Survey on Community Detection in Multi-node Multi-relational Hybrid Networks (Postprint)

Authors: Jiang Lu, Chen Yunwei

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

Abstract

[Purpose/Significance] This study aims to systematically review community detection methods for multi-node multi-relationship heterogeneous networks, analyze the problems and difficulties in existing methods, and foresee future development trends. [Method/Process] We systematically survey recent research on community detection methods for heterogeneous networks with multiple node types and relationship types, elaborating on the detection approaches from five perspectives: probabilistic generative models, meta-paths, seed nodes, extended modularity, and heterogeneous network homogenization. We summarize commonly used evaluation metrics for community detection in heterogeneous networks, including Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and modularity Q, and identify three application scenarios: social media, academic networks, and fraud detection. [Results/Conclusion] We summarize the applicability, advantages, and disadvantages of community detection methods for multi-node multi-relationship heterogeneous networks, reveal the current development challenges, provide new perspectives for subsequent heterogeneous network analysis research, and prospect potential research directions for future expansion.

Full Text

A Review of Community Detection in Hybrid Networks with Multiple Nodes and Relationships

Jiang Lu[1,2], Chen Yunwei[1,2]

[1]Science Evaluation Research Center (SERC), Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041, China
[2]Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

Abstract: [Purpose/Significance] This paper aims to systematically review community detection methods for multi-node, multi-relationship hybrid networks, analyze existing problems and difficulties in current methods, and anticipate future development trends. [Method/Process] We systematically review recent research on community detection methods for hybrid networks with multiple node types and relationship types, elaborating on these methods from five perspectives: probabilistic generative models, meta-paths, seed nodes, extended modularity, and hybrid network isomorphism. We summarize commonly used evaluation metrics for hybrid network community detection: Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and modularity Q, and identify three application scenarios: social media, academic networks, and fraud detection. [Result/Conclusion] We summarize the applicability, advantages, and disadvantages of community detection methods for multi-node, multi-relationship hybrid networks, reveal current development challenges, provide new perspectives for subsequent hybrid network analysis research, and outline promising directions for future expansion.

Keywords: multi-node; multi-relationship; hybrid network; community; evaluation

Network science has emerged as one of the most active research fields in recent years [1], with successful applications in numerous domains including social sciences studying interpersonal relationships [2], biology examining gene and protein interactions [3], and neuroscience investigating brain structure and function [4]. Existing research reveals that entire networks are composed of several "communities." A community, also called a cluster or module, is a set of vertices with common attributes or identical roles in the network [5]. Connections within each community are relatively dense, while connections between communities are relatively sparse [6-7]. Identifying community structure not only reveals similarities between nodes but also uncovers the operating principles within communities, helping to understand network structural characteristics and latent semantic information. Therefore, community detection is considered a fundamental approach for understanding and analyzing networks [8], enabling the prediction of community structures from observed network topology and node attribute information [9] and clustering densely connected nodes into communities [5,10].

Currently, few scholars have systematically reviewed community detection methods for multi-node, multi-relationship hybrid networks; most reviews focus on homogeneous networks [11-12]. Li et al. [13] surveyed community detection methods in complex networks from five perspectives: modularity optimization, label propagation, local expansion, streaming analysis, and deep learning, but most discussed methods only apply to single-node, single-relationship networks. Zhang et al. [14] separately reviewed community detection methods for single-node single-relationship, single-node multi-relationship, and multi-node multi-relationship hybrid networks, but their review of multi-node multi-relationship methods was incomplete, unsystematic, and failed to clearly identify development trajectories, advantages, and disadvantages. Therefore, this paper focuses on multi-node, multi-relationship hybrid networks, retrieving relevant domestic and international literature. First, we clarify the definitions of multi-node type, multi-relationship type hybrid networks and conceptually differentiate them from other networks. Second, we introduce community detection methods and main evaluation metrics for multi-node, multi-relationship hybrid networks, revealing the development process, advantages, disadvantages, and applicability of each algorithm type to provide researchers with a clearer and more comprehensive understanding of the field and theoretical foundations for deeper algorithm research. Finally, we elaborate on application scenarios and development challenges, offering insights for constructing community detection algorithms with strong network applicability, low complexity, and integration of network topology and textual information, while also prospecting future directions for dynamic network community detection methods and standard evaluation metrics for hybrid networks.

1. Definitions of Various Networks

Numerous network definitions have emerged over time, with various formulations summarized in Table 1 .

Complex networks are characterized by self-organization, self-similarity, attractors, small-world properties, and scale-free characteristics [15], representing networks with massive scale, complex connections, and heterogeneous nodes [16]. Heterogeneous networks are defined as networks where the number of object types $|A| > 1$ and relationship types $|R| > 1$ [17]. Homogeneous networks, corresponding to heterogeneous networks, contain only one node type and one relationship type [18]. Hybrid networks contain multiple node types or multiple relationship types [14].

