Postprint of a Review on Hybrid Network Community Detection Methods for Scientific Structure Analysis

Authors: Zhang Ruihong, Chen Yunwei, Deng Yong

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Research on community structures in complex networks has gradually become a powerful tool for scientists to conduct studies on the structure of science using bibliographic data, and the varying effectiveness of community detection exerts a decisive influence on the interpretation of scientific structures. This paper provides a systematic review of community detection methods for heterogeneous networks, aiming to offer reference and guidance for related research in this domain. [Method/Process] Through literature review, we elucidate the concept and types of heterogeneous networks, summarize community detection research for various types of heterogeneous networks from the perspectives of network construction and algorithmic innovation, and briefly introduce classical algorithms that underpin community detection in heterogeneous networks. [Results/Conclusion] By systematically reviewing and summarizing community detection work for different types of heterogeneous networks, we provide research perspectives and methodological insights for subsequent network analysis studies, while revealing the challenges faced and practical significance in the study of scientific structures, and prospecting potential directions for future research expansion.

Full Text

A Review of Community Detection Methods in Hybrid Networks for Scientific Structure Analysis

Zhang Ruihong[1,2], Chen Yunwei[1,2], Deng Yong[1,2]
[1]Scientometrics & Evaluation Research Center (SERC), Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041
[2]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract

[**Purpose/Significance**] Research on community structure in complex networks has gradually become a powerful tool for scientists to conduct scientific structure analysis using literature data, and different community detection results have a significant impact on the interpretation of scientific structure. This paper reviews community detection methods in hybrid networks to provide reference and insights for related research in this field. [**Method/Process**] Through literature research, this paper clarifies the concept and types of hybrid networks, summarizes community detection research on various types of hybrid networks from the perspectives of network construction and algorithm innovation, and briefly introduces classical algorithms supporting hybrid network community detection. [**Result/Conclusion**] By systematically reviewing community detection work on different types of hybrid networks, this study provides new perspectives and methods for subsequent network analysis research, reveals the challenges and practical significance in scientific structure research, and prospects for future research directions that may be further expanded.

Keywords: hybrid network; community detection; cluster analysis; collaboration; citation

1. Introduction

The increasing complexity and interdisciplinary nature of scientific research have blurred disciplinary boundaries, making scientific structure increasingly difficult to comprehend. Scientific structure is an objective existence formed over the long term, inherent and independent of human will, reflecting the internal logic of science through its disciplinary framework, subject structure, and knowledge architecture. Although the intrinsic essence of science remains objective and constant, its external manifestation continuously evolves as human understanding deepens. Effectively discovering scientific structure has become a focal issue in knowledge discovery research, holding significant value for exploring disciplinary evolution, identifying interdisciplinary infiltration, and uncovering frontier directions.

Since the formal introduction of the community concept in 2002, numerous studies on collaboration networks have emerged. For instance, Lambiotte and Panzarasa (2009) investigated how scientific collaboration patterns facilitate knowledge creation and diffusion through community detection in collaboration networks. Moline et al. (2017) studied the evolution of scientist communities in the talent management domain, enriching talent management dynamics. Zheng et al. (2017) conducted evolutionary analysis of communities based on author co-authorship networks in single journals, discovering more effective composite indices and lifecycle strategies for analyzing collaborator community evolution, providing new insights for dynamically observing academic community evolution through collaboration networks.

