Evolutionary Characteristics of Scientific Communities in Disciplinary Fields Based on Evolutionary Event Detection: A Case Study of Library and Information Science (Postprint)

Authors: Li Gang, Tang Jing, Mao Jin, Tian Yunpei, Zhang Bin

Date: 2023-04-01T16:02:55+00:00

Abstract

[Purpose/Significance] As important knowledge collectives in contemporary scientific research, studying the evolutionary characteristics of research communities during disciplinary development holds significant importance for exploring domain development patterns and promoting knowledge innovation.

[Method/Process] To investigate the dynamic evolutionary characteristics of research communities, this study takes the field of library and information science as an example, approaches from the perspective of evolutionary event detection, employs the Leiden algorithm to partition research communities, and constructs research community evolution paths and evolution trees. On this basis, it identifies research community evolution events and reveals the evolution patterns and characteristics of research communities from three aspects: overall analysis of research community evolution, analysis of research community evolution path and evolution tree features, and statistical feature analysis of research community evolution events.

[Results/Conclusion] The study reveals that research community scale demonstrates a vigorous development trend, research community evolution trees exhibit two evolution patterns, growth-type evolution events mostly occur in large-scale research communities with higher publication volumes, birth and extinction evolution events both occur in small-scale research communities with higher publication volumes, while evolution events such as merging, partial merging, splitting, and attenuation have relatively small average community scales and lower publication volumes, further demonstrating that cooperation and exchange among research communities tend to be more frequent, and research community evolution is becoming increasingly complex.

Full Text

Preamble

Research on the Evolution Characteristics of Scientific Research Communities in Subject Fields Based on Evolutionary Event Detection—Taking Library and Information Science as an Example

Li Gang[1], Tang Jing[1], Mao Jin[1], Tian Yunpei[1], Zhang Bin[2]
[1]Center for the Studies of Information Resources, Wuhan University, Wuhan 430072
[2]School of Information Management, Nanjing University, Nanjing 210023

Abstract:
[Purpose/Significance] Scientific research communities serve as important knowledge groups in contemporary scientific research. Investigating their evolutionary characteristics during disciplinary development holds significant value for exploring field development patterns and promoting knowledge innovation. [Method/Process] To explore the dynamic evolution characteristics of scientific research communities, this study takes the field of Library and Information Science (LIS) as an example. From the perspective of evolutionary event detection, the Leiden algorithm is employed to partition research communities and construct their evolution paths and evolution trees. On this basis, community evolution events are identified, and the evolution patterns and characteristics of research communities are revealed through three aspects: overall analysis of community evolution, analysis of evolution paths and tree characteristics, and statistical analysis of evolution events. [Result/Conclusion] The findings indicate that the scale of research communities shows vigorous development momentum. Evolution trees exhibit two distinct evolution modes. Growth-type evolution events mostly occur in large communities with high publication output, while formation and dissolution events occur in small communities with relatively high publication output. Evolution events such as merge, partial merge, split, and shrink have small average community sizes and low publication output, further demonstrating that cooperation and exchange between research communities are becoming increasingly frequent and that community evolution is growing more complex.

Keywords: scientific research community; Leiden algorithm; evolution events; evolution characteristics
Classification Number: G251
DOI: 10.13266/j.issn.0252-3116.2021.17.008

With the gradual formation of the fourth research paradigm and the rise of interdisciplinary research, scientific collaboration has become an effective approach for researchers to address complex problems. Research teams, as the fundamental carriers of scientific collaboration, constitute the main body of scientific and technological research—physical groups composed of researchers with complementary advantages who are willing to cooperate and collaborate toward common research purposes, goals, and methodologies [1]. However, as research problems become increasingly complex and systematic, virtual research communities dominated by collaborative relationships have emerged [2]. In this paper, scientific research communities refer to research groups formed by researchers through co-authorship relationships [3]. Identifying these communities and deeply exploring their evolution characteristics is crucial for understanding field development patterns, formulating science and technology development policies, and cultivating innovative talent, holding significant academic value and strategic importance.

