Postprint: Advances in Community Detection for Online Social Networks

Authors: Haitao Zhang, Zhou Honglei, Zhang Xinrui, Sun Tong

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

Abstract

[Purpose/Significance] Taking online social networks as the research object, this study accurately captures the development trends and research hotspots of community detection through literature review, and explores approaches to mining hidden communities in large-scale social networks, which holds theoretical and practical significance. [Method/Process] Using the China National Knowledge Infrastructure (CNKI) database, Web of Science Core Collection, and related international conference literature as data sources, we apply the CiteSpace visualization analysis tool to conduct quantitative research from the perspectives of hotspot keywords, topic evolution paths, and co-cited literature, and provide detailed review and commentary on the literature content across three dimensions: community detection methods, algorithm implementation, and application practice. [Results/Conclusions] The current research field still possesses broad development prospects; future research should emphasize algorithm optimization and innovation, differentiation and expansion of application scenarios, and cross-disciplinary research integrating interdisciplinary knowledge and cutting-edge technical methods.

Full Text

Research Progress in Community Detection of Online Social Networks

Zhang Haitao[1,2], Zhou Honglei[1], Zhang Xinrui[1], Sun Tong[1]

[1]School of Management, Jilin University, Changchun 130022
[2]Information Resource Research Center, Jilin University, Changchun 130022

Abstract: *[Purpose/Significance]* This study examines online social networks as its research object. Through systematic literature review, it aims to accurately capture the development trends and research hotspots of community detection, and explore how to mine hidden communities in large-scale social networks, which holds both theoretical and practical significance. *[Method/Process]* Using the CNKI database, Web of Science Core Collection, and relevant international conference literature as data sources, we applied the CiteSpace visualization tool to conduct quantitative research from the perspectives of hotspot keywords, topic evolution paths, and co-cited literature. We also provided detailed qualitative reviews of the literature from three dimensions: community detection methods, algorithm implementation, and application practice. *[Result/Conclusion]* The current research field still has broad development space. Future research should focus on algorithm optimization and innovation, differentiation and expansion of application scenarios, and interdisciplinary research that integrates knowledge and cutting-edge technological methods from multiple fields.

Keywords: online social network; community detection; dynamic community evolution; research progress

Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2020.09.016

---

1. Introduction

In recent years, online social platforms have exerted significant influence on individuals' learning and daily life, national economic development, and social stability. Social media platforms such as Zhihu and Weibo have broadened channels for netizens to acquire knowledge and stay informed about news. However, as online social network platforms have become fully integrated into people' s lives, the proliferation of false rumors and fraudulent activities by internet promoters has disrupted normal internet order. Fang Binxing et al. [1] define online social networks as social structures formed by group collections and individual connections on information networks, comprising three elements: network groups, relational structures, and network information, characterized by the interpenetration of interpersonal communication and virtual interaction.

Research has revealed that community structures exist in online social networks similar to those in real society—that is, the entire social network consists of several sub-communities [2]. Simply put, community detection means discovering sub-communities with certain associations within the overall network structure. However, as research has progressed, the criteria for community division in online social networks have expanded. Community structures can now be characterized through interest networks, tag information, and node links, making the concept still evolving without a clear definition to date.

Based on the above logic, this paper conducts quantitative analysis of relevant literature and categorizes it into three dimensions according to research hotspots: community detection methods, algorithms, and application practice. It analyzes and extracts the development trends and research frontiers in this field both domestically and internationally, aiming to clarify future research entry points. The specific research framework is shown in Figure 1 [Figure 1: see original paper].

2. Bibliometric Analysis of Community Detection Literature in Online Social Networks

Discovering these potential communities has practical significance for studying information dissemination hierarchies, friend recommendations, and network public opinion monitoring. In academic research, the structural evolution of knowledge networks also reflects the development process of disciplinary research themes. Community detection methods have been widely applied to topics such as author collaboration networks and disciplinary knowledge flow [3-4], providing new perspectives for studying scholar collaboration networks and academic topic evolution. Community detection is also known as community identification, community recognition, or social group discovery. M. E. J. Newman [5] defines community detection as dividing the entire network structure into several groups based on network nodes, making connections within groups relatively dense and connections between groups relatively sparse, where these groups represent deeply mined hidden communities or sub-communities.

We used the visualization software CiteSpace to conduct bibliometric analysis of the literature, covering four aspects: temporal distribution of publications, keyword statistics, topic development paths, and co-citation analysis.

2.1 Data Sources Our data primarily came from three sources: authoritative Chinese and English databases, international conference proceedings, and online academic communities plus self-media platforms.

(1) English literature sources. Based on the Web of Science Core Collection with a time span of 1990-2019, the search query was: TS=("Online social network"AND ("Community discovery"OR "community identif"*OR " Community detect*")). This yielded 885 documents, which were filtered by theme and abstract content to obtain 460 highly relevant papers. An additional 17 papers were obtained through similar searches on Elsevier, and 52 recent top-tier journal and conference papers were referenced during the writing process.

(2) Chinese literature sources. Primarily from three databases: CNKI, Wanfang, and VIP Chinese Science and Technology Journals. The search query was: SU=('social network'+ 'social media platform') * ('community discovery'+ 'community detection' + 'community identification' + 'social group discovery'), retrieving 439 Chinese documents. After deduplication and screening, 402 highly relevant documents were obtained.

