Computing and Informatics, Vol. 43, 2024, 94–125, doi: 10.31577/cai_2024_1_94

LOUVAIN-BASED FUSION OF TOPOLOGY AND ATTRIBUTE STRUCTURE OF SOCIAL NETWORKS

Qin Zhao, Yaru Miao, Jie Lian, Xiaoya Li, Dongdong An*

Shanghai Egineering Research Center of Intelligent Education and Big Data
Shanghai Normal University, Shanghai 200234, China
e-mail: andongdong@shnu.edu.cn

Abstract. With the increasing diversity and complexity of online social networks, effectively dividing communities presents a growing challenge. These networks are characterized by their large scale, sparse structure, and numerous isolated points. Traditional community detection methods lack consideration of node attribute information, thereby negatively impacting the accuracy of community detection. To address these challenges, this paper presents a novel Louvain-FTAS community detection algorithm that integrates topology and attribute structure. The proposed algorithm first selects attributes with positive effects to account for attribute heterogeneity. Subsequently, it utilizes a semi-local strategy to calculate topology similarity and information entropy to calculate attribute similarity. These values are combined to obtain the final node similarity matrix, which is then fed into the Louvain algorithm to maximize modularity and incorporate multi-dimensional attribute features to enhance community detection accuracy. The proposed model is evaluated through comparative experiments on two real datasets and artificial synthetic networks, demonstrating its rationality and effectiveness.

Keywords: Attribute networks, community detection, Louvain method, multi-dimensional fusion

Mathematics Subject Classification 2010: 68-T30

* Corresponding author

1 INTRODUCTION

The rapid expansion of online social networks has resulted in a greater number of connections between individuals, making it increasingly challenging to identify similar users. Consequently, the task of community detection has gained significant importance [1]. Conventionally, community detection methods have relied on topological information to identify users with similar characteristics. In order to assess the effectiveness of network partitioning into communities, Newman introduced the concept of modularity [2]. It serves as a measure to evaluate the quality of a network's division into communities, ensuring that nodes within a community exhibit denser connections compared to nodes across different communities. This criterion plays a crucial role in establishing a coherent and meaningful community structure. One of the most popular methods for modularity optimization is the Louvain method [3], which is a greedy algorithm that iteratively merges nodes into communities based on the modularity gain. This process aims to maximize the modularity score, thereby refining the community structure. By employing this iterative procedure, the Louvain method effectively enhances the accuracy and efficiency of community detection in online social networks.

While traditional topology-based methods for community detection have been widely used, they often overlook valuable user attributions that can aid in identifying user types. Recognizing the significance of this supplemental information, many researchers have incorporated user attributions to enhance the effectiveness of community detection. In line with this approach, Zhao et al. have proposed a hybrid method that incorporates user attributions to calculate user similarity [4, 5, 6]. This method takes into account the semantic similarity, which acts as a latent connection between users. By considering the semantic aspects of user attributes, the method enables a more comprehensive understanding of user relationships, thus improving the accuracy of community detection. This integration of semantic similarity provides a valuable extension to traditional topology-based methods, enhancing the overall performance of community detection algorithms.

In recent years, the significance of deep learning in community detection has grown exponentially. Deep learning techniques, with their ability to effectively capture intricate relationships among users, have revolutionized the field. In particular, methods based on Graph Neural Networks (GNNs) have gained widespread popularity for obtaining user embeddings that aid in user clustering [7, 8, 9, 10, 11]. However, it is important to acknowledge that these GNN-based methods often require a substantial amount of time and computational resources, rendering them impractical for application on large-scale social networks. The computational demands associated with training and applying deep neural networks can be overwhelming, making them unaffordable for real-world scenarios with extensive social network data.

To address the challenges discussed earlier, this paper introduces a novel approach called Louvain-FTAS (Fusion of Topology and Attribute Structure), which leverages the Louvain method for community detection. The authors recognize that node attribute information may not always have a positive impact on topology, prompting them to propose attribute heterogeneity as a method to select attributes with beneficial effects.

In Louvain-FTAS, a semi-local strategy is employed to obtain topological similarity, while information entropy is utilized to calculate attribute similarity. These two measures are combined to create a final node similarity matrix, effectively balancing the contributions of node attributes and topology information. By fusing both types of information, Louvain-FTAS optimizes the utilization of multi-dimensional attribute features, leading to improved accuracy in community detection. To achieve this, the paper integrates the node similarity matrix with the maximization of modularity using the Louvain algorithm. This integration enables Louvain-FTAS to fully exploit the advantages offered by multi-dimensional attribute features, enhancing the overall effectiveness of community detection.