Both complex networks and hypernetworks refer to broad categories emphasizing high complexity without specifying particular nodes and relationships. Heterogeneous networks emphasize topological complexity, focusing more on heterogeneous relationships while paying less attention to homogeneous relationships between similar object types [19]. Homogeneous networks only reflect connections from a single perspective, limiting their utility in studying scientific structures, identifying research fronts, and detecting technological opportunities [20]. Hybrid networks essentially belong to the category of heterogeneous networks but emphasize the mixing of multiple nodes and multiple relationships, fully considering both heterogeneous and homogeneous relationships and demonstrating functional richness.

Therefore, this study adopts the "hybrid network" definition to shift researchers' focus from network topology construction to functional enhancement while maintaining alignment with real-world networks.

Multi-node type, multi-relationship type hybrid networks contain multiple node types and relationship types, with characteristics manifested in two aspects: (1) Node diversity—multiple node types exist. In academic networks, nodes can be authors, papers, keywords, journals, etc.; in healthcare networks, nodes can be doctors, medicines, patients, etc. (2) Relationship richness—in academic networks, relationships can include author collaboration, author citation, paper citation, and author-paper affiliation; in healthcare networks, relationships can include doctor prescription and patient medication relationships. Multi-node, multi-relationship hybrid networks integrate multiple perspectives and relationship types, leveraging the rich semantics of multi-type nodes and links to discover knowledge from interconnected data and capture fundamental semantic information from the real world [21]. Therefore, comprehensively understanding scientific structure information within a field necessitates in-depth research on multi-node, multi-relationship hybrid networks.

2. Community Detection in Multi-Node Multi-Relationship Hybrid Networks

Research on community detection for multi-node, multi-relationship hybrid networks remains insufficient. Current research concentrates on two approaches: (1) Extending existing algorithms to directly process hybrid networks, and (2) Reducing hybrid networks to homogeneous networks before performing community detection [23-25]. Based on these two approaches, community detection methods for multi-node, multi-relationship hybrid networks mainly fall into five categories:

2.1 Methods Based on Probabilistic Generative Models These methods include ranking-based approaches and probabilistic statistical model approaches.

2.1.1 Ranking-Based Methods

In probabilistic generative model methods, some algorithms combine ranking problems with community detection: good ranking enhances community detection results, and good communities improve ranking [26]. RankClus [27] was the earliest ranking-based clustering algorithm for hybrid networks, but it only applied to two node types. Y. Z. Sun et al. proposed NetClus [28] based on RankClus, which generates high-quality network clusters using links between multi-type nodes and achieves better clustering results, but it only applies to star network structures and requires prior knowledge of representative objects in the dataset.

Since the above algorithms lack universality, M. Ji et al. proposed the RankClass [29] algorithm for hybrid networks with arbitrary schemas that can fully utilize label information of any data object. Zhao Huan improved the classic NetClus algorithm and proposed the MAO-NetClus algorithm, which achieves Web service clustering based on multi-node, multi-relationship hybrid networks for three node types (Web services, providers, users) and designs a Web service recommendation system prototype [30]. Additionally, for dynamic hybrid network community detection revealing evolution processes of each node type, M. Gupta et al. [31] proposed the ENetClus algorithm, which performs evolutionary clustering using temporal smoothing to show clustering changes over time. C. H. Qiu et al. [32] proposed the OcdRank algorithm, which supports incremental data updates with low time complexity.

2.1.2 Probabilistic Statistical Model Methods

Ranking-based methods require pre-setting the number of communities, leading to instability. Therefore, scholars proposed using probabilistic statistical models for community detection. These methods employ Bayesian models, prior probabilities, and posterior probabilities to calculate the probability of nodes belonging to communities. Chen Yi [33] proposed a multi-dimensional Bayesian nonparametric mixture model (MBNPM) that fuses structural features captured from each dimension and obtains community information through clustering models, automatically exploring the number of communities and achieving superior results. Yin Haoxiao et al. proposed the DirCom method based on information dimension statistics in hybrid networks, which learns Dirichlet distribution parameters on information dimensions to characterize communities after information dimension rollup and achieves community detection using maximum a posteriori probability [34]. S. Sengupta and Y. Chen proposed a spectral clustering method for hybrid networks based on stochastic block models, applying variational EM algorithms for posterior inference in large networks that allows different node types to have multiple memberships [35], but this algorithm does not solve the overlapping community problem.

These algorithms only target hybrid networks containing heterogeneous relationships, but real-world networks are more complex, also containing homogeneous relationships between similar node types. For such networks, Tong Hao et al. proposed the RankCoClus algorithm based on RankClus, combining ranking clustering with co-clustering methods, selecting four node types (papers, authors, terms, conferences) and two relationship types (conference-author, author-author), with experiments demonstrating its effectiveness [36]. R. Wang et al. proposed the ComClus algorithm, which organizes hybrid networks using a star schema with self-loops and employs probabilistic models to represent object generation probabilities [37], with experiments showing superior clustering results.

In summary, ranking-based methods have low time complexity and enable dynamic network community detection but require prior knowledge to specify community numbers, making accurate prediction difficult for large networks and causing result instability. Probabilistic statistical model methods do not require prior knowledge, offer strong stability, and are suitable for large networks but do not solve overlapping community problems.

