If node  $i$  connects to node  $j$  and node  $j$  connects to node  $k$ , then node  $i$  is likely to connect to node  $k$  in the future. The transitivity mechanism is a core feature of social networks, describing a phenomenon where two nodes with a common neighbor are more likely to form a direct connection. This transitive relationship indicates network sensitivity, where small local changes (such as a change in the state of a node) can trigger a chain reaction, similar to the "butterfly effect," potentially leading to a global adjustment of the entire network structure.

Transitivity in social networks is crucial for rapid information dissemination. It facilitates the spread of information along existing connection paths, enabling new knowledge, ideas, or messages to efficiently penetrate all corners of the network. Research collaboration networks are also significantly influenced by the transitivity mechanism. Studies have shown that in academic journal collaboration networks, the transitivity mechanism is more pronounced compared to conference paper collaboration networks, indicating that research collaborations tend to develop along existing collaboration chains.

In research collaboration networks, due to the directional nature of edges, the transitivity mechanism manifests as follows: if entity  $i$  exerts a leading influence on  $j$ , and  $j$  exerts a leading influence on  $k$ , then the likelihood of  $i$  exerting a leading influence on  $k$  in the future increases. This indirect transmission of leadership reflects the step-by-step diffusion of innovation leadership within the network. Based on the above theoretical discussion and empirical evidence, this study proposes the following hypothesis:

H2. The transitivity mechanism plays a crucial role in the formation and evolution of edges in research collaboration networks.

If a node has many connections, other nodes are more likely to connect with it in the future. According to social capital theory, entities with higher social capital are more likely to benefit from it over time. Social capital, as an intangible asset, facilitates information exchange and resource sharing among individuals or organizations. In social networks, the number of followers is often seen as an indicator of social capital: users with many followers have greater influence and can attract more attention and interactions. Thus, other users prefer to connect with those who have a large following, demonstrating the "preferential attachment" phenomenon.

Similarly, in research collaboration, a subject's past cooperation experience, especially the number of established collaborations, is considered part of its social capital. Studies have shown that the number of new collaborators an innovator attracts is positively correlated with its current network size, indicating a preference for partnering with well-connected innovators, again reflecting "preferential attachment."

Applying this to research collaboration networks, the frequency of a subject's leadership-participation interactions can be seen as its social capital. This social capital may influence whether other subjects choose to establish leadership-participation relationships with it. Specifically, a subject with a history of high influence—having led many other subjects—is more likely to attract new collaborators who perceive its past leadership as a sign of future potential. Thus, this study hypothesizes:

H3. The preferential attachment mechanism significantly promotes the formation and evolution of edges in research collaboration networks.

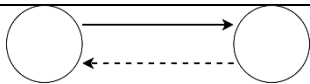
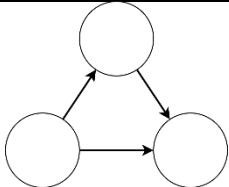
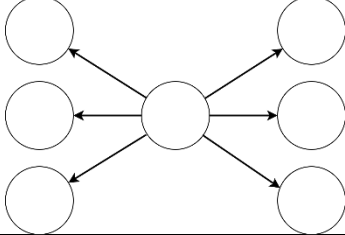
Nodes in a network tend to connect with others that have similar attributes. The homophily mechanism, a fundamental principle driving the development of complex networks, is encapsulated by the adage "birds of a feather flock together." McPherson and Smith-Lovin (2001) defined homophily as the tendency for individuals with similar characteristics to form connections more frequently than those with differing traits. This mechanism promotes consistent behavior among similar individuals, reducing communication barriers and naturally attracting them to one another.

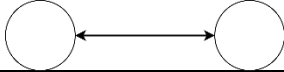
Homophily is evident in various contexts. In diverse racial environments, individuals form closer trust relationships with those of the same race, more likely sharing personal feelings with them. Similarly, in the field of technological innovation, subjects with similar professional backgrounds or experiences are more likely to form collaborative alliances. This similarity reduces initial information asymmetry and cognitive friction, enhancing mutual understanding in goal setting and strategy execution.

In research collaboration networks, the homophily principle remains influential. Innovators with similar backgrounds are more likely to connect, forming networks rich in shared ideas, complementary resources, and collaborative innovation. These connections result from the natural aggregation of similar forces and are effective pathways for spreading scientific progress and leadership. Overall, the homophily mechanism reduces interaction costs, strengthens behavioral consistency, and fosters strong intra-group connections, significantly impacting the structure and dynamics of social and technological networks. Based on this understanding, the hypothesis is proposed:

H4. The homophily mechanism significantly promotes the formation and evolution of edges in research collaboration networks.