The major contributions of this paper are summarized as follows:

- To obtain attribute enhancement networks in social networks, this paper proposes a selection approach based on two criteria: influence degree and information entropy. The aim is to identify homogeneous attributes that can effectively enhance the community detection process.
- To enhance the accuracy of community detection methods on social networks, this paper utilizes both topology information and attribute information to obtain fusion similarity. A new objective function is formulated, which incorporates the fusion similarity into the existing Louvain method. This objective function serves as a guiding principle for the optimization process, ensuring that the community structure is refined based on both topology and attribute information.
- To validate the rationality and validity of the proposed model, this paper conducts comparative experiments on two real datasets and artificial synthetic networks. The experimental results demonstrate the superiority of the algorithm proposed in this paper. The proposed model outperforms other algorithms in terms of accuracy and precision, providing more accurate community detection results.

The rest of the paper is organized as follows. Section 2 describes some related work about community detection. In Section 3, we propose our model Louvain-FTAS. The experimental results and the influencing factors are discussed in Section 4. Finally, conclusions are described in Section 5.

2 RELATED WORK

In the 1960s, Herbert Simon first proposed the concept that complex systems have modular structural characteristics [12]. In the field of sociology, researchers have found that communities generally exist in various complex networks [13]. In recent years, with the rise of social networks, the attention in the field of social network analysis has greatly increased [14, 15], including research on community detection algorithms. Currently, community detection algorithms can be mainly divided into topology-based community detection algorithms, attribute-based community detection algorithms, and hybrid algorithms that integrate topology and attributes.

2.1 Topology Based Community Detection Algorithm

Topology-based community detection methods can be classified into optimization methods and heuristic methods. Optimization methods [16, 17] typically set an objective function and iteratively calculate the optimal value. Representative algorithms of optimization methods include the spectral algorithm and modular maximization algorithm. The spectral algorithm [18, 19, 20, 21] transforms the community detection problem into a simple quadratic optimization problem and obtains the approximate optimal network partition by solving the eigenvector of the Laplace matrix. The modular maximization method [22, 23] finds the maximum value of the modularity function in the network, whose representative algorithms include the Louvain method and the simulated annealing algorithm. The Louvain algorithm [24, 25] has demonstrated good results in efficiency and effectiveness and can find hierarchical communities. Its optimization goal is to maximize the modularity of the whole network. The simulated annealing algorithm [26, 27] solves the local optimal problem. In contrast, heuristic methods determine the optimal division of communities by setting heuristic rules, and their representative algorithm is the GN algorithm [28]. In addition to the small world and scale-free characteristics, complex networks generally exhibit community characteristics.

Additionally, the LPA algorithm [29] proposed by Raghavan et al. is a graph-based semi-supervised learning method, which main idea is to predict the label information of unmarked nodes with the help of node labels. Although it runs very fast, it has randomness. Currently, the commonly used improved LPA algorithms are the SLPA algorithm [30] and COPRA algorithm [31], which can be used in overlapping communities. Furthermore, Rosvall and Bergstrom proposed the infomap algorithm based on random walk [32] and proposed an objective function based on information entropy through random sampling. At present, most algorithm improvements are based on the above algorithms, but only for topology.

Moving on to attribute-based community detection algorithms, there are relatively few community detection algorithms that only use attribute information. The K-means algorithm [33] is a classical clustering method that only considers attribute information in the clustering process. The K-SNAP algorithm [34] is a typical clustering algorithm based on attribute similarity.

2.2 Community Detection Algorithm Based on Attributes and Topology

There are several community detection algorithms combining attribute information and topology information: SA-Cluster [35] uses random walk to distance topology information and node attribute information, but the complexity is high. Inc-Cluster algorithm [36] decomposes the random walk distance matrix into multiple submatrixes, and uses the incremental update method for the submatrix to improve the calculation efficiency. GBAGC algorithm [37] uses Bayesian to model with variational method; BIGCLAM [38] modeled the network adjacency matrix and node attribute matrix respectively, and then optimized the model using gradient descent method.

As aforementioned, many existing methods for combining topology and attribute information in social networks only assign different weights to these two types of information, overlooking their underlying relationship. Other methods employ deep learning models to fuse these information sources, but they often struggle with the computational complexity associated with complex social networks.

To address these limitations, we propose a novel approach called Louvain-Based Fusion of Topology and Attribute Structure. By leveraging the low complexity of the Louvain method, our approach aims to achieve improved performance at a lower computational cost.

3 LOUVAIN ALGORITHM BASED FUSION OF TOPOLOGY AND ATTRIBUTE STRUCTURE

3.1 Formal Modeling and Definitions

Social networks can be modeled as graphs, in which nodes and edges correspond to users and social relationships, respectively. In this study, we incorporate attribute information into the graph representation. Specifically, a graph is defined as $G(V, E, A)$, where V denotes the node set, E denotes the edge set, and A denotes the attribute set of the network.