2.2 Methods Based on Meta-Paths Multiple node types are connected by multiple links, each containing different semantics, forming meta-paths [38]. Meta-paths are effective semantic capture tools that can capture rich semantic information within hybrid networks [39-40] and represent unique features and feature extraction methods for hybrid networks [41]. Therefore, meta-path-based community detection methods have emerged successively for multi-node, multi-relationship hybrid networks.

PathSim [40] was the earliest meta-path-based algorithm, proposed for homogeneous networks, performing well in measuring similarity between similar node types. J. Li et al. noted that most meta-path-based hybrid network community detection methods have two problems: (1) similarity obtained directly from meta-paths is often a biased measure, and (2) fusing similarities from different meta-paths is challenging [42]. Therefore, they proposed a flexible fusion mechanism to dynamically optimize results based on PathSim similarity standardization to eliminate bias, achieving better community detection. C. Shi et al. [43] proposed the HetSim algorithm based on meta-paths to measure similarity between similar or different node types, using bidirectional random walks to calculate similarity and outperforming traditional algorithms in query and clustering tasks. However, HetSim only applies to single meta-path environments, cannot capture multiple semantic information in hybrid networks [44], and has high complexity unsuitable for large-scale networks. Subsequently, X. F. Meng et al. [45] proposed a dual random walk process based on given meta-paths and reverse meta-paths to calculate similarity between two objects—the AvgSim algorithm—which can be applied in large-scale networks with excellent clustering results.

Different meta-paths contain different information, and selecting different meta-paths leads to different community detection results. Determining the number of meta-paths to select or identifying optimal meta-paths remains challenging. Y. Z. Sun et al. [46] proposed the PathSelClus algorithm, which can assign different weights to different meta-paths in hybrid networks. Wu Yao et al. proposed a multi-graph fusion hybrid network embedding method that can automatically learn key meta-paths in networks [47]. C. Shi [41] introduced the HRank meta-path-based random walk method to evaluate node and meta-path importance, with experimental results demonstrating meta-paths' unique advantages.

Evidently, meta-path-based methods in hybrid networks are mostly improved from homogeneous network PathSim methods. These methods are simple and intuitive, with multiple meta-paths capturing rich information in hybrid networks, but they suffer from high algorithmic complexity and often produce biased similarity measures [42], making them less suitable for large-scale networks. Moreover, different meta-paths contain different information, making it challenging to accurately calculate node similarity to reveal rich semantic relationships and to select optimal meta-paths for optimal partitioning results.

2.3 Methods Based on Seed Nodes Seed-centric approaches have become an emerging trend in community detection algorithms [48]. The basic idea is to identify certain specific nodes called seed nodes and build communities around them [49-51].

Z. Yakoubi et al. first proposed the seed-driven community detection algorithm Licod, which selects nodes with higher centrality than most direct neighbors as seed nodes, performs local community calculations around these nodes, and then partitions from the local community set [52], but this method only applies to homogeneous networks. M. Hmimida et al. [48] extended the Licod algorithm to hybrid networks, called mux-Licod, which considers different relationship types between nodes in different layers, with experimental results showing good practicality. Xue Weijia proposed the NS-Clus algorithm based on seed node clustering, which selects seed nodes according to node importance and second-order neighbors, performs initial community partitioning through similarity measures, and adds non-seed nodes to communities based on node membership probability to obtain final partitioning results, demonstrating effectiveness on DBLP and ACM datasets [38].

Seed node-based methods are local computation methods that are easy to understand and suitable for processing large-scale and dynamic networks [52]. However, no consensus exists on how to efficiently select effective seed nodes, and merging non-seed node communities causes problems of excessive merging of large communities and too many small communities.

2.4 Extended Modularity Algorithms Modularity was originally proposed as an indicator to evaluate community detection results. With deepening research, community detection algorithms based on modularity emerged [6,53-54]. Extended modularity algorithms expand modularity algorithms applicable to homogeneous networks to hybrid networks.

M. E. J. Newman et al. first proposed the FN modularity optimization algorithm, which treats each node as a community, calculates modularity values after combining communities pairwise, and iteratively combines communities with maximum modularity increase or minimum decrease until modularity no longer increases [6], but this only applies to single-node-type networks. R. Guimerà et al. proposed an extended modularity algorithm for bipartite networks that can independently identify nodes with similar output connections and nodes with similar input connections [55], but it lacks universality. T. Murata et al. [56] proposed a modularity algorithm for k-core networks that suffers from the resolution limit problem common to general modularity algorithms and does not apply to general-form hybrid networks. X. Liu et al. [57] proposed a composite modularity method that decomposes hybrid networks into multiple sub-networks, integrates modularity across sub-networks, and optimizes composite modularity based on the Louvain algorithm to achieve community detection without requiring prior knowledge and applicable to large-scale general-form networks.

Obviously, extended modularity algorithms evolved from homogeneous network algorithms, offering high stability and suitability for large-scale networks. However, they suffer from poor network applicability and cannot avoid modularity maximization limitations—the resolution limit—preventing detection of small communities in large networks [5,58-59].