Table 1: ERGM Statistical Terms and Assumptions

| Statistical items    | Corresponding statistical terms | Structure                                                                            | Hypothesis |
|----------------------|---------------------------------|--------------------------------------------------------------------------------------|------------|
| Reciprocity          | Mutual                          |  | H1         |
| Transitivity         | Transitive triple closure       |  | H2         |
| Priority linkability | Preferential attachment         |  | H3         |

|             |           |                                                                                    |    |
|-------------|-----------|------------------------------------------------------------------------------------|----|
| Homogeneity | Homophily |  | H4 |
|-------------|-----------|------------------------------------------------------------------------------------|----|

ANALYSIS OF RESULTS

This study constructed three types of ERGM (network structure model, node attribute model, and comprehensive model) to analyze the impact of endogenous network structure characteristics and exogenous node attributes on the formation and evolution of research collaboration network ties. The network structure model includes endogenous features (such as node degree, clustering coefficient, community structure) to explore how these inherent characteristics shape the formation and evolution of network ties. The node attribute model incorporates exogenous node attributes (such as node type, size, innovation capability, geographic location) within the null model framework to reveal how individual node characteristics affect the network ties' formation and evolution. The comprehensive model considers the interaction of both factors. The models were implemented using the R package statnet, with AIC and BIC values used to evaluate model fit; a smaller AIC indicates a better fit. Comparing the AIC and BIC values of the models helps identify the relative importance of network structure and node attributes, both individually and interactively, in the formation and evolution of research collaboration network ties.

Finally, we drew this diagram, and the number of nodes is gradually increasing, and the scale of the network also reflects a rapid development growth rate. Meanwhile, the research cooperation network in the field of artificial intelligence has undergone a clear differentiation process during the period.

Specifically, the number of nodes and edges is increasing year by year. The network nodes in 2020 are scattered and the edges are sparsely distributed. This indicates a relatively low number of collaborations and inadequate leadership relationships. The prominent change in the 2021 network compared to the 2020 network is the increase in core nodes. This shows that scientific collaborations are more frequent. The distinguishing feature of the 2022 network is the emergence of obvious aggregation networks, and the small world network has become denser. This shows that the frequency of leadership-participation relationships in some groups has been significantly improved. Compared with the networks in previous periods, a significant feature of the network in 2023 is the sharp increase in the number of network connections and the stratification of the network. The network changes from its original sparse and dispersed state to a closely connected state. Most nodes are in the largest connected component of the network and are closely connected to each other. The other connected components in the periphery are relatively free. This shows that most researchers have been integrated into the collaboration. In 2024, the layering phenomenon of the network will further deepen. The network is growing rapidly in the number of nodes and edges. This shows that more and more scientific researchers are forming collaborations.