(3) Online academic communities and self-media platforms. We examined practical effects and efficiency evaluations of commonly used algorithms such as label propagation and Louvain on current scientific academic communities and self-media platforms such as Jianshu, Toutiao, and CSDN.

2.2 Temporal Distribution of Publications Foreign scholars first explored community structures in social networks in 1999, proposing numerous pioneering ideas and foundational theories. As shown in Figure 2 [Figure 2: see original paper], the overall publication volume growth trends are synchronized between domestic and international research. Domestic research on this topic began in 2004, mainly consisting of reviews of foreign literature, and adaptation and optimization of authoritative foreign algorithms. In 2008, with the rapid development of online social platforms, entertainment blogs, and knowledge-sharing learning communities, the number of publications showed significant growth for the first time. Since 2010, the field has experienced a research boom, with a large number of algorithm and methodology studies emerging, and practical application scenarios enriched, mainly involving telecommunications, e-commerce, and network security.

2.3 Keyword Statistics The domestic keyword distribution is shown in Figure 3 [Figure 3: see original paper]. Besides "community detection" and "social network," the most important node is "complex network," appearing 51 times. Other high-frequency nodes include Weibo (40), overlapping community (29), label propagation (27), community division (20), community structure (18), collaborative filtering (18), recommendation system (16), topic model (11), modularity (11), matrix factorization (10), and friend recommendation (10), providing references for the following thematic content analysis.

The foreign keyword distribution is shown in Figure 4 [Figure 4: see original paper]. Similar to domestic research, besides "community detection" and "social network," important high-frequency nodes include complex network (105), social network analysis (78), modularity (57), algorithm (50), overlapping community (43), community structure (27), graph (27), cluster (21), label algorithm (17), centrality (16), and data mining (12).

2.4 Topic Development Path Analysis This section uses timezone view to illustrate the evolution path of research topics. Domestic "community detection" research began in 2004, when Yang Nan et al. [6] reviewed Web community detection technologies. The theme of "complex network" has continued from 2004 to the present, with its small-world properties, scale-free properties, and network clustering characteristics being thoroughly studied. From 2005-2008, research themes were relatively sparse. Starting in 2010, the field gained widespread attention, with community detection methods widely applied in information retrieval, recommendation systems, and user analysis. Social media platforms, particularly Weibo, became important platforms for online social network research, with user tendency analysis, interest expansion, and topic detection receiving attention. Network community evolution, community structure identification, and "dynamic community" and "overlapping community" detection algorithms were all hot topics, as shown in Figure 5 [Figure 5: see original paper].

Foreign scholars began studying random networks in 1999. Adjacent matrix analysis and centrality derived from this theme were crucial for implementing community detection methods (see Figure 6 [Figure 6: see original paper]). Other topic clusters also basically began in 2004. Analysis shows that foreign research is more practice-oriented, with numerous algorithms and technologies integrated into specific real-world domains such as social media, medical viruses, telecommunications, and e-commerce, tending to combine virtual network community detection with real-world group analysis to address big data challenges and provide better user experiences. Simultaneously, foreign research places greater emphasis on theoretical foundations, such as game theory, graph theory, decision theory, and fuzzy adaptive resonance theory. These theories are closely related to community detection, topology, and label propagation algorithms in complex networks, especially game theory-based community detection methods, which have received widespread attention in overlapping community identification [7]. Similarly, foreign research also emphasizes network community structure identification, overlapping community detection, and dynamic community evolution. In terms of algorithms, domestic and foreign research are broadly similar. The Bat Algorithm (BA algorithm) shown in Figure 6 was proposed by X. S. Yang [8] in 2010 as an effective method for searching global optimal solutions.

2.5 Analysis of Key Literature Co-cited literature nodes play a connecting role in thematic development and evolution. Using CiteSpace' s burst detection function and analyzing highly cited Chinese literature from the past decade year by year, we identified representative scholars and key literature as shown in Figure 7 [Figure 7: see original paper] and Table 1 .

Table 1. Summary of Core Literature

Time Period	Representative Scholars	Key Literature	Main Contributions
2005-2015	F. Radicchi et al. [9-10]	Community structure in social and biological networks	Earliest studies on community structure in social networks, proposed the important concept of "modularity"

Time Period	Representative Scholars	Key Literature	Main Contributions
2004-2010	M. Girvan, M. E. J. Newman et al. [12-14]	Finding and evaluating community structure in networks	Divided network communities into strong and weak communities based on structural tightness
2010-2019	S. Gregory, Wu Xiaolan, Zhang Chengzhi, Liu Shichao, Xin Yu et al. [19-22]	Reviews: Comparative analysis and improvement suggestions for community detection methods	Early proposal of community detection algorithms that attracted scholarly attention
2007-2011	U. N. Raghavan, R. Albert, M. Rosvall et al. [15-16]	Near linear time algorithm to detect community structures	First defined overlapping communities, becoming a new research focus
2013-2019	F. Radicchi, L. Danon, A. Lanchichinetti et al. [24-26]	Finding overlapping communities in networks	Continuous proposal and improvement of algorithm benchmarks and evaluation standards
	L. Danon, Wang Li, Li Jianhua et al.	Community detection in dynamic social networks	Explored dynamic community detection algorithms to track community structure evolution
	J. Xie, Z. Zhao, He Jing et al.	Incremental method to detect communities	New research perspective: communities are formed by a set of closely related "links" rather than nodes

Time Period	Representative Scholars	Key Literature	Main Contributions
	G. Palla, I. Derényi et al. Y. Y. Ahn et al.	Link communities reveal multiscale complexity	Developed new community detection schemes for network evolution, including combination and innovation of multiple theories or models

Analysis of the above key literature reveals that the academic community places greater emphasis on overlapping community detection, dynamic community identification, and comparative analysis of community detection methods.