Early studies on research communities primarily adopted static perspectives, identifying communities in static co-authorship networks and analyzing their quantitative characteristics and topological structures using informetrics and social network analysis methods [4-5]. However, actual research communities change dynamically over time, with members joining or leaving and new collaborations forming or existing ones ending. Traditional static methods cannot adequately analyze these continuous evolution processes, making dynamic exploration of community evolution a key focus for researchers [6].

Current studies mainly investigate community evolution by defining and identifying evolution events and analyzing network attributes across time windows. However, most research offers incomplete definitions of evolution events and fails to deeply reveal the evolution characteristics and patterns of research communities. Addressing this gap, this paper takes LIS as a case study, employs the Leiden algorithm to partition research communities, identifies 12 types of evolution events, and analyzes overall evolution patterns, evolution paths and tree characteristics, and statistical features of evolution events to reveal the intrinsic evolution characteristics and modes of research communities.

2 Related Research

Most existing studies convert networks into static graphs across different time windows to explore community evolution, typically following two steps: community identification and evolution analysis.

2.1 Community Identification

Research communities are formed by researchers and their co-authorship relationships. Previous studies have identified them based on either topical or topological structures. From a topical perspective, researchers form communities through shared research interests. For instance, Chen Wenjie proposed a community detection algorithm integrating author-topic features [7], while S. Jung clustered authors with common topic preferences into "common interest groups" to explore topic evolution [8]. J. Mao et al. argued that research communities possess both topological and topical coherence, proposing a topical community identification method to study community characteristics [3].

From a topological perspective, co-authorship networks exhibit clear community structures due to varying collaboration strengths. Researchers commonly employ community detection algorithms from complex network theory, such as Infomap [9], Label Propagation [10], and Louvain [11]. The Louvain algorithm is a modularity-based graph clustering algorithm comprising two iterative steps: node local movement and network aggregation. In the local movement phase, nodes are moved to neighboring communities to maximize modularity gain. In the aggregation phase, each community is collapsed into a single node to form an aggregated network, with edge weights recalculated. These phases repeat until network modularity is maximized. Compared to other algorithms, Louvain offers high efficiency and effective partitioning [12], making it a preferred choice for community identification. For example, Yu Houqiang [13] and Zou Bentao [14] used Louvain to identify research teams in artificial intelligence and explore their characteristics. Jiang Wenhua et al. [15] applied Louvain to partition researchers based on research interest similarity networks in library science.

However, the Louvain algorithm may identify communities with poor internal connectivity when nodes serve as bridges between communities. To address this, V. Traag et al. proposed the Leiden algorithm, which employs a fast local movement process for initial community partitioning to improve speed and adds a refinement step to optimize partitions. This retains Louvain's advantages while enhancing partitioning quality [16].

2.2 Community Evolution Analysis

Research on community evolution analysis falls into two categories: (1) examining structural changes to reflect evolution characteristics [17-18], such as Zou Bentao et al.'s study calculating extreme distributions of four network metrics (nodes, edges, network density, average clustering coefficient) to explore evolution patterns of high-productivity teams in AI [14]; and (2) identifying key evolution events to investigate community evolution patterns, such as S. Asur et al.'s classification into five event types (continue, merge, split, form, dissolve) [20], M. Takaffoli et al.'s addition of growth and shrink events [21], and G. Palla et al.'s correlation analysis between events and community size/duration [22].

