Connector and Provincial Hub Dichotomy in Scientific Collaborations Identified by Reinforcement Learning Algorithm⋆

Feifan Liu[1,2], Shuang Zhang[1] and Haoxiang Xia[1,2,*]

[1]Institute of Systems Engineering, Dalian University of Technology, No.2 Linggong Road, Dalian, 116024, Liaoning, China
[2]Institute for Advanced Intelligence, Dalian University of Technology, No.2 Linggong Road, Dalian, 116024, Liaoning, China

Abstract

Scientific problem-solving relies on effective organizational patterns of research collaboration. To recognize the more complex cross-community collaboration patterns of researchers in modern science, this study probes the central core structure of co-authorship networks at the mesoscale, aiming at understanding the emerging structural characteristics and functional performance of the effectiveness of complex research and innovation systems. Taking the field of physics as an example, combining the deep reinforcement learning pre-training model with the hub role information of the complex network, this study identifies both the provincial hub and global connector hub and the emergence of multi-core structures at the mesoscopic level of the scientific collaboration network. The existence of the multi-core structure reflects the spontaneous formation of "local centrality and global decentrality" in the scientific collaboration system, which makes the knowledge creation system economical at the structural level and efficient in the functions of global collaboration and knowledge diffusion. Through an analysis of the structural and functional characteristics and mesoscale collaborative organizational structures of researchers, this study enhances comprehension and insights into the inherent factors propelling scientific development and the dynamics of collective knowledge creation. The findings contribute valuable perspectives for the establishment of inclusive scientific research management policies, fostering a more sophisticated scientific research and innovation system.

Keywords

Scientific collaborative behavior, complex network analysis, deep reinforcement learning, hub role identification

1. Introduction

The scientific research and innovation system embodies a form of "collective intelligence," where individual scholars possessing specialized knowledge and intellectual capacity collaboratively tackle intricate real-world challenges through self-organizing coordination, thereby propelling the advancement of knowledge domains [1]. Within this context, the scientific collaboration network constitutes a fundamental component of the overall innovation framework, embodying the interactive and cooperative dynamics among researchers [2, 3, 4, 5]. Ongoing development in knowledge engineering and the science of science discipline center on unraveling the nature of collective collaborative behavior, uncovering emergent collaborative patterns, and elucidating the underlying mechanisms driving knowledge creation system [6, 7, 8, 9].

Existing studies has demonstrated that co-authorship networks typically exhibit typical heterogeneity, confirming that these networks feature a high degree of uneven distribution in connectivity [10, 11, 12]. This implies that scientists with extensive social ties wield significant influence over the network as a whole, often engaging preferentially in collaborations with other highly influential peers, thus giving rise to the formation of "rich clubs" [13, 14].

Despite this, the investigation into the diversity of pivotal actors within expansive scientific collaboration networks remains underexplored, particularly concerning the identification of mesoscale core structures that bolster global efficiency within large-scale social systems. There is a dearth of research addressing how researchers with varying levels of social capital or differing types of social linkages contribute to the social division of cognitive labor in scientific communities.

This study aims to address these pressing issues by identifying and examining multi-core structures within co-authorship networks using a mesoscopic lens that taps into the inherent community structure. Leveraging a pre-trained reinforcement learning algorithm[15], it focuses on identifying key players within the co-authorship milieu. By combining the complex network topology theory, the study distinguishes between provincial hub scientists—those central within their respective communities—and connector hub scientists who bridge different communities. Moreover, it delves deeper to discern and analyze the multifaceted club-like properties and functions of members within these two core structural typologies. The results of this study promise to enrich our comprehension of the intricate collaborative patterns in a large-scale social innovation ecosystem.

2. Dataset

This study focuses on the field of physics. We use the scientific publications in the journals of the American Physical Society (APS) from the period 1985 to 2009 [16]. After the necessary pre-processing procedure, the dataset finally contains 104,484 researchers and their 848,231 edges established by coauthorship relations.

3. Results

In this study, we propose an interpretable framework to detect and analyze crucial core structures within large-scale co-authorship networks, integrating previously mentioned research concepts alongside club structure detection algorithms. This approach involves applying a second-stage "key player" detection algorithm, which ranks nodes in the co-authorship network based on their "criticality."