3. Content Analysis of Community Detection Themes in Online Social Networks

Through the above quantitative analysis and literature review, we found that research on community detection methods, algorithm implementation, and application practice is very rich and can basically cover the relevant research themes, though no scholar has yet provided a comprehensive review. Therefore, this paper divides the "community detection" theme in online social networks into three dimensions for qualitative review: community detection methods, algorithm implementation, and application practice research.

3.1 Community Detection Methods Experts and scholars have explored community detection methods from different perspectives, keeping pace with the development and changes of online social networks. Based on the prominent community characteristics of online social networks in recent years, community detection algorithms and evaluation standards have received further attention and optimization. Since the following sections focus on reviewing community detection methods and algorithms, this section only briefly summarizes the core literature content, as shown in Table 2 .

Table 2. Summary of Community Detection Methods

Method Category	Representative Literature	Core Idea	Advantages/Disadvantages
Graph theory-based community detection	Y. R. Lin et al.	Community discovery through decomposing incremental meta-graphs to handle time-varying relationships	Well-developed theory, simple and feasible method, but has limitations in identifying dynamic networks
	J. Chen et al.	Extracting meaningful dense subgraphs from given sparse graphs	No need to pre-specify the number of clusters
Mathematical model-based community detection	P. K. Gopalan	Bayesian network model combined with network subsampling for dynamic updating of community estimates	Precise computation, high-quality identified community structures, but suffers from parameter oversensitivity during expansion

Method Category	Representative Literature	Core Idea	Advantages/Disadvantages
	Zhang Qin et al.	Using grey theory, density peak clustering, and rough set theory for community detection	
Semantic-based community detection	Z. Xia et al.	Mining semantic information from comment content to build similar topic networks	Can utilize key semantic information to reflect people' s preferences and topics of interest, with strong applicability
	M. M. Anwar	Discovering time-sensitive activity-driven user groups based on same-topic input queries	

Method Category	Representative Literature	Core Idea	Advantages/Disadvantages
Link analysis-based community detection	Y. Y. Ahn et al.	Considering communities as formed by "links" rather than nodes, using link similarity for detection	Can effectively detect social groups using potential links without considering network topology
	W. Liu et al.	Markov network framework for community detection through link analysis	
Local optimization and expansion-based community detection	K. Tang et al.	Local influence-first strategy, assigning candidate nodes to communities based on marginal gain	No need to construct complete network topology, reducing high dimensionality or sparsity of adjacency matrices

Method Category	Representative Literature	Core Idea	Advantages/Disadvantages
Deep learning-based community detection	L. Yang et al.	Semi-supervised method based on modularity function for mining community structures	Strong high-dimensional data processing and data mining capabilities, easy to solve complex computational problems in large networks
	G. Sperli	Using deep learning algorithms to improve community detection	

3.1.1 Graph Theory-Based Community Detection Methods (2002-present) Graphs are commonly used to represent relationships in social networks, including elements such as nodes, edges, and degrees. Classical graph clustering methods, also known as graph partitioning, attempt to find the best cutting method to divide graphs into different parts (i.e., communities), often using minimum graph segmentation methods. Y. R. Lin et al. [27] studied community structures in media platforms, using meta-graphs to build multi-relational models and multi-dimensional social data, processing time-varying relationships through incremental meta-graph decomposition. J. Chen et al. [28] extracted meaningful dense subgraphs from given sparse graphs, interpreting "dense subgraphs" as communities, a method that does not require pre-specifying the number of clusters.

3.1.2 Mathematical Model-Based Community Detection Methods (2007-present) This research theme involves numerous theories and methods from mathematics. P. K. Gopalan [29] proposed a community detection method based on Bayesian models, allowing nodes to participate in multiple communities, which aligns with the characteristics of overlapping community identification while flexibly superimposing subsampling from networks and dynamically updating community estimates. Zhang Qin et al. [30] used grey theory to define global structural similarity, combined with density peak clustering algorithms to determine cluster centers, and introduced rough set theory to automatically select central nodes based on network structure, continuously adjusting distance ratio thresholds for iterative partitioning to identify overlapping community structures.

3.1.3 Semantic-Based Community Detection Methods (2009-present) Most community detection methods use common node attributes or topological structures for community division, failing to utilize key information such as semantics of nodes or edges, and cannot reflect people' s interests, preferences, or topics of concern. Semantic-based community detection methods consider node information content and are suitable for studying social media platforms. Z. Xia et al. [31] mined semantic information from comment content to build an entire semantic topic network, focusing on the impact of topic weights on each edge and emphasizing reduction of computational complexity, making it suitable for large-scale network processing. M. M. Anwar et al. [32] discovered time-sensitive activity-driven user groups for a given set of query topics, tracking topic attention within groups that tend to be similar over time.