Meanwhile, citation network analysis has continuously enriched its connotations and methods alongside the development of social network analysis methods, enabling more accurate revelation of scientific structure and development processes through community analysis. For example, when Girvan and Newman first proposed the community concept in 2002, they applied the GN algorithm to perform community detection on the collaboration network of scientists at the Santa Fe Institute during 1999-2000 (118 scientists) (see Figure 1 [Figure 1: see original paper]), dividing these scientists into four communities (agent-based models for economic and transportation issues, mathematical models in ecology, statistical physics, and RNA structures). Kajikawa et al. used the FN algorithm to analyze temporal changes in citation network communities to identify emerging research fields and expand disciplinary knowledge structures. Chen Yunwei proposed a sample-weighted citation network community detection method based on the Louvain algorithm, representing scientific papers using a vector space model and calculating similarity between adjacent papers via cosine similarity as weights for citation edges. Subsequently, by comprehensively considering node structure and content similarity, the network was reconstructed, yielding clear community detection results. A representative recent work is the CitNetExplorer software developed by Waltman and Eck at Leiden University in 2013, which integrates the SLM community detection algorithm for citation network community research.

However, despite extensive community research on co-authorship networks, citation networks, and others in the library and information science field, most studies have focused solely on single-type node networks (such as homogeneous networks with only authors or papers as nodes), suffering from limitations including single analysis objects, single relationships, and incomplete or imprecise revelation of scientific structure. To address this, researchers have recently begun investigating community detection in networks with multiple node types or multiple relationships to enhance scientific structure analysis.

The type of network significantly influences community detection effectiveness and the scientific structure to be revealed. Particularly, different types of objects and their diverse interrelationships in hybrid networks offer richer functionality in revealing the abundant semantics carried by networks and may yield various mining results. Therefore, this review focuses on community detection research in hybrid networks within the library and information science domain, aiming to trace the development of using hybrid network methods to study scientific structure and its evolution, as well as potential future trends. Additionally, community detection algorithms for multi-type, multi-relation networks present greater difficulty and challenges, concentrating many research efforts on algorithmic innovation, which will also be briefly summarized below. Finally, commonly used community detection algorithms in current library and information science research are concisely summarized for scholars interested in conducting related studies.

2. Concept and Types of Hybrid Networks

This paper defines "hybrid networks" as networks containing multiple node types or multiple relationships, where the network simultaneously includes two or more types of nodes such as authors and papers, or edges encompass two or more relationships such as collaboration, citation, or topical similarity. According to definitions by Han Jiawei and B. Taskar et al., hybrid networks essentially belong to the category of heterogeneous networks—networks composed of multiple node types and multiple relationship types are heterogeneous [12-13]. However, while the concept of heterogeneous networks emphasizes structural complexity, the networks analyzed in this paper—integrating multiple nodes and relationships—focus on functional richness. Therefore, we propose the concept of "hybrid networks" to enable library and information science researchers to concentrate on functional enhancement rather than merely increasing complexity.

Through comparison with community detection research on single-type node networks (such as citation networks and collaboration networks) in library and information science, hybrid networks can be classified into three categories based on node and edge types: The first category is single-type node multi-relation networks, such as author-only networks containing both collaboration and citation relationships. The second category is multi-type node multi-relation networks, such as networks containing both authors and papers as nodes with collaboration and citation relationships. The third category is multi-type node single-relation networks, such as networks containing both authors and papers but only citation relationships.

The following sections elaborate on and analyze community detection research related to these three types of hybrid networks.

3. Community Detection in Single-Type Node Multi-Relation Hybrid Networks

In single-type node multi-relation networks, multiple relationships between nodes can enrich the semantics of network edges for subsequent clustering or community detection. Current approaches to combining different relationships follow two directions: The first is multi-relation combination (relation combination), where different relationship types between nodes are directly superimposed in the network based on research objectives. The second is multi-relation fusion (relation fusion), where multiple relationships are integrated into a new relationship before analyzing the relational characteristics of research objects.