Current research has limitations: topology-based tracking only reveals overall evolution patterns, while previous studies oversimplify evolution event definitions [23], limiting in-depth analysis. Moreover, research on evolution paths and patterns remains insufficient, with indicators still focusing on basic quantitative features like community age and size [24], without incorporating research output dimensions. K. Mohammadmosaferi et al. defined 12 community evolution events considering partial evolution phenomena (e.g., subsets of two communities merging into a new one), providing deeper insights into evolution processes [23]. Building on this, our study proposes a method to explore research community evolution characteristics in subject fields, using LIS as a case study, employing the Leiden algorithm for community detection, and identifying these

12 evolution events to reveal intrinsic evolution patterns through network features, evolution paths, and event analysis.

3 Research Design

This study proposes a method to investigate research community evolution characteristics from an evolutionary event detection perspective, analyzing evolution processes based on comprehensive event definitions. Using LIS as an empirical case, the workflow is illustrated in Figure 1 [Figure 1: see original paper].

3.1 Data Collection and Cleaning

To resolve author name disambiguation, this study selects the Scopus database, which assigns unique author identifiers, as the data source. We collected all literature published between 2000-2019 from the top 100 LIS journals by CiteScore, yielding 78,483 documents. After data cleaning, 70,015 documents remained, involving 106,100 authors.

We filtered out low-frequency authors with fewer than two publications and divided the 20-year period into 10 time windows (two years each). For each window, we constructed co-authorship networks with authors as nodes and co-authorship relationships as edges, resulting in 10 networks.

3.2 Community Detection

Community identification is crucial for evolution analysis. Based on co-authorship network topology, we employed community detection algorithms. To address Louvain's connectivity issues, we used the Leiden algorithm from Python's igraph package, with modularity as the objective function and co-authorship frequency as edge weights, to identify communities in each network. Following the principle that triadic closure serves as a basic unit in knowledge network analysis [26], we analyzed communities with more than two members, naming them as "time window, community ID."

The Leiden algorithm improves upon Louvain through three steps: fast local movement, partition refinement, and network aggregation based on refined partitions. This ensures community quality while enhancing efficiency [25].

Finally, we input the 10 networks and community partition data into Cytoscape [27] to calculate network characteristics and visualize results.

3.3 Evolution Event Definition and Path Construction

After identifying communities, we extracted evolution paths and trees and recognized evolution events. Figure 2 [Figure 2: see original paper] illustrates a possible evolution scenario, where "au" represents community members and arrows indicate member joining or leaving.

We define Cn as the set of all communities in time window n:

$$C_n = \{C(n,1), C(n,2), \dots, C(n,m)\}, \quad T = n \tag{1}$$

where T=n denotes time window n and C(n,m) represents a specific community.

An evolution path P is a sequence of communities across adjacent time windows with common members, showing member flow from formation to dissolution:

$$P_x = [C(n,y), \dots, C(m-1,i), C(m,j)], \quad n < m-1 < m \tag{2}$$

where the intersection of C(m-1,i) and C(m,j) is non-empty.

An evolution tree comprises evolution paths sharing the same formative community:

$$\text{Tree} = \{P_1, \dots, P_j\} \tag{3}$$

For example, in Figure 2, the set {[C(2,3),C(3,4),C(4,3)], [C(2,3),C(3,4),C(4,4)]} shares formative community C(2,3), constituting an evolution tree.

Community evolution events represent significant changes during temporal development [28]. While previous studies typically categorized events into seven types (form, grow, split, merge, shrink, dissolve, continue), they overlooked partial evolution phenomena where communities form from subsets of multiple communities (e.g., C(3,4) in Figure 2 formed from subsets of C(2,3) and C(2,2)). Therefore, we adopt K. Mohammadmosaferi et al.'s 12 evolution events [23] (Table 1) and propose a common-member-based algorithm for event identification. All communities at t=0 are considered formative.

3.4 Community Evolution Characteristic Analysis

We analyze evolution characteristics from three perspectives: overall community evolution, evolution paths and tree features, and statistical characteristics of evolution events. For overall analysis, we visualize co-authorship networks using Cytoscape and calculate basic statistics (publications, community count, nodes, edges, average clustering coefficient) to understand structural changes.