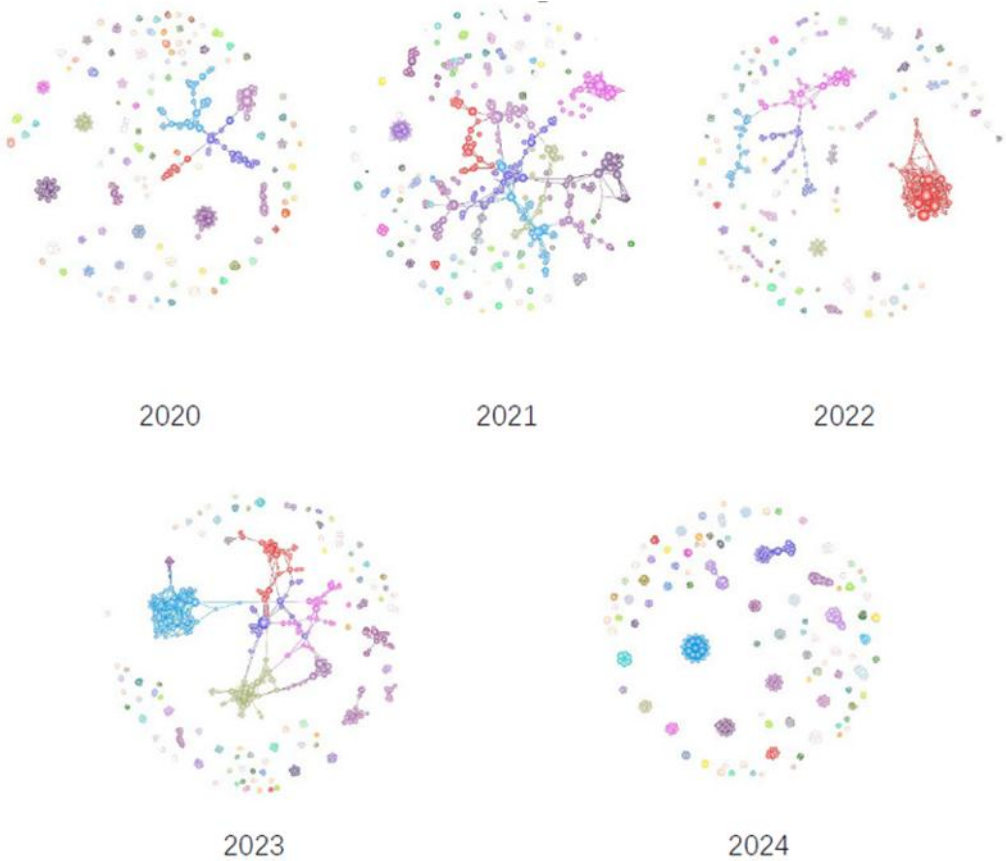


Figure 2: 2020-2024 Scientific research collaboration network diagram

Table 2: ERGM parameter estimation results of scientific research cooperation network

| Attributes                | Hypothesis | Model (1)          | Model (2)           | Model (3)          |
|---------------------------|------------|--------------------|---------------------|--------------------|
| Edges                     |            | -4.58***<br>(0.26) | -10.51***<br>(0.16) | -9.89***<br>(0.21) |
| Mutual                    | H1         | 3.15***<br>(-0.25) |                     | 2.28***<br>(0.44)  |
| Transitive triple closure | H2         | 1.96***<br>(0.40)  |                     | 1.42***<br>(0.28)  |
| Preferential attachment   | H3         | 0.81**<br>(0.07)   |                     | 0.59**<br>(0.06)   |
| Homophily                 | H4         |                    | 2.55**<br>(0.29)    | 1.87**<br>(0.16)   |
| AIC                       |            | 962                | 845                 | 769                |
| BIC                       |            | 973                | 858                 | 707                |

\*\*\*p<0.01;\*\*p<0.05;\*p<0.1

The study found that in Models 2 and 3, the estimated coefficients for the mutual effect (Mutual) were 2.55 and 1.87, both significant at the 0.01 level. This confirms the positive impact of the mutual effect on constructing research collaboration leadership-participation relationships. This means that when a lead author *i* has a leading role over partner author *j*, *j* tends to reciprocate, resulting in a close mutual leadership-participation relationship that enhances collaboration depth and stability. Thus, H1 is validated.

Regarding the transitive effect (Transitive triple closure), its coefficients in Models 1 and 3 were 1.96 and 1.42, significant at the 0.01 level. The results show a significant transitive effect in research collaboration networks: if *i* leads *j* and *j* leads *k*, then *i* is more likely to lead *k*. In this process, *i* acts as the knowledge source, and *j* as the intermediary, creating a low-risk, low-cost collaboration path from *i* to *k*. Therefore, H2 is validated.

The mutual and transitive effects have higher estimated coefficients than other variables, confirming them as key drivers in shaping leadership-participation relationships in research collaborations. This indicates that leadership-participation relationships are more likely to form between authors with direct or indirect connections, making it harder for new relationships to form between previously unconnected authors. Additionally, authors already in lead positions are more likely to secure new leadership-participation opportunities, further solidifying their network positions. These results show that once an author integrates into the research collaboration network and forms specific paths, they tend to maintain and strengthen these paths due to path dependence and inertia, thereby shaping the network structure and exacerbating its polarization.

In models 1 and 3, the estimated coefficients for Preferential attachment are 0.81 and 0.59 respectively, both significant at the 0.05 level, revealing the preferential attachment effect in international research collaboration networks. This suggests that lead authors in research projects prefer to collaborate with participating authors who have more research outputs, while the likelihood of collaborating with authors with fewer research outputs is lower. This reflects the "Matthew effect" in research: authors with more research outputs not only enjoy more collaboration opportunities but also enhance their research capabilities through these collaborations. Conversely, authors with fewer research outputs struggle to improve their research strength through collaborations and may lose opportunities for research collaboration, thereby widening the gap with high-output authors. Thus, H3 is confirmed.