3.1.4 Link Analysis-Based Community Detection Methods (2010-present) In social media platforms, user attention often exists in one-way situations with weak connection degrees, making community detection using topological structure characteristics sometimes unsatisfactory. This method allows network vertices to belong to multiple communities, helping to discover overlapping community structures. Different from most researchers' views, Y. Y. Ahn et al. [18] pointed out that highly overlapping communities may have more external links, arguing that communities are formed by a set of closely related "links" rather than nodes. This method uses hierarchical clustering and similarity between links to build dendrograms, analyzing discovered link communities at maximum partition density. W. Liu et al. [33] used a Markov network framework for community detection by analyzing links associated with social objects, using potential links to mine hidden dynamics in social group behavior, a method that does not require considering the topology of social networks.

3.1.5 Local Optimization and Expansion-Based Community Detection Methods (2012-present) When identifying virtual communities in large-scale social networks with numerous node information, local community detection algorithms do not require overall network information. These methods quickly locate target nodes through local structure information, starting from core nodes and incorporating surrounding nodes into identified virtual communities through local benefit functions and greedy strategies, mainly including four types of methods: local expansion optimization, clique filtering, label propagation, and local edge clustering optimization [14]. J. Tang et al. [34] adopted a local influence-first strategy when maximizing community propagation influence, using label propagation algorithms to detect community distribution in the first stage and assigning candidate node quantities to each community based on their marginal gain, on which basis they formulated dynamic encoding mechanisms for individual particles and discrete evolution rules for groups to identify high-influence nodes within communities.

3.1.6 Deep Learning-Based Community Detection Methods (2014-present) In the era of big data, deep learning algorithm research is burgeoning, widely applied in computer vision, machine reading comprehension, and large-scale data processing. A few scholars have proposed solutions for community detection methods, mostly based on dimensionality reduction processing of network node information or obtaining low-dimensional feature matrices by training network graph similarity matrices. L. Yang et al. [35] proposed a semi-supervised community detection method based on modularity functions, using unweighted graph matrices as autoencoder inputs to obtain low-dimensional embedding matrices with nonlinear mappings. G. Sperli [36] proposed a novel community detection method based on deep learning to address the high dimensionality or sparsity issues of adjacency matrices, fully considering dataset dimensions and topological characteristics of adjacency matrices.

3.2 Community Detection Algorithms Algorithm research is crucial for mining community structures. According to the above quantitative analysis, community detection algorithm research began in 2002. Scholars such as M. Girvan, M. E. J. Newman, and S. Fortunato were early pioneers in this field. After 2010, research on algorithm optimization and innovation gradually increased, with experts proposing or optimizing early community detection algorithms to cope with rapidly changing community structures. Meanwhile, algorithm benchmarking has been continuously improved, and exploration of overlapping and dynamic communities remains a key research focus. Therefore, combining the above key literature review with tracking of the latest algorithm research, we summarize as follows:

3.2.1 Early Authoritative Community Detection Algorithms Early community detection algorithms can be broadly divided into divisive algorithms, modularity optimization-based algorithms, and algorithms based on information theory concepts.

(1) The G-N algorithm is the most representative divisive algorithm. M. Girvan and M. E. J. Newman [9] focused on community structures in social and biological networks in 2002, first pointing out that network topological features facilitate community structure identification, and proposed a divisive hierarchical algorithm based on maximum edge betweenness (G-N algorithm). This algorithm does not require pre-determining the number of communities and improves upon the limitations of the Kernighan-Lin (KL) algorithm and spectral bipartitioning to some extent, but it requires repeated evaluation of each edge, making it costly.

(2) The Louvain algorithm [37] is a modularity-based community detection algorithm that does not require prior community information. It demonstrates good efficiency and effectiveness in studying large networks and is commonly used in blog-based social media platforms and citation network communities [38]. The CNM algorithm [39], based on greedy principles, essentially applies modularity concepts, using three data structures for community detection, where the sparse matrix represents variables indicating changes in node modularity. In addition, algorithms based on simulated annealing and extremal optimization also originate from modularity principles.

(3) M. Rosvall et al. approached from an information theory perspective, proposing the Infomap algorithm [16] in 2008 based on random walk coding and information compression coding. This algorithm uses an objective optimization function to group nodes with high similarity into the same community, achieving community detection by partitioning virtual communities with the shortest coding length, mostly used in scenarios requiring detailed community structure characterization.

3.2.2 Development and Optimization of Community Detection Algorithms The above algorithms assume each node exists in only a single community. However, taking online social networks as an example, each user has the possibility of belonging to different communities. Therefore, although some algorithms are still used today, they have limitations. G. Palla et al. [17] first demonstrated the phenomenon of duplicate nodes in social networks and proposed the Cluster Percolation Method (CPM) clique filtering algorithm. As the structural characteristics of network communities received widespread attention, various more applicable community detection algorithms emerged.

(1) Label Propagation Algorithm (LPA) and its optimization. LPA [40] is a semi-supervised learning method based on graphs, using labeled node information to predict other nodes. It has advantages such as simple implementation and low time complexity, but LPA only applies to non-overlapping static network communities. S. Gregory [19] optimized the label propagation algorithm by expanding node label types and information propagation paths, enabling information values to encompass multiple communities for overlapping community detection. Wu Xiaolan et al. [20] proposed the COPRA_{CD} algorithm based on contribution degree, using node contribution to communities to distinguish the closeness between nodes and their neighbor communities. Liu Shichao [21] proposed the LPPB overlapping community detection algorithm based on label propagation probability, which first assigns an independent label to each node, then sorts nodes by influence, calculates label propagation probability based on network propagation characteristics and node attribute feature values, and finally uses nodes' historical label records to correct detection results.