For evolution path analysis, we synthesize event frequency, path count, and start times to explore temporal evolution relationships. To identify evolution patterns, we compare key metrics across evolution trees: member count, average community size, average publication output, and event frequencies.

For evolution feature exploration, we conduct cross-analysis of events with community size and publication output to uncover deeper characteristics, performing significance tests and correlation analysis.

4 Results Analysis

4.1 Overall Analysis of LIS Research Communities

Overall analysis reveals temporal evolution trends. Figure 3 [Figure 3: see original paper] shows co-authorship networks for 2000-2001, 2008-2009, and 2018-

2019, with node colors representing communities. The networks remain disconnected, containing numerous independent small communities. The largest connected component shows clear community structures of varying scales based on collaboration intensity. Temporally, networks become denser, with the largest component expanding, reflecting vigorous development and strengthening inter-community connections.

Figure 4 [Figure 4: see original paper] displays a local network for 2008-2009, where node labels indicate community IDs. Community 5-61 exemplifies a mature community with 55 nodes and 82 edges, demonstrating dense internal connections and validating Leiden's superiority.

Table 2 presents basic network statistics. Publication volume grows annually, indicating deepening research. Increasing node and edge counts reflect expanding network scale. The rising average clustering coefficient (from 0.146 to 0.400) shows intensifying collaboration and stabilized "small group" phenomena.

Community-level statistics show co-evolution with the overall network. The proportion of publications within communities rose from 60.1% to 79.3%, and edge proportion from 73.5% to 94.6%, indicating strengthening collaborative intensity and productivity.

4.2 Evolution Path Analysis in LIS

Evolution paths reveal temporal relationships. We constructed 5,796 evolution paths in LIS. Path length and starting distribution (Figure 5 [Figure 5: see original paper]) show 59.2% of paths exceed length 8, demonstrating strong continuity.

Based on our event definitions, we identified evolution events across time windows (Figure 6 [Figure 6: see original paper]). Events like partial merge & grow, split & grow, and partial survive & grow increased over time, reflecting growing complexity and enhanced cooperation. High frequencies of form and dissolve events indicate continuous community entry and exit. Overall, growth-type events dominate, showing dynamic, vigorous development.

To identify evolution patterns, we classified trees with >10 communities as large (37 trees, 1,363 communities) and others as small (1,543 trees, 3,003 communities). Figures 7 [Figure 7: see original paper] and 8 [Figure 8: see original paper] compare their characteristics (form and dissolve excluded due to community count effects).

Large trees (2.34% of total) show higher average community size and publication output. Their evolution pattern centers on "partial merge & grow" (45.2%) and "split & grow" (30.5%), indicating intensive inter-community exchange and continuous new member integration, boosting research output.

Small trees, though numerous, have lower average size and output, consisting mainly of small communities. Their pattern features "partial merge & grow" and "partial survive & grow," with higher continue and shrink rates, indicating weaker inter-community collaboration.

We visualized evolution paths using Sankey diagrams (Figures 9 [Figure 9: see original paper] and 10 [Figure 10: see original paper]). Node length represents member count; branch width shows member flow to next-stage communities.

The large community tree (Figure 9) involves 197 communities and 352 paths (average length: 6.23), showing strong continuity. Events center on "partial merge & grow" and "split & grow," with dense member exchange and expanding scale.

The small community tree (Figure 10) involves fewer communities (average path length: 2.67), showing weaker continuity and higher member exit rates. Events are split-dominated with lower collaboration intensity.

4.3 Evolution Event Characteristics in LIS

Evolution event detection is vital for understanding dynamic patterns. We conducted multi-factor ANOVA on time windows, community size, publication count, and events, plus correlation analysis between size and publications (Tables 3 and 4).

Table 3 shows significant differences across time windows and event types ($p<0.05$). Table 4 reveals a weak positive correlation ($r=0.633$) between member count and publications, indicating size is not the sole output determinant.