2.5 Hybrid Network Isomorphism Methods Since community detection methods for homogeneous networks are relatively mature, hybrid networks can be reduced to homogeneous networks before applying homogeneous network community detection methods. Dimensionality reduction methods mainly include Non-negative Matrix Factorization (NMF) [60], topic models [14], Principal Component Analysis (PCA) [61], and Linear Discriminant Analysis (LDA) [62].

2.5.1 Non-negative Matrix Factorization Methods

Non-negative matrix factorization can decompose any given non-negative matrix into two non-negative matrices: a basis matrix and a coefficient matrix, using the coefficient matrix to replace the original matrix for dimensionality reduction. S. Tafavogh proposed a hybrid network community detection method based on matrix factorization and semantic paths [64], demonstrating its effectiveness. X. C. Zhang proposed a non-negative matrix tri-factorization method, HMFClus, which integrates information between similar object types into HMFClus using similarity regularization and can simultaneously cluster all object types in hybrid networks [65]. Huang Ruiyang et al. discovered community structures in dynamic hybrid networks by fusing multi-relationship similarity matrices with non-negative matrix factorization models, but this algorithm is effective yet complex [66]. J. Liu et al. proposed a penalized alternating factorization (PAF) algorithm from a matrix factorization perspective for multi-layer attributed networks, which not only achieves good community detection results but also demonstrates strong network morphology applicability [67].

2.5.2 Topic Model Methods

Introducing topic models can mine hidden topic information in textual data to improve community detection results [68]. Q. Z. Mei et al. fully utilized the advantages of statistical topic models and discrete regularization, achieving community detection through regularized improved topic models [69]. Wang Ting proposed the topic-aware LDA-light algorithm, which reduces hybrid networks to homogeneous or bipartite networks and uses label propagation for community detection, producing communities with semantic information and strong universality applicable to practical scenarios [70].

2.5.3 Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)

Both methods belong to linear dimensionality reduction approaches that use linear projection to map high-dimensional data to low-dimensional space. The difference is that PCA ensures reduced data retains more original information, while LDA makes reduced data more easily distinguishable [71]. Existing research only applies these two methods to single-node-type networks [72-74] or bipartite networks [75].

Hybrid network isomorphism methods are easy to understand, but the process of reducing hybrid networks to homogeneous networks is complex and prone to information distortion. Non-negative matrix factorization methods offer strong network applicability but suffer from excessive complexity. Topic model methods utilize semantic information for more reliable results with strong universality. PCA and LDA methods have poor network applicability.

Increasingly, research is not limited to a single community detection method, and fusing multiple methods yields better results. Gao Changjie et al. calculated node similarity using a semantic-based meta-path model and obtained community detection results by minimizing objective function values [77]. Chen Changgeng proposed the PathLPA algorithm—a label propagation algorithm based on meta-path similarity calculation [76]—and applied it to DBLP hybrid networks for author node community detection, achieving good results. Zhang Zhenglin proposed the Hete_{MESC} overlapping community detection algorithm based on meta-path extraction and seed communities, where users select central nodes, extract multi-way networks about central nodes for community detection, use partitioning results as seed communities, and finally achieve community detection for all nodes based on similarity between other node types and seed communities [26], applicable to any network morphology with low complexity.

In summary, existing community detection methods for multi-node, multi-relationship hybrid networks are mostly based on probabilistic models and meta-paths, while seed node, extended modularity, and hybrid network isomorphism methods remain exploratory. Furthermore, various methods still have many unresolved issues, indicating substantial room for further research on multi-node, multi-relationship hybrid network community detection methods.

3. Commonly Used Evaluation Metrics for Community Detection

Numerous evaluation metrics exist for community detection, with different metrics used for different experimental requirements. This paper introduces the three most widely used metrics in community detection research: Normalized Mutual Information (NMI) [76], Adjusted Rand Index (ARI) [78], and modularity Q [6]. NMI and ARI evaluate against known ground-truth community partitions, while modularity Q evaluates without ground truth. Their comparison is shown in Table 2 .

Normalized Mutual Information (NMI) is a similarity measure based on information theory and probability theory [76], commonly used to detect differences between actual and ground-truth partitions, intuitively showing community detection quality. The NMI formula is as follows:

NMI = [formula with proper formatting]

where A and B are partition result sets, N is the total number of nodes, C_A and C_B represent the number of communities in A and B respectively, N_{ij} is the number of shared nodes between two communities, and N_i (N_j) is the sum of elements in row i (column j) of N. NMI ranges from [0,1], with higher values indicating more accurate community detection.

ARI broadly measures the agreement between two data distributions by comparing whether each node pair remains consistent across different community partitions [79]. The definition is:

ARI = [formula with proper formatting]

where a_{11} represents the number of node pairs in the same community in both ground-truth and actual partitions, a_{00} represents node pairs in different communities in both partitions, a_{10} represents node pairs in the same ground-truth community but different actual communities, and a_{01} represents node pairs in different ground-truth communities but the same actual community [79]. ARI ranges from [-1,1], with higher values indicating greater consistency between actual and ground-truth partitions. Compared with NMI, ARI has higher discriminability.