As depicted in Figure 1a-c, the network resilience experiment shows the significance of the key members detected using the deep reinforcement learning algorithm. Figure 1a illustrates the ratio of the maximum connected subgraph size to the potential maximum size after systematically re-

Joint Workshop of the 5th Extraction and Evaluation of Knowledge Entities from Scientific Documents and the 4th AI + Informetrics (EEKE-AII2024), April 23 24, 2024, Changchun, China and Online

*Corresponding author.
liufeifan@dlut.edu.cn (F. Liu); shuang94@mail.dlut.edu.cn (S. Zhang); hxxia@dlut.edu.cn (H. Xia)

moving nodes according to their ranked criticality. The sheer size and complexity of the entire co-authorship network make visualizing it challenging. Therefore, Figures 1b-c present embeddings of the co-authorship network graphs between select communities (7, 10, and 14) to exemplify the influence of key members on the network's architecture. The findings reveal that eliminating just the top 28% of key members causes a near-total collapse of the network, reducing the maximum connected subgraph size to almost zero. These results suggest a three-phase impact of key members on the overall network resilience. From Figures 1b-c, it becomes evident that the "key members" recognized by the deep reinforcement learning algorithm play a significantly more pivotal role in maintaining the network structure compared to randomly chosen nodes.

To assess the overlap between the "key members" and the "pivotal players" in the co-authorship network and verify if they support one another, the study conducts statistical analyses. Given that real-world networks tend to display hierarchical modularity [17, 18], we calculate the modularity of each collaborative community, partitioning them further into sub-communities using a co-authorship network community detection algorithm [19]. We then identify "pivotal" roles within these sub-communities. With an average modularity of around 0.71 across the 20 sub-communities, this suggests a prevalent hierarchical modular organization pattern within the co-authorship network.

Moreover, the multi-scale hierarchical modular structure observed in the co-authorship network reflects the inherent hierarchical structure of domain knowledge. Research directions, topics, subfields, and disciplines compose the knowledge hierarchy in a discipline, and researchers adaptively form co-authorships that embed research problems within different knowledge system scales.

Figure 1d presents the variation of club coefficients within the sub-communities of the multi-scale physical domain as a function of the proportion of deleted nodes (f). It demonstrates that each sub-community contains both global connector hubs and local provincial hubs, with global connector hubs exhibiting a stronger cohesive core structure relative to provincial hubs from the complex network system view.

Figures 1e-f summarize the density and number distribution of "pivotal role" members in the "key member" sequence groups. Key observations include: 1) A significant majority of globally and locally pivotal members are concentrated in the initial sequence subgroups of "key members." This indicates that the higher the criticality rank, the greater the proportion of "pivotal role" members. 2) There is a descending order correlation between the criticality of "pivotal role" member classification. 3) Nodes with high degree are more critical and occur in larger numbers across both the global collaborative communities and the sub-communities, demonstrating a consistent pattern in terms of importance and centrality within the network structure.

4. Discussion and Conclusion

Scientific collaborative behavior is a cornerstone of large-scale knowledge exploration among researchers and significantly influences their academic productivity and impact. Co-authorship networks serve as a primary analytical tool for deciphering collaboration patterns among researchers within a knowledge landscape. As network science theories and methodologies have evolved, so too has the examination

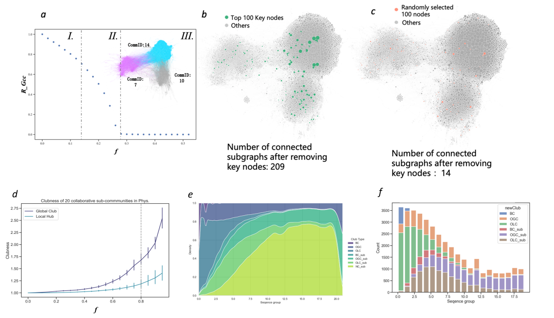

Figure 1: Structural Characteristics of Connector and Provincial Hub in Scientific Collaborations Identified by Pre-trained Reinforcement Learning Algorithm

of co-authorship networks' macroscopic and mesoscopic attributes, including scale-free, small-world, modularity, and club structures. While the modularity-based and collective collaboration aspects of these networks have received substantial attention, the in-depth analysis of core structures within co-authorship networks from the modularization and collaboration perspective remains an open issue.

Recent research has demonstrated that mesoscopic core structures have indeed been detected and studied in various domains like biology, transportation, and power systems [20, 21, 22], playing a pivotal role in global information integration and subsystem coordination. This study extends this line of inquiry by exploring the existence of similar mesoscopic core structures in co-authorship networks and analyzing their associated network structural traits and functional implications.

By harnessing the interpretability of complex topology theory and the representational power of deep learning techniques, this study introduces an interpretable framework to identify and analyze the key cohesive structures in co-authorship networks. The study reveals the coexistence of two distinct core structures: local provincial hubs that primarily consolidate community members with sparse interconnections among themselves, and global connector hubs that act as bridges between researchers across different research areas within the collaborative community, maintaining tight interconnections.

These two types of hubs exhibit minimal overlap and possess unique network structural characteristics, exerting varying degrees of influence on other network members. The provincial hubs demonstrate a star-shaped, centralized structure, whereas the connector hubs showcase a flatter and less centralized pattern of close collaborations.