Figures 11 [Figure 11: see original paper] and 12 [Figure 12: see original paper] present heatmaps of average community size and publication output by event type and time window.

Form and dissolve events involve small communities (average size: 3-4 members), occurring primarily in small groups. Growth-type events show larger average sizes, especially partial merge & grow events, which are largest and increasing over time. This suggests researchers actively recruit new members and strengthen inter-community cooperation when facing complex problems. Strict merge and partial merge events involve smaller communities, indicating these occur mainly among small groups. Split and shrink events also show small sizes due to member exits.

Publication output is higher for growth-type events, particularly partial merge & grow, which consistently ranks highest, demonstrating strong research capacity. Form and dissolve events also show notable output, likely due to high frequency. Merge, split, and continue events show lower but rising output over time, indicating improving research levels.

5 Conclusions and Discussion

This study proposes a method to explore research community evolution characteristics, using LIS literature (2000-2019) as data. We partitioned communities using the Leiden algorithm, analyzed network features temporally, extracted evolution paths, identified events, and summarized evolution patterns. The research enriches community evolution theory and reveals LIS community evolution patterns, aiding discovery, formation, and management.

Limitations include: (1) not considering non-continuous publication periods for researchers, and (2) not integrating topical features into evolution analysis. Future work will explore co-evolution of topics and communities for more comprehensive insights.

References

[1] Ren N, Zhou J N. Research on discovery and evaluation of scientific research teams in weighted co-authorship networks[J]. New Technology of Library and Information Service, 2015(9): 68-75.

[2] Evans T S, Lambiotte R, Panzarasa P. Community structure and patterns of scientific collaboration in business and management[J]. Scientometrics, 2017, 112(4): 851-875.

[3] Mao J, Cao Y, Lu K, et al. Topic scientific community in science: a combined perspective of scientific collaboration and topics in the digital library research community[J]. Information Processing & Management, 2019, 56(4): 1452-1465.

[4] Liu X, Bollen J, Nelson M L, et al. Co-authorship networks in the digital library research community[J]. Information Processing & Management, 2005, 41(6): 1462-1480.

[5] Li L, Zhu Q H. Empirical study of social network analysis in co-authorship analysis[J]. Information Science, 2008, 26(4): 549-555.

[6] Masmoudi A, Mezghani E, Bellaaj H, et al. A web-based knowledge management system for scientific research team[C]//2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises. Poland: IEEE, 2017.

[7] Chen W J. Research on community detection integrating node topic features[J/OL]. Information Theory and Practice: 1-10[2021-01-22]. http://kns.cnki.net/kcms/detail/11.1762.g3.20201223.0841.002.html.

[8] Jung S, Yoon W C. An alternative topic model based on common interest authors for topic evolution analysis[J]. Journal of Informetrics, 2020, 14(3): 101040.

[9] Rosvall M, Bergstrom C T. Maps of random walks on complex networks reveal community structure[J]. Proceedings of the National Academy of Sciences, 2008, 105(4): 1118-1123.

[10] Raghavan U N, Albert R, Kumara S. Near linear time algorithm to detect community structures in large-scale networks[J]. Physical Review E, 2007, 76(3 Pt 2): 036106.

[11] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 10: P10008.

[12] Sun Y. Research on community evolution event detection in social networks[D]. Shanghai: Shanghai Jiao Tong University, 2016.

[13] Yu H Q, Bai K, Zou B T, et al. Identification and leading team extraction of scientific research teams in artificial intelligence[J]. Library and Information Service, 2020, 64(20): 4-13.

[14] Zou B T, Wang Y F, Yu H Q. Evolution of high-productivity research teams in artificial intelligence[J]. Library and Information Service, 2020, 64(20): 23-33.

[15] Jiang W H, Xu J, Li G, et al. Analysis of Chinese library science research communities based on research interest similarity networks[J]. Journal of Modern Information, 2019, 39(9): 21-27.