The modularity function Q was proposed by M. E. J. Newman and M. Girvan. Optimizing modularity Q yields better community detection results, as it makes connections within communities denser, making it a community strength metric [38]. The definition is:

Q = [formula with proper formatting]

where i and j are arbitrary nodes, k_i and k_j are node degrees, m is the total number of edges, A_{ij} = 1 when nodes are directly connected and 0 otherwise, and δ(C_i, C_j) = 1 when nodes belong to the same community and 0 otherwise. Q ranges from [0,1], with higher values indicating more stable and effective community structures.

NMI, ARI, and Q are the most commonly used community detection evaluation metrics, but their effectiveness for multi-node, multi-relationship hybrid networks requires systematic and in-depth validation. In hybrid network community detection evaluation, some scholars use custom metrics such as paper keyword relevance, paper topic relevance, and author relevance [26]. Evidently, the field still lacks a standardized evaluation metric that simultaneously considers connection strength between similar and different node types. Developing such an evaluation metric represents a future research direction.

4. Applications of Community Detection in Multi-Node Multi-Relationship Hybrid Networks

Community detection research in multi-node, multi-relationship hybrid networks has both theoretical significance and practical feasibility and effectiveness. Researchers have applied community detection methods across various domains to discover community structures and solve practical problems. This paper selects three common domains—social media, academic networks, and fraud detection—to elaborate on commonly used community detection methods.

4.1 Social Media The rapid development of social media networks has created networks with numerous nodes and complex relationships. Community detection in these networks has practical significance for friend recommendation, public opinion monitoring, and understanding communities at the network level and associating them with real life [80]. A key task in social networks is recommendation systems, while community detection groups like-minded individuals. Probabilistic generative model methods are most commonly used in community structure-based recommendation systems, identifying similar users and making precise recommendations based on common characteristics to optimize collaborative filtering issues such as data overload and low recommendation efficiency [81]. Chen Yi applied the Bayesian nonparametric mixture model BNPM method to friend recommendation [33], achieving better results and higher efficiency compared with traditional friend recommendation algorithms.

4.2 Academic Networks As scientific structure research deepens, constructing hybrid networks about authors, papers, keywords, and relationships such as author collaboration, paper citation, and paper-keyword affiliation, then performing community detection, can reveal more scholar structure information and show community structures from different perspectives. This provides a basis for comprehensively and clearly revealing scientific communities, research structures, and disciplinary development trajectories, becoming a new perspective in academic network research. Many scholars have applied probabilistic generative models, meta-paths, and seed node methods to the DBLP academic literature dataset to validate their proposed multi-node, multi-relationship hybrid network community detection algorithms [38,82-83]. Zhang Zhenglin constructed a hybrid network containing four node types (papers, authors, keywords, journals) and four relationship types (paper citation, paper-author authorship, paper-keyword inclusion, paper-journal publication), and after community detection based on meta-path extraction and seed nodes, compared author communities and paper communities, finding that "paper communities are smaller with single research fields, while author communities are larger with dispersed research fields" [26].

4.3 Fraud Detection Fraud detection has enormous applications in real life such as telecommunications networks and healthcare. Various fraud detection scenarios involve numerous nodes, large data volumes, and uneven distribution. Traditional anomaly detection methods struggle to detect anomalies, while multi-node, multi-relationship hybrid network community detection methods effectively simplify problems while focusing more on node relationships, providing new directions for fraud detection. Extended modularity algorithms are most commonly used in this scenario [84-85]. Luan Tingting modeled doctors and medicines in general inpatient data as a hybrid weighted network, proposed the FNO modularity optimization algorithm to partition doctors and medicines into corresponding communities, and then identified abnormal doctors by comparing doctor and medicine communities to detect fraud in medical insurance [84].

5. Discussion and Future Directions

Multi-node, multi-relationship hybrid networks break the single limitation of traditional homogeneous networks, enabling the mining of rich hidden information through analysis. However, their characteristics pose numerous challenges for community detection algorithms: (1) Networks have multiple node and relationship types—how to fuse multi-layer networks and effectively utilize topological structure and node attribute information is the primary challenge [28]; (2) Large network scale—real-world networks have numerous nodes with sparse relationships, making it more difficult to design algorithms suitable for large-scale networks with good partitioning results; (3) Existence of unconnected similar node types or relationships—this is unfavorable for similarity measurement calculation; (4) Few algorithms identify overlapping communities, yet in real networks, a node may belong to multiple communities simultaneously, requiring effective algorithms for differentiation. These factors make heterogeneous network community detection algorithms highly challenging [26] and represent problems requiring solutions in future research.

Moreover, current applications of multi-node, multi-relationship hybrid network community detection methods in social media, academic networks, and fraud detection mostly target static networks without considering dataset changes. Future research should address how to construct community detection algorithms for dynamic hybrid networks to deeply reveal dynamic changes in scientific structures. Frontier research on multi-node, multi-relationship hybrid networks extends beyond community detection to include link prediction, node classification, and semantic search, which have practical significance for hybrid network research and represent future research directions.