3.1 Multi-Relation Combination Methods

Applying multi-relation combination methods to scientific structure analysis in disciplinary fields primarily includes: (1) Combination of citation and co-word relationships. The most representative work is H. Small' s 1998 study combining citation and co-word relationships to reveal both direct and indirect connections between documents, serving as a framework involving hierarchical clustering, cluster ranking, and common coordinate projection methods to support visualization of science mapping [14]. Other works include: C. Calero-Medina et al. (2008) used combined co-word and citation analysis to identify influential articles in a field over time, analyzing knowledge creation and flow processes in scientific publications by linking these articles to early influential traditional research, providing inspiration for subsequent multi-method integration studies [15]; Hou Yuefang et al. (2007) applied content word and co-citation clustering analysis, revealing both the development status of gestational diabetes research and validating clustering effectiveness, pioneering the application of combination methods to topical studies [16]; Zhang Han et al. (2007) combined co-word analysis with citation counts to comprehensively explore the development process of the peptic ulcer field using PubMed data [17], demonstrating that combining topical co-word analysis with citation frequency more easily detects disciplinary hotspots.

(2) Combination of collaboration and citation (including co-citation, bibliographic coupling) relationships. For example, K. Larsen et al. (2008) combined co-authorship and co-citation relationships to measure central points in knowledge networks of solar cell research, highlighting the importance of distinguishing between early and late stages of new research field development and the need for systematic observation of learning processes and knowledge dissemination in science and technology [18]; Chen Wei et al. (2014) constructed co-authorship and citation networks with China' s "985" universities as nodes, conducting joint analysis of basic structural characteristics, network correlation properties, community features, and important nodes, revealing the complexity characteristics and development trends of research collaboration networks among "985" universities [19], opening new perspectives for studying inter-university collaboration and citation.

These studies all combined two or more relationships, approaching from both the relationship and cognition dimensions to combine networks with lower similarity for richer information revelation.

3.2 Multi-Relation Fusion Methods

Unlike multi-relation combination methods, multi-relation fusion methods integrate multiple relationships, originating from webpage clustering and classification research. Based on different fusion stages, two types can be distinguished: One is cluster merging, where different data sources are clustered separately and then merged into new clusters through specific algorithms. The other is kernel fusion, where similarity or distance matrices from multiple sources are integrated into a new unified matrix for clustering or other multivariate statistical analysis.

In cluster merging research for single-type node multi-relation hybrid networks, X.X. Yin et al. (2015) proposed a simple semi-supervised method called CROSS-CLUS, which performs multiple clustering evaluations on multi-relational objects based on user-selected features relevant to clustering objectives [21]. L. Wei et al. (2015) used correlation analysis methods for multi-relational data, calculating distances between different clusters as distances between cluster centroids with assigned weights, ensuring both clustering efficiency and precision [22]. Ding Zhijun et al. (2017) proposed a partial multi-relation clustering method, a typical study of clustering ensemble relation fusion. This method clusters entities based on different relationships, weights the importance of different relationships according to clustering results, and finally integrates them into a single-relationship network for re-clustering. Experiments on multiple public datasets demonstrated its effectiveness in improving clustering precision [23]. These clustering methods have enhanced both efficiency and accuracy, providing more reliable and accurate approaches for scientific structure visualization through clustering.

In kernel fusion research for single-type node multi-relation hybrid networks, representative recent work involves hybrid clustering methods that comprehensively consider both textual and link attributes of academic papers, such as the series of studies around W. Glanzel's "citation-text" hybrid clustering algorithm integrating citation coupling and textual similarity [24-25]. These studies proved that hybrid clustering methods achieve higher accuracy than single-method community detection. First, Glanzel et al., drawing from webpage content and link analysis, combined word-based relationships with citation coupling relationships between documents, demonstrating this approach's effectiveness in revealing research field structures [26]. Second, Zhang Lin et al. used journal cross-citation-based clustering algorithms to validate and improve journal-based disciplinary classification schemes [27]. Furthermore, Glanzel's team integrated journal cross-citation with text mining to validate and enhance existing topical classification schemes [28]. Additionally, Wang Xiaomei's recently published series of "Science Structure Maps" also adopted Glanzel's team's hybrid clustering method.

In relation fusion research, Glanzel's team focused primarily on mining the complementary information of citation relationships and text, without exploring other pairwise independent relationships. For instance, how hybrid clustering performs between citation networks (based on citation relationships) and collaboration networks (based on social cognition) requires further investigation and exploration.