The coexistence of local centrality and global decentralization in co-authorship networks reflects a delicate balance between cost-effectiveness, stability, and flexibility within the large-scale researcher-driven knowledge exploration process. Future research aims to delve into the potential universal patterns of scientific meso-core structures across various disciplines and career stages, drawing upon comprehensive academic datasets covering multiple fields and historical periods.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No.71871042 and 72371052).

References

[1] A. Zeng, Z. Shen, J. Zhou, J. Wu, Y. Fan, Y. Wang, H. E. Stanley, The science of science: From the perspective of complex systems, Physics Reports 714-715 (2017) 1–73. doi:10.1016/j.physrep.2017.10.001.

[2] M. E. J. Newman, Coauthorship networks and patterns of scientific collaboration., Proceedings of the National Academy of Sciences of the United States of America 101 Suppl (2004) 5200–5. doi:10.1073/pnas.0307545100.

[3] Z. Xie, Z. Ouyang, J. Li, E. Dong, D. Yi, Modelling transition phenomena of scientific coauthorship networks, Journal of the Association for Information Science and Technology 69 (2018) 305–317. doi:10.1002/ASI.23935.

[4] S. Yu, F. Alqahtani, A. Tolba, I. Lee, T. Jia, F. Xia, Collaborative team recognition: A core plus extension structure, Journal of Informetrics 16 (2022) 101346. doi:https://doi.org/10.1016/j.joi.2022.101346.

[5] A. Hajibabaei, A. Schiffauerova, A. Ebadi, Women and key positions in scientific collaboration networks: analyzing central scientists'profiles in the artificial intelligence ecosystem through a gender lens, Scientometrics 128 (2023) 1219–1240. doi:10.1007/s11192-022-04601-5.

[6] Y. Ding, Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks, Journal of Informetrics 5 (2011) 187–203. doi:10.1016/j.joi.2010.10.008.

[7] W. Li, T. Aste, F. Caccioli, G. Livan, Early coauthorship with top scientists predicts success in academic careers, Nature Communications 10 (2019) 5170. doi:10.1038/s41467-019-13130-4.

[8] Y. Zhu, L. Quan, P.-Y. Chen, M. C. Kim, C. Che, Predicting coauthorship using bibliographic network embedding, Journal of the Association for Information Science and Technology (2022). doi:10.1002/asi.24711.

[9] D. Aguirre-Guerrero, R. Bernal-Jaquez, A methodology for the analysis of collaboration networks with higher-order interactions, Mathematics 11 (2023). doi:10.3390/math11102265.

[10] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, Evolution of the social network of scientific collaborations, Physica A: Statistical Mechanics and its Applications 311 (2002) 590–614. doi:10.1016/S0378-4371(02)00736-7.

[11] R. Guimera, Team assembly mechanisms determine collaboration network structure and team performance, Science 308 (2005) 697–702. doi:10.1126/science.1106340.

[12] A. M. Petersen, Quantifying the impact of weak, strong, and super ties in scientific careers, Proceedings of the National Academy of Sciences of the United States of America 112 (2015) E4671–E4680. doi:10.1073/pnas.1501444112.

[13] V. Colizza, A. Flammini, M. A. Serrano, A. Vespignani, Detecting rich-club ordering in complex networks, Nature Physics 2 (2006) 110–115. doi:10.1038/nphys209.

[14] M. Szell, R. Sinatra, Research funding goes to rich clubs., Proceedings of the National Academy of Sciences of the United States of America 112 (2015) 14749–50. doi:10.1073/pnas.1520118112.

[15] C. Fan, L. Zeng, Y. Sun, Y. Y. Liu, Finding key players in complex networks through deep reinforcement learning, Nature Machine Intelligence 2 (2020) 317–324. doi:10.1038/s42256-020-0177-2.

[16] S. American Physical, Aps data sets for research, 2018. URL: https://journals.aps.org/datasets.

[17] A. Trusina, S. Maslov, P. Minnhagen, K. Sneppen, Hierarchy measures in complex networks, Physical Review Letters 92 (2004). doi:10.1103/PhysRevLett.92.178702.

[18] P. L. Krapivsky, S. Redner, Emergent network modularity, J. Stat. Mech (2017). doi:10.1088/1742-5468/aa7a3f.

[19] V. D. Blondel, J. L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008 (2008) P10008–P10008. doi:10.1088/1742-5468/2008/10/P10008.

[20] R. F. Betzel, J. D. Medaglia, D. S. Bassett, Diversity of meso-scale architecture in human and non-human connectomes, Nature Communications 9 (2018). doi:10.1038/s41467-017-02681-z.

[21] S. Kojaku, M. Xu, H. Xia, N. Masuda, Multiscale core-periphery structure in a global liner shipping network, Scientific Reports 9 (2019) 404–415. doi:10.1038/s41598-018-35922-2.

[22] C. Seguin, O. Sporns, A. Zalesky, Brain network communication: concepts, models and applications, Nature Reviews Neuroscience 24 (2023). doi:10.1038/s41583-023-00718-5.