References: [1] Newman M E J. Networks[M]. Oxford: Oxford University Press, 2018.
[2] Wasserman S, Galaskiewicz J. Advances in social network analysis: research in the social and behavioral sciences[M]. Los Angeles: Sage, 1994.
[3] Bader G D, Hogue C W. An automated method for finding molecular complexes in large protein interaction networks[J]. BMC bioinformatics, 2003, 4(2): 1-27.
[4] Sporns O. Networks of the brain[M]. Cambridge: MIT Press, 2010.
[5] Fortunato S. Community detection in graphs[J]. Physics reports, 2009, 486(3): 75-174.
[6] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical review e, 2004, 69(2): 026113.
[7] Duch J, Arenas A. Community detection in complex networks using extremal optimization[J]. Physical review e statistical nonlinear & soft matter physics, 2005, 72(2): 027104.
[8] Schlitt T, Brazma A. Current approaches to gene regulatory network modelling[J]. BMC bioinformatics, 2007, 8(6): 1-22.
[9] Zheng Yanyan, Wang Mingsheng, Shi Chuan, et al. Research on meta-path-based community discovery algorithms in heterogeneous information

networks[J]. Journal of Chinese Information Processing, 2018, 32(9): 132-142.
[10] Porter M A, Onnela J P, Mucha P J. Communities in networks[J]. Notices of the American Mathematical Society, 2009, 56(9): 4294-4303.
[11] Javed M A, Younis M S, Latif S, et al. Community detection in networks: a multidisciplinary review[J]. Journal of network and computer applications, 2018, 108(4): 87-111.
[12] Karatas A, Sahin S. Application areas of community detection: a review[C]//2018 International congress on big data, deep learning and fighting cyberterrorism. Ankara: IEEE, 2018: 65-70.
[13] Li Hui, Chen Fucai, Zhang Jianpeng, et al. Survey on community discovery algorithms in complex networks[J]. Application Research of Computers, 2021, 38(6): 1611-1618.
[14] Zhang Ruihong, Chen Yunwei, Deng Yong. Review of hybrid network community detection methods for scientific structure analysis[J]. Library and Information Service, 2019, 63(4): 135-141.
[15] Qian Xuesen, Yu Jingyuan, Dai Ruwei. A new scientific field—open complex giant systems and its methodology[J]. Chinese Journal of Nature, 1990, 13(1): 3-10.
[16] McCord M R. Urban transportation networks: equilibrium analysis with mathematical programming methods[J]. Transportation research part a: general, 1987, 21(6): 481-484.
[17] Sun Yizhou, Han Jiawei. Mining heterogeneous information networks: principles and methodologies[M]. Beijing: China Machine Press, 2016.
[18] Huo Chaoguang, Zhang Bin, Dong Ke. Review of academic behavior prediction research from a complex network perspective: topic selection, collaboration, and citation[J]. Information Studies: Theory & Application, 2021, 44(6): 180-188, 227.
[19] Ran W, Shi C, Yu P S, et al. Integrating clustering and ranking on hybrid heterogeneous information networks[C]//Pacific-Asia conference on knowledge discovery and data mining. Berlin: Springer, 2013: 583-594.
[20] Van Den Besselaar P, Heimeriks G. Mapping research topics using word-reference co-occurrences: a method and an exploratory case study[J]. Scientometrics, 2006, 68(3): 377-393.
[21] Han J W. Mining heterogeneous information networks: principles and methodologies[J]. ACM sigkdd explorations newsletter, 2012, 3(2): 1-159.
[22] Kang Yuhang. Technology opportunity analysis based on "coupling-co-citation" hybrid networks[J]. Journal of Intelligence, 2017(2): 170-179.
[23] Berlingerio M, Coscia M, Giannotti F. Finding and characterizing communities in multidimensional networks[C]//2011 international conference on advances in social networks analysis and mining. Taiwan: IEEE, 2011: 490-494.
[24] Suthers D, Fusco J, Schank P, et al. Discovery of community structures in a heterogeneous professional online network[J]. 2013 46th Hawaii international conference on system sciences, 2013(3): 3262-3271.
[25] Tang L, Liu H. Community detection and mining in social media[J]. Community detection and mining in social media, 2010, 2(1): 1-137.
[26] Zhang Zhenglin. Research on large-scale heterogeneous information