(2) Community detection algorithms based on spectral analysis and clustering. Spectral analysis-based algorithms are based on the idea that nodes within the same community show similarity in eigenvectors of the Laplacian matrix. These algorithms treat matrix eigenvectors corresponding to nodes as spatial coordinates, map network nodes to multi-dimensional vector spaces, and then apply classical algorithms like K-means or FCM for clustering. Such algorithms are costly but highly flexible as they can directly use traditional vector clustering results [41]. T. Ma et al. [42] proposed the LED algorithm for overlapping community structure identification, which converts structural similarity between vertices into network weights based on structural clustering, improving algorithm precision and computational efficiency. Zhang Junxiang [43] proposed the L1-ECDA algorithm based on smooth L1-norm deep sparse autoencoders, which preprocesses network graph adjacency matrices for dimensionality reduction and uses three-layer neural networks and K-means algorithm for matrix clustering to discover hidden network communities.

(3) Community detection algorithms integrating multi-dimensional information. As user communication gradually shifts to online media and social relations increasingly appear on online social platforms, algorithm research targeting platform characteristics has become a hot topic in recent years. Algorithms improve accuracy by integrating multi-dimensional information, including user behavior, node content, link weights, social relations, and geographic attributes. Liu Bingyu et al. [44] proposed the DC-DTM algorithm, mapping Weibo networks to directed weighted networks where edge directions reflect attention relationships between nodes, assigning semantic similarity between nodes as connection weights to effectively solve the sparsity of Weibo networks and the backflow problem of traditional LPA algorithms. Tian Bo et al. [45] proposed a community detection algorithm based on user interaction behavior, using forwarding, commenting, and mentioning relationships between Weibo users to build weighted interaction networks, continuously merging node pairs that maximize modularity function gain until the gain becomes negative. Xin Yu et al. [22] proposed the LBTC algorithm for semantic community detection, using the LDA model as the semantic information model, combining semantic features with social relationship features, and achieving quantification of semantic information by defining semantic link weights.

(4) Dynamic community detection algorithms. The overall scale of social networks continues to expand while nodes and links disappear. Community evolution analysis mainly explores the phased changes in network structure, which better reflects the dynamic interaction behaviors of network users. However, scholars mostly abstract real networks as static networks in certain time windows, with algorithms only applicable to situations with small numbers of added nodes or edges. J. Xie et al. [24] proposed the LabelRank algorithm to optimize the randomness problem of early algorithms, introducing a set of operators to control and stabilize propagation dynamics, thereby identifying all communities being detected in the network. Compared with static community detection algorithms, LabelRank significantly improves the quality of detected communities.

Z. Zhao et al. [25] have been committed to studying edge weight update rules for years, proposing a "completely independent" subgraph update strategy, dividing dynamic communities into two methods: birth of new communities and expansion of original communities, to timely mine community evolution processes. He Jing et al. [26] proposed the LUA algorithm considering incremental nodes, updating topological potential values based on topological potential field theory, calculating neighbor nodes within the influence range of incremental nodes, and dynamically adjusting community 归属 of changed network parts.

3.2.3 Evaluation Metrics for Community Detection Algorithms (1) In 2004, M. E. J. Newman and M. Girvan [10] proposed the famous "modularity" as a standard for measuring community division, using the G-N algorithm for experiments. The proposal of modularity brought new opportunities for algorithm research and evaluation, and later scholars also proposed corresponding modularity evaluation formulas for overlapping community detection algorithms.

(2) F. Radicchi [11] first quantitatively divided network communities into "strong communities and weak communities" based on subgraphs, enabling judgment of community structure tightness. On this basis, they proposed a splitting algorithm considering only local variables to improve upon the resolution limitations and excessive dependence on global network features in modularity optimization.

(3) In terms of algorithm performance, L. Danon et al. [12] compared community detection algorithms at that time based on sensitivity and computational cost, suggesting that both aspects should be comprehensively measured when selecting algorithms. In addition, precision and time complexity are also commonly used indicators for measuring algorithm performance.

(4) Normalized Mutual Information (NMI) is also a widely used community division evaluation indicator. This method defines a confusion matrix where rows correspond to real communities and columns to discovered communities, judging the actual effectiveness of algorithms by measuring the similarity between discovered and real communities [12].

(5) As link direction and algorithm weights receive widespread attention, A. Lanchichinetti et al. [23] proposed in 2011 a class of undirected and unweighted benchmark graphs that explain the heterogeneity of node degree and community size distributions, depicting heterogeneous distribution sizes of node degrees and communities, and fully considering the possibility of nodes belonging to multiple communities, making it suitable for testing overlapping community algorithms.

In summary, community detection algorithms continue to be updated, and testing standards for algorithms are continuously increasing, but no unified standard has been achieved to date. Especially in dynamic communities where various situations such as merging, splitting, shrinking, expanding, emerging, and disappearing exist [13], evaluation based solely on community similarity at adjacent time points is insufficient. Secondly, most algorithms are verified through synthetic networks or self-selected test networks, making evaluation results subjective.