[16] Traag V, Waltman L, Van Eck N J. From louvain to leiden: guaranteeing well-connected communities[J]. Scientific Reports, 2019, 9(1): 5233.

[17] Han T Q, Wang L M, Xu X. Evolution of scientific collaboration networks in Yangtze River Delta urban agglomeration—An empirical analysis based on SCIE and SSCI papers[J]. Information Studies: Theory & Application, 2020, 43(10): 151-156.

[18] Li G, Li C Y, Li X. Research on scientific team discovery based on social network analysis[J]. Library and Information Service, 2014, 58(7): 63-70, 82.

[19] Wang Y F, Li D Q, Yu H Q. Evolution of scientific collaboration networks and growth characteristics of high-impact scholars across lifecycle stages[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(2): 121-131.

[20] Asur S, Parthasarathy S, Ucar D. An event-based framework for characterizing the evolutionary behavior of interaction graphs[J]. ACM Transactions on Knowledge Discovery from Data, 2009, 3(4): 1-36.

[21] Takaffoli M, Fagnan J, Sangi F, et al. Tracking changes in dynamic information networks[C]//International Conference on Computational Aspects of Social Networks. Spain: IEEE, 2011.

[22] Palla G, Barabási A L, Vicsek T. Quantifying social group evolution[J]. Nature, 2007(446): 664-667.

[23] Mohammadmosaferi K K, Naderi H. Evolution of communities in dynamic social networks: an efficient map-based approach[J]. Expert Systems with Applications, 2020(147): 113221.

[24] Tang Q. Research on community discovery and evolution characteristics in scientific research networks[D]. Xi'an: Xidian University, 2015.

[25] Baes S, Dhall A, West J D, et al. Scalable and efficient flow-based community detection for large-scale graph analysis[J]. ACM Transactions on Knowledge Discovery from Data, 2017, 11(3): 1-30.

[26] Hu C P, Chen G. Research on triadic relations in conceptual knowledge networks from a hierarchical perspective[J]. Library and Information Service, 2014, 58(4): 11-16.

[27] Shannon P. Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. Genome Research, 2003, 13(11): 2498-2504.

[28] Xu B, Zhao Y W, Xu Y X. Community tracking algorithm based on associated group evolution similarity[J]. Complex Systems and Complexity Science, 2019, 16(1): 14-25.

Author Contributions

Li Gang: Conceptualized the research and provided revision guidance
Tang Jing: Designed the research, processed and analyzed data, and wrote the manuscript
Mao Jin: Proposed research questions and reviewed/revised the manuscript
Tian Yunpei: Processed and analyzed data
Zhang Bin: Analyzed data and revised the manuscript

Research on the Evolution Characteristics of Scientific Research Communities in Subject Fields Based on Evolutionary Event Detection—An Example of LIS

Li Gang[1], Tang Jing[1], Mao Jin[1], Tian Yunpei[1], Zhang Bin[2]
[1]Center for the Studies of Information Resources, Wuhan University, Wuhan 430072
[2]School of Information Management, Nanjing University, Nanjing 210023

Abstract: [Purpose/significance] Scientific research communities are important knowledge groups in contemporary science. Studying their evolution characteristics is significant for exploring field development laws and promoting knowledge innovation. [Method/process] Taking Library and Information Science as an example, this paper uses the Leiden algorithm to detect communities and construct evolution paths and trees. It identifies 12 evolution events and reveals evolution modes and characteristics through overall analysis, path/tree analysis, and event statistics. [Result/conclusion] Results show vigorous community scale development and two evolution modes. Growth events occur mostly in large, high-output communities, while formation and dissolution occur in small but relatively high-output communities. Merge, partial merge, split, and shrink events involve small communities with low output, proving increasingly frequent inter-community cooperation and growing evolution complexity.

Keywords: scientific research community; Leiden algorithm; evolution events; evolution characteristics