network community discovery algorithms and community characteristics[D]. Beijing: Beijing University of Posts and Telecommunications, 2017.
[27] Sun Y Z, Han J W, Zhao P X, et al. RankClus: integrating clustering with ranking for heterogeneous information network analysis[C]//ACM sigkdd international conference on knowledge discovery & data mining. New York: ACM Press, 2009: 565-576.
[28] Sun Y Z, Han J W. Ranking-based clustering of heterogeneous information networks with star network schema[C]//ACM sigkdd international conference on knowledge discovery & data mining. New York: ACM Press, 2009: 797-806.
[29] Ji M, Han J W, Danilevsky M. Ranking-based classification of heterogeneous information networks[C]//ACM sigkdd international conference on knowledge discovery & data mining. New York: ACM Press, 2011: 1298-1306.
[30] Zhao Huan. Research on Web service recommendation system based on heterogeneous network clustering[D]. Chongqing: Chongqing University, 2015.
[31] Gupta M, Aggarwal C C, Han J, et al. Evolutionary clustering and analysis of bibliographic networks[C]//International conference on advances in social networks analysis and mining. Taiwan: IEEE, 2011: 63-70.
[32] Qiu C H, Chen W, Wang T J, et al. Overlapping community detection in directed heterogeneous social network[J]. Web-age information management, 2015(6): 490-493.
[33] Chen Yi. Community structure analysis of complex networks based on statistical inference[D]. Harbin: Harbin Institute of Technology, 2016.
[34] Yin Haoxiao, Li Chuan. Research on heterogeneous information network probability models and community discovery algorithms[J]. Modern Computer (Professional Edition), 2016(3): 3-6.
[35] Sengupta S, Chen Y. Spectral clustering in heterogeneous networks[J]. Statistica sinica, 2015, 25(3): 1081-1106.
[36] Tong Hao, Yu Chunyan. Collaborative clustering algorithm for heterogeneous information networks based on ranking distribution[J]. Small Microcomputer Systems, 2014, 35(11): 2445-2449.
[37] Wang R, Shi C, Yu P S, et al. Integrating clustering and ranking on hybrid heterogeneous information networks[C]//Pacific-Asia conference on knowledge discovery and data mining. Berlin: Springer, 2013: 583-594.
[38] Xue Weijia. Research on clustering-based community discovery methods in heterogeneous information networks[D]. Baotou: Inner Mongolia University of Science and Technology, 2020.
[39] Sun Y Z, Han J W, Yan X F, 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.
[40] Shi C, Li Y T, Zhang J W, et al. A survey of heterogeneous information network analysis[J]. IEEE transactions on knowledge and data engineering, 2017, 29(1): 17-37.
[41] Shi C, Yu P S. Heterogeneous information network analysis and applications[M]. Switzerland: Springer International Publishing, 2017.
[42] Li J, Sun P Y, Mao Q R, et al. Path-Graph fusion based community detection over heterogeneous information networks[C]//2018 IEEE 20th inter-

national conference on high performance computing and communications. New Jersey: IEEE, 2018: 274-281.
[43] Shi C, Kong X N, Yu P S, et al. Relevance search in heterogeneous networks[C]//Proceedings of the 15th international conference on extending database technology. New York: ACM Press, 2012: 180-191.
[44] Ding Pingjian. Research on meta-path-based heterogeneous information network mining methods[D]. Changsha: Hunan University, 2015.
[45] Meng X F, Shi C, Li Y T, et al. Relevance measure in large-scale heterogeneous networks[C]//Asia-Pacific Web conference. Berlin: Springer, 2014: 636-643.
[46] Sun Y Z, Norick B, Han J W, et al. Pathselclus: integrating meta-path selection with user-guided object clustering in heterogeneous information networks[J]. ACM transactions on knowledge discovery from data, 2013, 7(3): 1-23.
[47] Wu Yao, Shen Derong, Kou Yue, et al. Multi-graph fusion for heterogeneous information network embedding[J]. Journal of Computer Research and Development, 2020, 57(9): 1928-1938.
[48] Hmimida M, Kanawati R. Community detection in multiplex networks: a seed-centric approach[J]. Networks & heterogeneous media, 2015, 10(1): 71-85.
[49] Kanawati R. LICOD: leader identification for community detection in complex networks[C]//2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. New Jersey: IEEE, 2011: 577-582.
[50] Papadopoulos S, Kompatsiaris Y, Vakali A. A graph-based clustering scheme for identifying related tags in folksonomies[C]//Proceedings of the 12th international conference on data warehousing and knowledge discovery. Berlin: Springer, 2010: 65-76.
[51] Shah D, Zaman T. Community detection in networks: the leader-follower algorithm[C]//Workshop on networks across disciplines in theory and applications. 2010, 1050(2): 1-8.
[52] Yakoubi Z, Kanawati R. LICOD: a leader-driven algorithm for community detection in complex networks[J]. Vietnam journal of computer science, 2014, 1(4): 241-256.
[53] Tang L, Wang X F, Liu H. Uncovering groups via heterogeneous interaction analysis[C]//Ninth IEEE international conference on data mining. New Jersey: IEEE Computer Society, 2009: 503-512.
[54] Nicosia V, Mangioni G, Carchiolo V, et al. Extending the definition of modularity to directed graphs with overlapping communities[J]. Journal of statistical mechanics: theory and experiment, 2009(3): 3166-3168.
[55] Guimerà R, Marta S P, Amaral L A. Module identification in bipartite and directed networks[J]. Physical review e statistical nonlinear & soft matter physics, 2007, 76(2): 036102.
[56] Murata T, Ikeya T. A new modularity for detecting one-to-many correspondence of communities in bipartite networks[J]. Advances in complex systems, 2010, 13(1): 19-31.
[57] Liu X, Liu W C, Murata T, et al. A framework for community detection