4. Community Detection in Multi-Type Node Multi-Relation Hybrid Networks

Multi-type node multi-relation networks are defined in contrast to traditional networks—networks containing multiple entity types and associations can be regarded as heterogeneous information networks. In library and information science, literature information networks are heterogeneous information networks with multiple entity types and relationships, primarily involving four entity types: articles, journals, authors, and keywords. Articles have relationships with journals, authors, and keywords. Since information flows differently among heterogeneous nodes and relationships compared to homogeneous networks, many homogeneous network-based analysis methods are unsuitable for heterogeneous information networks. Consequently, clustering or community detection research on such networks focuses primarily on algorithmic innovation and improvement. Currently, three main approaches exist for community detection in multi-type node multi-relation networks: ranking-based methods, meta-path-based methods, and heterogeneous-to-homogeneous conversion methods.

4.1 Ranking-Based Methods

Applying ranking methods to community detection or clustering allows ranking and clustering to complement each other. The first ranking-clustering algorithm for heterogeneous information networks was RankClus [29], which iteratively clusters and ranks different nodes in the network until clustering objectives become clear. Subsequently, many similar ranking-clustering algorithms emerged, such as NetClus [30], ENetClus [31], and ComClus [32]. NetClus primarily targets star-structured networks and can efficiently generate both clustering and ranking results. Zhao Huan improved the NetClus algorithm and proposed the MAO-Netclus algorithm for heterogeneous networks, performing clustering analysis on multi-type node multi-relation networks composed of three object types in Web service systems to improve Web service recommendations [33]. Tong Hao et al. proposed the RankCoClus algorithm for heterogeneous information networks based on ranking and co-clustering, with experimental results demonstrating superior clustering performance [34].

4.2 Meta-Path-Based Methods

Meta-path-based methods target link relationships, as different links in networks convey different information that influences clustering effectiveness. Different link paths in heterogeneous networks constitute different meta-paths. A representative method is PathSim, proposed by Y. Sun et al. in 2011 [35], which is a meta-path-based similarity measurement method. Since this method only calculates similarity for same-type nodes, subsequent methods emerged for measuring similarity between different node types, continuously improving algorithmic performance. Meta-path-based clustering methods have also proliferated, including PathSelClus [36], which studied the impact of different meta-paths on node clustering but requires strong assumptions in meta-path selection. The GenClus algorithm [37] is a clustering method considering link relationship strength, which determines node attributes and link relationships through user guidance and automatically learns to construct different link strengths to improve clustering effectiveness. Li Li proposed a heuristic search and pruning strategy based on meta-path methods to effectively select paths consistent with user guidance while avoiding information loss from breadth-first traversal searches. Building on this, Li Li extended homogeneous network community detection algorithms by proposing a community detection framework combining relation extraction with meta-path weighting, validating its effectiveness and accuracy on real datasets [38]. Wang Rui also proposed a weighted meta-path community detection algorithm HCD, which effectively identifies communities across multiple meta-paths and can detect overlapping communities [39].

4.3 Heterogeneous-to-Homogeneous Conversion Methods

Since community detection algorithms for homogeneous networks are relatively mature, dimensionality reduction and reconstruction of heterogeneous networks into homogeneous networks represent a feasible approach. Current dimensionality reduction methods for heterogeneous networks primarily include Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), Non-negative Matrix Factorization (NMF), and topic models. Reconstruction methods mainly involve transforming heterogeneous networks into bipartite graphs. Based on these methods, Wang Ting (2016) proposed an efficient and fast heterogeneous network community detection algorithm [40], which first reduces dimensions of heterogeneous social network data, then reconstructs the heterogeneous network as a bipartite graph. To prevent information loss during community detection, the method employs label propagation for community partitioning [37]. This approach is generalizable and can be extended to many practical scenarios.