3.3 Application Practice of Community Detection

3.3.1 Applications in Social Media Platforms Identifying information dissemination trends, discovering community structure evolution patterns, and detecting evolution anomalies are valuable for monitoring network group events and dynamic public opinion evolution. Particularly, the approach to dynamic community identification based on time node algorithm design aligns with contextual research directions in social media platforms. K. Gu et al. [46] used edge tightness and node interest information as reference standards for community division to study personal willingness and information dissemination schemes in Weibo, YouTube, and Digg communities. C. Li et al. [47] proposed a Weibo community detection algorithm based on tag correlation and interactive behavior, using an improved maximum marginal relevance model to accurately assign user tags for mining Weibo user group characteristics. Li Gang et al. [48] borrowed community detection ideas to partition co-word networks, forming "topic communities" describing different hot topics, providing new approaches for identifying, measuring, and analyzing the evolution of hot topics on Weibo. Research shows that social platforms like Weibo, Facebook, and Twitter can provide personal information (node attributes), including friend records, interest preferences, and location information. Conducting community detection research on these platforms helps timely grasp group trends and improve user experience.

3.3.2 Applications in Recommendation Systems Community detection methods can identify similar users in large-scale user groups for precise push based on common user characteristics. In particular, local community detection and network community link prediction can further optimize traditional collaborative filtering, which suffers from data overload across the entire user network and low recommendation efficiency. K. Xin Chang et al. [49] alleviated the cold-start problem in recommendation systems by constructing user relationship networks, using personal information to establish user relationship matrices, and employing edge centrality-based community detection methods to provide precise push services for new users. Zhang Jidong et al. [50] improved the efficiency and accuracy of friend recommendations in mobile social networks by constructing an information service recommendation model based on community division and user similarity. Community detection methods significantly improve recommendation efficiency while ensuring recommendation accuracy, providing new methods for precise recommendation services in online social networks.

3.3.3 Applications in Internet Marketing Users establish social connections through network media, and information dissemination activities form a natural interactive behavior. With the continuous expansion of network user scale, internet marketing has become mainstream. However, conducting promotional marketing in huge social networks alone is inefficient and costly. Therefore, community detection methods can provide solutions: by mining network community structure characteristics, effectively selecting marketing activity scope, avoiding information dissemination overlap, and improving network service efficiency. Y. C. Chen et al. [51] developed a high-influence group identification framework based on influence maximization problems, allocating seed quantities according to community size, and maximizing information dissemination scope through three steps: identifying community structure, selecting candidates, and determining seed nodes. The influence maximization problem often refers to determining the minimum set of nodes for influence propagation, but requires high algorithm efficiency and practicality.

4. Research Challenges and Future Prospects

Research findings show that although there are many algorithm papers, most focus on optimization and adaptation of early authoritative algorithms. Early algorithms failed to accurately explain significant interference situations during community identification. Specifically, when analyzing dynamic community evolution, can source historical information be accurately preserved? How can we effectively reduce the phenomenon of large communities swallowing weak communities rather than simply judging them as overlapping communities? Meanwhile, issues such as the high cost of the G-N algorithm, instability of the LPA algorithm, and high cost of other algorithms, as well as the trade-offs between their advantages like identification accuracy, overlapping community identification, or dynamic community evolution tracking, urgently need unified standards for balancing. Secondly, the current application scope of this field is narrow, and practical application scenarios are not rich enough. Domestic online social network research platforms are currently limited to content-sharing platforms like Weibo, with application domains mostly distributed in virtual communities. Therefore, combined with current research difficulties, we propose the following future prospects:

4.1 Optimization and Innovation of Community Detection Algorithms Existing algorithms are mostly highly complex, making them unsuitable for mining network communities in large-scale social networks and causing a certain degree of information loss. Secondly, network community structures update rapidly, making algorithms based on large-scale static data measurement no longer applicable. Finally, as users gradually migrate from content-creative social networks to message-stream social networks, location-based detection algorithms fail to reflect the temporal characteristics of social media. Therefore, future algorithm research can focus on:

(1) Research on fast algorithms and fuzzy recognition algorithms. For example, community detection algorithms based on local network nodes, key tag information, and multi-dimensional node information fusion should receive further attention in the future.

(2) Optimize the adaptability of early authoritative algorithms and construct evaluation mechanisms for community detection algorithms. Unify evaluation standards for community detection algorithms, assign corresponding weights to different indicators, especially improve the evaluation of dynamic community detection algorithms, and comprehensively evaluate community evolution and identification quality across different time windows.

(3) Propose community detection algorithms based on spatiotemporal features. Combine time and space information to optimize location-based detection algorithms, and study the spatiotemporal trajectory distribution of topics or user groups in social media.

4.2 Differentiation and Expansion of Community Detection Method Application Scenarios **(1) Differentiate application scenarios to select optimal methods.** For example, for small community analysis, algorithms with high accuracy but low efficiency can be selected; conversely, when studying networks with tens of millions of nodes, besides identification accuracy, budget and time costs should also be considered. The selection of community detection methods should not blindly pursue identification precision.

(2) Expand research platform types. There is currently little research on shopping platforms like Taobao and Amazon, online health communities, and Q&A communities like Zhihu. Community detection research can be conducted based on various platforms.

(3) Broaden practical application scenarios. Many foreign literature studies involve virus transmission networks, communication networks, and identification of cybercrime groups. Future research can expand practical application scenarios of community detection, such as user privacy protection in community structure analysis, identifying anomalous user groups to solve network security problems, and identifying emerging themes in virtual academic communities.