in heterogeneous multi-relational networks[J]. Advances in complex systems, 2014, 17(6): 1450018.
[58] Lancichinetti A, Fortunato S. Limits of modularity maximization in community detection[J]. Physical review e statistical nonlinear & soft matter physics, 2011, 84(6): 066122.
[59] Fortunato S, Barthélemy M. Resolution limit in community detection[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(1): 36-41.
[60] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788-791.
[61] Jolliffe I T. Principal component analysis[M]. Berlin: Springer-Verlag, 2002.
[62] Schölkopf T B, Müller T K-R. Fisher discriminant analysis with kernels[J]. Neural networks for signal processing ix, 1999, 1(1): 41-48.
[63] Liu T L, Gong M M, Tao D C. Large-cone nonnegative matrix factorization[J]. IEEE transactions on neural networks and learning systems, 2016, 28(9): 2129-2142.
[64] Tafavogh S. Community detection on heterogeneous networks by multiple semantic-path clustering[C]//IEEE international conference on computational aspects of social networks. New Jersey: IEEE, 2014: 7-12.
[65] Zhang X C, Li H X, Liang W X, et al. Multi-type co-clustering of general heterogeneous information networks via nonnegative matrix tri-factorization[C]//2016 IEEE 16th international conference on data mining. New Jersey: IEEE, 2016: 1353-1358.
[66] Huang Ruiyang, Wu Qi, Zhu Yuhang. Dynamic heterogeneous network community discovery method based on joint matrix factorization[J]. Application Research of Computers, 2017, 34(10): 2989-2992.
[67] Liu J, Wang J Z, Liu B H. Community detection of multi-layer attributed networks via penalized alternating factorization[J]. Mathematics, 2020, 8(2): 239-258.
[68] Liu Peiqi, Sun Jiehan. Label propagation algorithm based on LDA topic model[J]. Computer Applications, 2012, 2(2): 403-406.
[69] Mei Q Z, Cai D, Zhang D, et al. Topic modelling with network regularization[C]//Proceeding of the 17th international conference on World Wide Web. New York: ACM Press, 2008: 101-110.
[70] Wang Ting. Research on community discovery algorithms in heterogeneous social networks[D]. Beijing: China University of Mining and Technology, 2016.
[71] Bao Lihong. Comparison of dimensionality reduction between principal component analysis and linear discriminant analysis[J]. Statistics and Application, 2020, 9(1): 47-52.
[72] Lin L, Xia Z M, Li S H, et al. Detecting overlapping community structure via an improved spread algorithm based on PCA[C]//International conference on computer science and software engineering. Lancaster: DEStech, 2014: 115-121.
[73] Li L, Fan K F, Zhang Z Y, et al. Community detection algorithm based on local expansion K-means[J]. Neural network world, 2016, 26(6): 589-605.

[74] Yuan P Y, Wang W, Song M Y. Detecting overlapping community structure with PCA technology and member index[C]//Proceedings of the 9th EAI international conference on mobile multimedia communications. New York: ACM Press, 2016: 121-125.
[75] Liu W, Chen L. Community detection in disease-gene network based on principal component analysis[J]. Tsinghua science and technology, 2013, 18(5): 454-461.
[76] Chen Changgeng. Node importance measurement and community discovery based on meta-paths in heterogeneous information networks[D]. Kunming: Yunnan University, 2019.
[77] Gao Changjie, Peng Dunlu. Academic community discovery algorithm for DBWorld data mining[J]. Application Research of Computers, 2017(7): 2059-2062.
[78] Santos J M, Embrechts M. On the use of the adjusted rand index as a metric for evaluating supervised classification[M]. Berlin: Springer, 2009.
[79] Wang Yiwen. Complex network node influence model and its application[D]. Hangzhou: Zhejiang University, 2015.
[80] Karatas A, Sahin S. Application areas of community detection: a review[J]. 2018 international congress on big data, deep learning and fighting cyberterrorism (ibigdelft), 2019(1): 65-70.
[81] Zhang Haitao, Zhou Honglei, Zhang Xinrui, et al. Research progress on community discovery in online social networks[J]. Library and Information Service, 2020, 64(9): 142-152.
[82] Qiao Y Q, Niu K, Du S, et al. Community detection analysis of heterogeneous network[J]. 2015 international conference on cyber-enabled distributed computing and knowledge discovery, 2015(10): 509-512.
[83] Huang W H, Liu Y, Chen Y G. Mixed membership stochastic blockmodels for heterogeneous networks[J]. Bayesian analysis, 2019, 15(3): 711-736.
[84] Luan Tingting. Research on medical abuse detection based on heterogeneous network community partition[D]. Jinan: Shandong University, 2019.
[85] Liu Dianzhong. Fraud detection for dynamic financial complex data[D]. Qingdao: Qingdao University, 2020.

Author Contributions:

Jiang Lu: Collected and organized materials, wrote and revised the paper.
Chen Yunwei: Proposed the review writing ideas and revision suggestions.

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