Typical heterogeneous information network clustering algorithms include meta-path-based methods and ranking-based methods. The former eliminates cumbersome ranking iteration processes but requires user prior experience guidance, while the latter demands less prior knowledge but involves more complex iteration. Heterogeneous-to-homogeneous conversion methods are intuitive but involve complex processes. To observe scientific structures in complex heterogeneous networks, applying ranking-based methods, meta-path-based methods, or heterogeneous-to-homogeneous conversion methods during preliminary network processing is necessary.

5. Community Detection in Multi-Type Node Single-Relation Hybrid Networks

Multi-type node single-relation networks are characterized by diverse node types. Research on community detection for such networks is scarce but can provide preliminary references for community detection from a network construction perspective. Conducting community detection or clustering on such networks requires understanding the meaning of node types. Community detection or clustering in these networks better facilitates clear judgment of inheritance and subordination relationships in scientific structure formation, representing a potential future research focus.

5.1 Nodes with Multiple Attributes

In some networks, multi-type nodes essentially reflect multiple attributes of entities. For instance, in collaboration networks, nodes are typically authors or researchers without distinguishing their social attributes. However, strictly speaking, researchers have multiple attributes, including article attributes (keywords, topics, etc.), characteristic attributes (age, title, etc.), and social attributes (student, faculty). Wang Yan et al. explored theories and methods for expert academic networks using different attributes of scholars, constructing monograph expert collaboration networks, expert topic networks, and patent expert collaboration networks based on multi-source data to more accurately characterize explicit and implicit collaboration networks among experts [41]. Lei Xue et al. distinguished first authors from other authors based on contribution degree, constructing directed collaboration networks with two node types and comparing them with traditional undirected collaboration networks to explore more effective research analysis methods [42]. Tan Zongying et al., when studying international collaboration, differentiated between leading and other countries, constructing a China-led directed international collaboration network and conducting topical content analysis [43].

5.2 Nodes Representing Multiple Entities

In some networks, multi-type nodes essentially represent different entities. Wang Peng et al., when studying university-industry collaboration networks, constructed relationship networks between researchers and nano-related patents, revealing the topological structure of the Tsinghua University-centered nano-technology collaboration network [44]. Ma Yanyan et al. further expanded research objects, depicting university-enterprise patent application collaboration networks using Chinese university and enterprise patent application data, and analyzing network characteristics to find substantial room for improvement in China' s university-industry collaboration [45].

It is evident that research on multi-type node single-relation networks has primarily focused on collaboration networks. Currently, research on constructing and conducting community detection in multi-type node single-relation networks remains relatively scarce. However, community detection or clustering in such networks better facilitates clear judgment of inheritance and subordination relationships in scientific structure formation, representing a potential future research focus.

6. Brief Introduction to Algorithms Supporting Hybrid Network Community Detection

Community detection methods are crucial for studying complex network structures. In 2002, M. Girvan and M.E.J. Newman proposed a divisive algorithm—the GN algorithm [3]—sparking a surge in community research. The GN algorithm achieves community detection by continuously removing edges with the highest betweenness. From another perspective, M.E.J. Newman proposed an agglomerative greedy algorithm [46], where each node initially forms an independent cluster that progressively merges during the detection process. Subsequently, to evaluate community detection quality, Newman and Girvan introduced the modularity function Q in 2004 [47], where higher Q values generally indicate better community partitions. To address inefficiency in large network community discovery, Blondel et al. [48] proposed the Louvain community detection algorithm in 2008, which Rotta and Noack optimized in 2011 with multilevel refinement [49]. Building on this, Waltman and Eck improved and proposed the SLM algorithm in 2013, characterized by allowing already-partitioned nodes to be re-assigned to communities [11].

For detailed introductions and comparative studies of these community detection algorithms, readers may refer to works by Shi Jingjing [50] and Chen Yunwei and Zhang Ruihong [51].