4.3 Interdisciplinary Research Integrating Cross-Disciplinary Knowledge and Cutting-Edge Technical Methods **(1) Integrate interdisciplinary domain knowledge.** For example, drawing on network structure analysis methods used in studying protein network structures, predicting protein lesions, and inter-molecular influences [52], future research can integrate knowledge from medicine and biology to analyze biological cells and molecular structures using community detection methods. Or based on complex network and propagation dynamics theories, using community detection methods to study information and knowledge dissemination mechanisms, knowledge fusion patterns between disciplines, and track network public opinion propagation trends.

(2) Combine emerging methods like artificial intelligence. Complex network node information representation mostly uses manual feature extraction methods, while network representation learning algorithms transform network information into low-dimensional dense real vectors, achieving optimal parameter solving through gradient descent optimization algorithms, which can reduce computation and labor costs [53]. For example, the SOM neural network clustering algorithm [54] can assign corresponding weight values to different tags. Combined with community detection methods, it can accurately cluster users and effectively solve the sparsity of interest features or dimensionality disasters caused by too many tags during feature selection. Secondly, as people become increasingly interested in video media platforms, extracting semantic information from video media and focusing on user-generated content have become difficult points, resulting in limited research on community detection methods in image-text and video-sharing media. Referencing the successful applications of word2vec models and convolutional neural networks in image recognition and natural language processing, future research can apply these cutting-edge technical methods to improve community detection methods, helping to mine group structures in content-based multimedia communities.

References

[1] Fang Binxing, Jia Yan, Han Yi. Core scientific problems, research status and future prospects of social network analysis[J]. Bulletin of Chinese Academy of Sciences, 2015, 30(2): 187-199.
[2] Cheng Xueqi, Shen Huawei. Community structure in complex networks[J]. Complex Systems and Complexity Science, 2011, 8(1): 57-70.
[3] Wu Xiaolan, Zhang Chengzhi. Research on disciplinary knowledge flow patterns from the perspective of academic social media—A case study of ScienceNet[J]. Data Analysis and Knowledge Discovery, 2019, 3(4): 107-116.
[4] Shu Wenchen, Zhou Enguo, Li Daifeng, et al. Evolution analysis of informatics research topics based on co-authorship network community detection[J]. Information Science, 2020, 38(1): 75-81.
[5] Newman M E J. Networks[M]. New York: Oxford University Press, 2018.
[6] Yang Nan, Gong Dan Zhi, Li Min, et al. A survey of Web community detection technologies[J]. Journal of Computer Research and Development, 2005(3): 439-447.
[7] Zhou L, Lu K, Yang P, et al. An approach for overlapping and hierarchical community detection in social networks based on coalition formation game theory[J]. Expert systems with applications, 2015, 42(24): 9634-9646.
[8] Yang X S. A new metaheuristic bat-inspired algorithm[C]//Nature inspired cooperative strategies for optimization. Berlin: Springer-Verlag, 2010: 65-74.
[9] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences, 2002, 99(12):

7821-7826.
[10] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical review E, 2004, 69(2): 1-16.
[11] Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in networks[J]. Proceedings of the National Academy of Sciences, 2004, 101(9): 2658-2663.
[12] Danon L, Díaz-Guilera A, Duch J, et al. Comparing community structure identification[J]. Journal of statistical mechanics: theory and experiment, 2005, 2005(9): 1-10.
[13] Wang Li, Cheng Xueqi. Dynamic community detection and evolution in online social networks[J]. Chinese Journal of Computers, 2015, 38(2): 219-237.
[14] Li Jianhua, Wang Xiaofeng, Wu Peng. Research status of community detection methods based on local optimization[J]. Bulletin of Chinese Academy of Sciences, 2015, 30(2): 238-247, 180.
[15] 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): 1-11.
[16] 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.
[17] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435(7043): 814-818.
[18] Ahn Y Y, Bagrow J P, Lehmann S. Link communities reveal multiscale complexity in networks[J]. Nature, 2010, 466(7307): 761-764.
[19] Gregory S. Finding overlapping communities in networks by label propagation[J]. New journal of physics, 2010, 12(10): 1-26.
[20] Wu Xiaolan, Zhang Chengzhi. Research on overlapping community detection based on contribution degree of multi-label propagation[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(9): 949-957.
[21] Liu Shichao, Zhu Fuxi, Gan Lin. Overlapping community detection algorithm based on label propagation probability[J]. Chinese Journal of Computers, 2016, 39(4): 717-729.
[22] Xin Yu, Yang Jing, Xie Zhiqiang. A Link-Block algorithm for semantic overlapping community detection[J]. Journal of Software, 2016, 27(2): 363-380.
[23] Lanchichinetti A, Radicchi F, Ramasco J J, et al. Finding statistically significant communities in networks[J]. PloS one, 2011, 6(4): 1-18.
[24] Xie J, Szymanski B K. LabelRank: a stabilized label propagation algorithm for community detection in networks[C]//The 2013 IEEE 2nd international network science workshop. New York: IEEE, 2013: 138-143.
[25] Zhao Z, Li C, Zhang X, et al. An incremental method to detect communities in dynamic evolving social networks[J]. Knowledge-based systems, 2019, 163: 404-415.
[26] He Jing, Wang Zhixiao, Hou Mengnan, et al. Incremental dynamic community detection method based on topological potential[J]. Computer