Homophily has estimated values of 2.55 in model 2 and 1.87 in model 3, significant at the 0.05 level, indicating that homophily promotes the formation of research collaboration relationships. This suggests that higher homogeneity based on existing leadership-participation relationships facilitates the formation of mutual leadership-participation relationships. Thus, H4 is confirmed.

However, the estimated coefficients for Preferential attachment and Homophily did not significantly increase, confirming them as other important drivers shaping leadership-participation relationships in research collaborations. This indicates that the influence of Preferential attachment and Homophily on the formation of leadership-participation relationships is limited. This may be because collaborations in the field of artificial intelligence are not restricted to existing achievements and have high

innovation potential, attracting authors from various fields. This also reflects the breadth and depth of artificial intelligence research collaboration networks.

## SUMMARY AND DISCUSSION

Studying leadership-participation relationships is crucial for understanding research collaboration networks. Unveiling their mechanisms helps in comprehensively understanding research cooperation. This study focuses on directed leadership-participation relationships and constructs a research collaboration network based on these relationships. We analyze the formation mechanisms from the perspectives of exogenous attributes and endogenous mechanisms.

We provide empirical evidence that leadership-participation relationships are reciprocal and based on local hierarchies. If researcher *i* is led by *j*, *i* is likely to lead *j* in another collaboration, facilitating the exchange of academic resources. Additionally, if *iii* has led *j*, and *j* has led *k*, *i* is more likely to lead *k* in the future.

Lead authors are typically senior researchers, while participating authors are junior researchers. This hierarchy shows a significant preferential attachment effect, where participating researchers form relationships with highly connected lead authors. However, lead authors do not significantly tend to establish relationships with highly connected participating authors.

Cognitive and institutional proximity significantly influence tie formation. Higher cognitive proximity between lead and participating authors facilitates knowledge understanding and innovation, while higher institutional proximity reduces uncertainty and risk, promoting mutual trust and knowledge flow.

Our study advances research collaboration theory in two key ways. First, to our knowledge, it is the first to explore the formation mechanisms of research collaboration from the leadership-participation relationship perspective. Previous studies modeled research collaboration as undirected networks, capturing relationships between collaborators but neglecting the role of research leadership. Consequently, they only considered mechanisms like triadic closure, preferential attachment, and homophily/heterophily in undirected social networks. However, leadership-participation relationships are crucial in research collaboration.

Additionally, we examined the effects of reciprocity and preferential attachment to uncover the micro-mechanisms underlying leadership-participation relationships. Our analysis reveals that homophily, transitivity, and reciprocity significantly influence the formation of leadership-participation ties. We also found that out-degree preferential attachment affects these relationships. In summary, the directed leadership-participation perspective offers a deeper understanding of research collaboration, extending the theoretical mechanisms of directed social network formation to undirected networks, and more precisely revealing the inner workings of research cooperation.

Our study has important practical implications for policymakers, research teams, and individuals. First, both exogenous author attributes and endogenous network structure factors contribute to the formation of leadership-participation relationships. Policymakers should consider both factors. For example, researchers tend to collaborate with those from similar institutional backgrounds. It is recommended that policymakers enhance institutional culture integration and coordination to reduce barriers to effective leadership relationships. Additionally, there is a negative cyclic triadic closure effect in leadership-participation relationships: if researcher *i* leads *j* and *j* leads *k*, then *k* is less likely to lead *i* in the future.

Since research leadership is a form of academic social exchange, policymakers should encourage senior researchers to engage in junior researchers' projects to promote their development and avoid hierarchical obstacles.

Second, researchers should pay attention to relationship types and collaboration roles when forming partnerships and assigning tasks. For instance, while there is no clear gender bias in choosing research leaders, researchers tend to collaborate with those of the same gender. Research teams should aim for gender balance and emphasize cross-gender collaboration when assigning tasks and advancing projects. Researchers should avoid cyclic triadic closure in leadership relationships and consider others' leadership and participation roles.

Lastly, preferential attachment plays a crucial role in network formation, but it is out-degree rather than in-degree preferential attachment that affects leadership-participation relationships. Junior researchers are encouraged to take on more leadership roles in research collaborations, rather than just participating.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 72171217).

## REFERENCES

- Abbasi, A., Hossain, L., Uddin, S., & Rasmussen, K. J. (2011). Evolutionary dynamics of scientific collaboration networks: multi-levels and cross-time analysis. *Scientometrics*, 89(2), 687-710.