7. Discussion and Future Directions

This paper has reviewed the latest research progress on hybrid network community detection methods for scientific structure analysis in library and information science. For single-type node multi-relation hybrid networks, two approaches exist for handling multiple relationships: multi-relation combination and multi-relation fusion. Multi-relation combination is relatively simple, achieved by selecting two or more analytical methods. However, selection should be evidence-based, requiring scientific evaluation of different combination effects, preferably choosing methods from different dimensions. Multi-relation fusion methods concentrate on innovating hybrid network construction or community detection algorithms. However, relationship fusion is complex work, requiring further research on which relationships to fuse and how to evaluate fusion effectiveness.

For more complex multi-type node multi-relation networks, current research remains relatively limited due to diverse node attributes and complex relationships. Key research challenges include exploring principles and methods for mining multi-type node multi-relation networks, accurately constructing real-world models, and discriminating the importance of multiple nodes or relationships. The research frontier extends beyond network construction and community detection/clustering exploration to information diffusion, semantic search, and intelligent query. Due to the difficulty of mining heterogeneous information networks, such research is more challenging and practically valuable, representing an important future direction for information network research.

In revealing scientific structure, single-network community detection research is relatively mature, while hybrid network community detection research is still developing. For hybrid networks that break traditional single-network limitations, subsequent network analysis research is provided with new perspectives and methods, capable of mining rich information hidden among different links between entities, representing a completely new theoretical and practical advancement. Meanwhile, hybrid network community detection still has many issues worth exploring in analyzing scientific structure, describing knowledge development, and analyzing disciplinary intersections. Scientific research is a complex system, and the data discussed in this paper are all literature-based, representing only part of research output. Scientific research also involves substantial information related to science and technology strategies, planning, projects, funding, etc., all closely related to scientific structure. Future research can expand the data foundation, utilizing richer data types and relationship types from more comprehensive perspectives to fully understand and reveal scientific structure.

References

[1] Wei Junchao, Wei Haiyan. Review of methods for scientific structure and evolution analysis [J]. Library and Information, 2011(4): 48-52.

[2] Zhao Hongzhou. On scientific structure [J]. Academic Journal of Zhongzhou, 1981(3): 59-65, 133.

[3] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks [J]. PNAS, 2002, 99(12): 7821-7826.

[4] LAMBIOTTE R, PANZARASA P. Communities, knowledge creation, and information diffusion [J]. Journal of informetrics, 2009, 3(3): 180-190.

[5] MOLINER L A, GALLARDO-GALLARDO E, PUELLES P G D. Understanding scientific communities: a social network approach to collaborations in talent management research [J]. Scientometrics, 2017, 113(3): 1439-1462.

[6] ZHENG J, GONG J Y, LI R, et al. Community evolution analysis based on co-authorship network: a case study of academic communities in the journal of "Annals of the American Geographers"[J]. Scientometrics, 2017, 113(2): 845-865.

[7] Chen Yunwei. Analysis of research progress on citation network evolution [J]. Information Science, 2016, 34(8): 171-176.

[8] Qiu Junping, Dong Ke. Deep aggregation methods and empirical research of literature in citation networks—taking XML research papers in WOS database as an example [J]. Journal of Library Science in China, 2013, 39(2): 111-120.

[9] KAJIKAWA Y, YOSHIKAWA J, TAKEDA Y, et al. Tracking emerging technologies in energy research: toward a roadmap for sustainable energy [J]. Technological forecasting and social change, 2008, 75(6): 771-782.

[10] CHEN Y W, XIAO X, DENG Y, et al. A weighted method for citation network community detection [EB/OL]. [2018-05-20]. http://ir.csdl.ac.cn/handle/12502/9556?mode=full&submit_

[11] WALTMAN L, ECK N J V. A smart local moving algorithm for large-scale modularity-based community detection [J]. European physical journal b, 2013, 86(11): 471.