Engineering and Design, 2019, 40(1): 45-52.
[27] Lin Y R, Sun J, Castro P, et al. MetaFac: community discovery via relational hypergraph factorization[C]//Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2009: 527-536.
[28] Chen J, Saad Y. Dense subgraph extraction with application to community detection[J]. IEEE transactions on knowledge and data engineering, 2010, 24(7): 1216-1230.
[29] Gopalan P K, Blei D M. Efficient discovery of overlapping communities in massive networks[J]. Proceedings of the National Academy of Sciences, 2013, 110(36): 14534-14539.
[30] Zhang Qin, Chen Hongmei, Feng Yunfei. An overlapping community detection method based on rough set and density peaks[J]. Application research of computers, 2018, 35(6): 1633-1637, 1698.
[31] Xia Z, Bu Z. Community detection based on a semantic network[J]. Knowledge-based systems, 2012, 26: 30-39.
[32] Anwar M M, Liu C, Li J. Discovering and tracking query-oriented active online social groups in dynamic information network[J]. World Wide Web, 2019, 22(4): 1819-1854.
[33] Liu W, Yue K, Wu H, et al. Markov-network based latent link analysis for community detection in social behavioral interactions[J]. Applied intelligence, 2018, 48(8): 2081-2096.
[34] Tang J, Zhang R, Yao Y, et al. An adaptive discrete particle swarm optimization for influence maximization based on network community structure[J]. International journal of modern physics C, 2019, 30(6): 1-21.
[35] Yang L, Cao X, He D, et al. Modularity based community detection with deep learning[C]//Proceedings of the Twenty-Fifth international joint conference on artificial intelligence (IJCAI-16). New York: AAAI Press. 2016: 2252-2258.
[36] Sperli G. A deep learning based community detection approach[C]//Association for Computing Machinery. Proceedings of the 34th ACM/SIGAPP symposium on applied computing. New York: ACM, 2019: 1107-1110.
[37] 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, 2008(10): P10008.
[38] Zhang Haitao, Liu Yashu, Zhang Xiaohui, et al. Topic discovery and netizen sentiment fluctuation analysis based on modularity—A case study of the "Sino-US trade friction" topic on Sina Weibo[J]. Library and Information Service, 2019, 63(4): 6-14.
[39] Clauset A, Newman M E J, Moore C. Finding community structure in very large networks[J]. Physical review E, 2004, 70(6): 066111.
[40] Zhu X J, Ghahramani Z, Lafferty J. Semi-supervised learning using Gaussian fields and harmonic functions[C]//Proceedings of the 20th international conference on machine learning. Washington DC: ICML. 2003: 912-919.
[41] Xing Xiangrui. Graph feature extraction method: Spectral Clustering detailed explanation[EB/OL]. [2019-11-10]. https://blog.csdn.net/weixin_{36474809}/article/details/89669623?

[42] Ma T, Wang Y, Tang M, et al. LED: a fast overlapping communities detection algorithm based on structural clustering[J]. Neurocomputing, 2016, 207: 488-500.
[43] Zhang Junxiang, Li Shuqin, Liu Bin. Community identification algorithm based on smooth L1-norm deep sparse autoencoder[J/OL]. [2020-02-05]. https://doi.org/10.19734/j.issn.1001-3695.2018.09.0743.
[44] Liu Bingyu, Wang Cuirong, Wang Cong, et al. Weibo community detection algorithm integrating multi-dimensional data based on dynamic topic model[J]. Journal of Software, 2017, 28(2): 246-261.
[45] Tian Bo, Fan Lingling. Research on community detection methods in online social networks based on interaction behavior[J]. Journal of Intelligence, 2016, 35(11): 183-188.
[46] Gu K, Wang L, Yin B. Social community detection and message propagation scheme based on personal willingness in social network[J]. Soft computing, 2019, 23(15): 6267-6285.
[47] Li C, Bai J, Du S, et al. Combining tag correlation and interactive behaviors for community discovery[J]. The computer journal, 2018, 62(5): 785-800.
[48] Li Gang, Chen Sijing, Mao Jin, et al. Spatiotemporal comparative analysis of natural disaster event hot topics on Weibo[J]. Data Analysis and Knowledge Discovery, 2019, 3(11): 1-15.
[49] Xin Chang K, Vilakon One P, Park D S. Movie recommendation algorithm using social network analysis to alleviate cold-start problem[J]. Journal of information processing systems, 2019, 15(3): 616-631.
[50] Zhang Jidong, Cai Xue. Research on friend information service recommendation based on community division and user similarity[J]. Information Studies: Theory & Application, 2019, 42(4): 151-157, 165.
[51] Chen Y C, Zhu W Y, Peng W C, et al. CIM: Community-based influence maximization in social networks[J]. ACM transactions on intelligent systems and technology, 2014, 5(2): 1-31.
[52] Kovács I A, Luck K, Spirohn K, et al. Network-based prediction of protein interactions[J]. Nature communications, 2019, 10(1): 1-8.
[53] Tu Cunchao, Yang Cheng, Liu Zhiyuan, et al. Survey on network representation learning[J]. Scientia Sinica Informationis, 2017, 47(8): 980-996.
[54] Ding Yonggang, Zhang Yuqin, Fu Qiang, et al. Precise recommendation of book resources based on SOM neural network and ranking factorization machine[J]. Information Studies: Theory & Application, 2019, 42(9): 133-138, 170.

Author Contributions:

Zhang Xinrui: Literature review and organization;
Zhou Honglei: Data collection, visualization analysis, and paper writing;
Zhang Haitao: Research direction and framework design, paper guidance and revision;
Sun Tong: Paper revision.

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