- Balahurovska, I. (2023). Network leadership is a promising direction in management science. *Marketing i menedžment innovacij*, 14(1), 133-145.
- Carayol, N., Bergé, L., Cassi, L., & Roux, P. (2019). Unintended triadic closure in social networks: The strategic formation of research collaborations between French inventors. *Journal of Economic Behavior & Organization*, 163, 218-238.
- Chen, K., Zhang, Y., & Fu, X. (2019). International research collaboration: An emerging domain of innovation studies? *Research policy*, 48(1), 149-168.
- Chinchilla-Rodríguez, Z., Sugimoto, C. R., & Larivière, V. (2019). Follow the leader: On the relationship between leadership and scholarly impact in international collaborations. *PloS one*, 14(6), e0218309.
- Clark, B. Y., & Llorens, J. J. (2012). Investments in scientific research: Examining the funding threshold effects on scientific collaboration and variation by academic discipline. *Policy Studies Journal*, 40(4), 698-729.
- Dagnino, G. B., Levanti, G., & Mocciaro Li Destri, A. (2016). Structural dynamics and intentional governance in strategic interorganizational network evolution: A multilevel approach. *Organization Studies*, 37(3), 349-373.
- González-Alcaide, G., Park, J., Huamaní, C., & Ramos, J. M. (2017). Dominance and leadership in research activities: Collaboration between countries of differing human development is reflected through authorship order and designation as corresponding authors in scientific publications. *PloS one*, 12(8), e0182513.
- He, C., Liu, F., Dong, K., Wu, J., & Zhang, Q. (2023). Research on the formation mechanism of research leadership relations: An exponential random graph model analysis approach. *Journal of Informetrics*, 17(2), 101401.
- He, C., Wu, J., & Zhang, Q. (2021). Characterizing research leadership on geographically weighted collaboration network. *Scientometrics*, 126, 4005-4037.
- Jiang, L. A., Zhu, N., Yang, Z., Xu, S., & Jun, M. (2018). The relationships between distance factors and international collaborative research outcomes: A bibliometric examination. *Journal of Informetrics*, 12(3), 618-630.
- Jiang, Q., Tan, C.-H., Phang, C. W., & Wei, K. K. (2012). Using sequence analysis to classify web usage patterns across websites. 2012 45th Hawaii International Conference on System Sciences,
- Kaltenberg, M., Jaffe, A. B., & Lachman, M. E. (2023). Invention and the life course: Age differences in patenting. *Research policy*, 52(1), 104629.
- Klavans, R., & Boyack, K. (2010). Toward an objective, reliable and accurate method for measuring research leadership. *Scientometrics*, 82(3), 539-553.
- Lester, J., & Kezar, A. (2017). Strategies and challenges for distributing leadership in communities of practice. *Journal of leadership studies*, 10(4), 17-34.
- Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2), 404-409.
- Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American journal of sociology*, 110(4), 1132-1205.
- Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59, 425-465.
- Sekara, V., Deville, P., Ahnert, S. E., Barabási, A.-L., Sinatra, R., & Lehmann, S. (2018). The chaperone effect in scientific publishing. *Proceedings of the national academy of sciences*, 115(50), 12603-12607.
- Stimson, R., Stough, R. R., & Salazar, M. (2009). Leadership and institutions in regional endogenous development. In *Leadership and Institutions in Regional Endogenous Development*. Edward Elgar Publishing.
- Sutanto, J., Jiang, Q., & Tan, C.-H. (2021). The contingent role of interproject connectedness in cultivating open source software projects. *The Journal of Strategic Information Systems*, 30(1), 101598.
- Walther, J., Sochacka, N. W., Benson, L. C., Bumbaco, A. E., Kellam, N., Pawley, A. L., & Phillips, C. M. (2017). Qualitative research quality: A collaborative inquiry across multiple methodological perspectives. *Journal of Engineering Education*, 106(3), 398-430.
- Xia, F., Chen, Z., Wang, W., Li, J., & Yang, L. T. (2014). MVCWalker: Random walk-based most valuable collaborators recommendation exploiting academic factors. *IEEE Transactions on Emerging Topics in Computing*, 2(3), 364-375.

- Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *The leadership quarterly*, 12(4), 451-483.
- Zhou, X., Ding, L., Li, Z., & Wan, R. (2017). Collaborator recommendation in heterogeneous bibliographic networks using random walks. *Information Retrieval Journal*, 20, 317-337.