[12] Sun Yizhou, Han Jiawei. Mining heterogeneous information networks: principles and methods [M]. Beijing: Mechanical Industry Press, 2016.

[13] TASKAR B, ABBEEL P, KOLLER D. Discriminative probabilistic models for relational data [J]. Algorithmic bioinformatics, 2002, 7(7): 485-492.

[14] SMALL H. A general framework for creating large-scale maps of science in two or three dimensions: the sciviz system [J]. Scientometrics, 1998, 41(1/2): 125-133.

[15] CALERO-MEDINA C, NOYONS E C M. Combining mapping and citation network analysis for a better understanding of the scientific development: the case of the absorptive capacity field [J]. Journal of informetrics, 2008, 2(4): 272-279.

[16] Hou Yuefang, Cui Lei, Wu Di. Application of co-citation clustering-content word analysis to study disciplinary development [J]. Journal of the China Society for Scientific and Technical Information, 2007, 26(2): 309-314.

[17] Zhang Han, Wang Xiaoyu, Cui Lei. Combining co-word analysis with citation counts to study development trends in topical fields [J]. Information Studies: Theory & Application, 2007, 30(3): 378-380, 426.

[18] LARSEN K. Knowledge network hubs and measures of research impact, science structure, and publication output in nanostructured solar cell research [J]. Scientometrics, 2008, 74(1): 123-142.

[19] Chen Wei, Zhou Wen, Lang Yifu, et al. Research collaboration network analysis based on co-authorship and citation networks [J]. Information Studies: Theory & Application, 2014, 37(10): 54-59.

[20] YAN E, DING Y. Scholarly network similarities: how bibliographic coupling networks, citation networks, co-citation networks, topical networks, co-authorship networks, and co-word networks relate to each other [J]. Journal of the American Society for Information Science and Technology, 2012, 63(7): 1313-1326.

[21] YIN X X, HAN J, YU P S. CrossClus: user-guided multi-relational clustering [J]. Data mining and knowledge discovery, 2007, 15(3): 321-348.

[22] WEI L, LI Y. Multi-relational clustering based on relational distance [C]//Web information system and application conference. Jinan: IEEE, 2016: 297-300.

[23] Zeng Yanyu, Ding Zhijun. Partial multi-relational clustering method considering importance weighting [J]. Small Microcomputer Systems, 2017, 38(6): 1227-1230.

[24] GLÄNZEL W, JANSSENS F, THIJS B. Combining full text and bibliometric information in mapping scientific disciplines [J]. Information processing & management, 2005, 41(6): 1548-1572.

[25] JANSSENS F, GLÄNZEL W, DE MOOR B. A hybrid mapping of information science [J]. Scientometrics, 2008, 75(3): 607-631.

[26] JANSSENS F. Clustering of scientific fields by integrating text mining and bibliometrics [EB/OL]. [2017-12-09]. https://repository.libis.kuleuven.be/dspace/bitstream/1979/847/5/PhD_

[27] ZHANG L, JANSSENS F, LIANG L M, et al. Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research [J]. Scientometrics, 2010, 82(3): 687-706.

[28] JANSSENS F, ZHANG L, DE MOOR B D, et al. Hybrid clustering for validation and improvement of subject-classification schemes [J]. Information processing & management, 2009, 45(6): 683-702.

[29] SUN Y, HAN J, ZHAO P, et al. RankClus: integrating clustering with ranking for heterogeneous information network analysis [C]//Proceedings of the 12th international conference on extending database technology: advances in database technology. Petersburg: ACM, 2009: 565-576.

[30] BARALIS E, BIANCO A, CERQUITELLI T, et al. NetCluster: a clustering-based framework for internet tomography [C]//IEEE international conference on communications. Dresden: IEEE, 2009: 1-6.

[31] GUPTA M, AGGARWAL C C, HAN J, et al. Evolutionary clustering and analysis of bibliographic networks [EB/OL]. [2018-05-20]. http://charuaggarwal.net/asonam-cluster.pdf.

[32] WANG R, SHI C, YU P S, et al. Integrating clustering and ranking on hybrid heterogeneous information network [C]//Pacific-Asia conference on knowledge discovery and data mining. Gold Coast: Springer, 2013: 583-594.

[33] Zhao Huan. Research on Web service recommendation system based on heterogeneous network clustering [D]. Chongqing: Chongqing University, 2015.

[34] Tong Hao, Yu Chunyan. Co-clustering algorithm for heterogeneous information networks based on ranking distribution [J]. Small Microcomputer Systems, 2014, 35(11): 2445-2449.

[35] SUN Y, HAN J, YAN X, et al. Pathsim: meta-path-based top-k similarity search in heterogeneous information networks [J]. Proceedings of the VLDB endowment, 2011, 4(11): 992-1003.

[36] SUN Y, NORICK B, HAN J, et al. Pathselclus: integrating meta-path selection with user-guided object clustering in heterogeneous information networks [J]. ACM transactions on knowledge discovery from data (TKDD), 2013, 7(3): 1-23.

[37] SUN Y, AGGARWAL C C, HAN J. Relation strength-aware clustering of heterogeneous information networks with incomplete attributes [J]. Proceedings of the VLDB endowment, 2012, 5(5): 394-405.

[38] Li Li. Community mining algorithm for heterogeneous information networks based on meta-path selection and fusion [D]. Xi' an: Xidian University, 2014.

[39] Wang Rui. Research and implementation of community detection in heterogeneous information networks [D]. Beijing: Beijing University of Posts and Telecommunications, 2017.

[40] Wang Ting. Research on community detection algorithms in heterogeneous social networks [D]. Beijing: China University of Mining and Technology (Beijing), 2016.

[41] Wang Yan, Wei Ruibin. Research on constructing expert academic networks based on multi-source data [J]. Journal of Intelligence, 2016, 35(12): 121-126, 138.

[42] Lei Xue, Wang Lixue, Zeng Jianxun. Construction and analysis of directed co-authorship networks [J]. Library and Information Service, 2015, 59(5): 94-99.

[43] Lu Jingjing, Tan Zongying, Liu Xiaoling, et al. Construction and analysis of country-led collaboration networks in international cooperation [J]. Journal of Intelligence, 2015, 34(12): 60-66, 100.

[44] Wang Peng, Zhang Xu, Zhao Yunhua, et al. Complex network analysis of university-enterprise research collaboration [J]. Information Studies: Theory & Application, 2010, 33(6): 89-93.

[45] Ma Yanyan, Liu Fengchao, Sun Yutao. Research on university-enterprise patent application collaboration networks in China [J]. Studies in Science of Science, 2011, 29(3): 390-395, 332.

[46] NEWMAN M E. Fast algorithm for detecting community structure in networks [J]. Physical review E, 2004, 69(6): 066133.

[47] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in networks [J]. Physical review E, 2004, 69(2): 026113.

[48] BLONDEL V, GUILLAUME J, LAMBIOTTE R, et al. Fast unfolding of communities in large networks [J]. Journal of statistical mechanics: theory and experiment, 2008(10): 155-168.

[49] ROTTA R, NOACK A. Multilevel local search algorithms for modularity clustering [J]. Journal of experimental algorithmics, 2011, 16(2): Article No. 2.3.

[50] Shi Jingjing. Research on three classical complex network community structure partitioning algorithms [J]. Computer and Information Technology, 2011, 19(4): 32-43, 79.

[51] Chen Yunwei, Zhang Ruihong. Comparative study of typical network community detection algorithms for intelligence mining [J]. Data Analysis and Knowledge Discovery, 2018, 2(10): 84-94.

Author Contributions:

Zhang Ruihong: Collected and organized materials, wrote and revised the paper;
Chen Yunwei: Proposed the review framework and revision suggestions;
Deng Yong: Proposed modifications to the article structure and writing details.

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