Since large social networks can contain a significant amount of attribute information, we focus on the L most significant attributes. We construct the filtered attribute set $A' = \{A_1, A_2, \dots, A_L\}$ by retaining only these top L attributes, where A_i denotes the i^{th} filtered attribute.

To compute the node similarity matrix S in the filtered network $G(V, E, A')$, we first compute the topology similarity matrix S_{top} of the original network $G(V, E)$ using the semi-local strategy. We then determine the attribute similarity matrix S_{att} of the filtered network $G(V, E, A')$ by employing information entropy calculations.

Then we use the function $z(\cdot)$ to combine the two similarity degrees:

$$S = z(S_{top}, S_{att}). \tag{1}$$

Finally, we use the community detection algorithm to fuse the node similarity matrix of topology and attribute information to improve the modular optimization objective function to complete the community division:

$$G_{cm} = f(G(V, E, A'), Q, S), \tag{2}$$

where $G_{cm} = \{c_1, c_2, \dots, c_n\}$ is the set of divided communities, and n is the number of communities. $f(\cdot)$ is the community detection algorithm, and Q is the modularity function.

Given the inclusion of numerous terms and definitions of social networks within the proposed method described in this paper, it is imperative to first provide clear definitions for these terms before proceeding with the method's description.

Definition 1 (Network structure representation)**.** The network with attributes can be represented as an undirected graph, denoted as $G(V, E, A)$, where $V = \{v_i\}$ represents the node in the network, the number of nodes is $|V| = N$, $E = \{e_{ij}\}$ represents the edge between any node v_i and node v_j, and the number of edges is $|E| = M$, and $A = \{a_i\}$ represents the attribute of the network node. The k^{th} attribute value of each node v_i is represented as $v_i^{a_k}$, and the number of attributes is $|A| = K$. It is worth noting that in the node network graph G, the topology is represented by (V, E), while the attributes are represented by (V, A).

Definition 2 (Attribute representation)**.** In the network structure, it is common for nodes to possess multiple attributes, each corresponding to different attribute values. For instance, a node may have attributes such as ID, gender, and age, denoted as $A_i = \{a_1, \ldots, a_k\}$, where $k \geq 2$. This implies that there are multiple types of attributes associated with each node. The attribute value vector corresponding to each node v_i can be represented as $\{v_i^{a_1}, \ldots, v_i^{a_k}\}$. Here, $v_i^{a_1}$ represents the attribute value of node v_i for the first attribute a_1, $v_i^{a_2}$ represents the attribute value for the second attribute a_2, and so on.

Definition 3 (Attribute information matrix)**.** The attribute information matrix W of nodes in the network structure can be expressed as:

$$W = \left[A_1^T, A_2^T, \ldots, A_K^T\right] = \begin{bmatrix} v_1^{a_1} & v_2^{a_1} & \cdots & v_N^{a_1} \\ v_1^{a_2} & v_2^{a_2} & \cdots & v_N^{a_2} \\ \vdots & \vdots & \ddots & \vdots \\ v_1^{a_k} & v_2^{a_k} & \cdots & v_N^{a_k} \end{bmatrix}. \tag{3}$$

Definition 4 (Similarity representation)**.** Considering balanced attribute and topology information, multi-dimensional fusion is carried out with node similarity S, including direct neighbor similarity $S_N(V_i, V_j)$, indirect neighbor similarity $S_{NN}(V_i, V_j)$ and weighted attribute similarity $S_{att}(V_i, V_j)$. As shown in Figure 1, v_5's direct neighbours are v_1, v_2, v_3, v_4, and the other nodes are indirect neighbours of v_5.

3.2 Model Overview

This paper presents the Louvain-FTAS community detection algorithm, which utilizes both the topology and node attributes to enhance the accuracy of community partitioning. The algorithm addresses the enrichment of network structure and outlier detection by incorporating attribute information. Additionally, it considers the

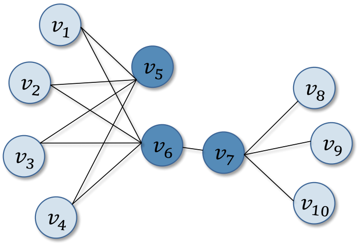


Figure 1. Example of direct/indirect neighbours

homogeneity of attributes and topology. To achieve a balanced integration of topological structure and attribute information, the paper proposes a multi-dimensional fusion approach to community division based on similarity.

The proposed algorithm consists of three main parts: attribute information filtering, similarity calculation, and multi-dimensional fusion community detection, which is shown in Figure 2. These parts can be further divided into the following three stages:

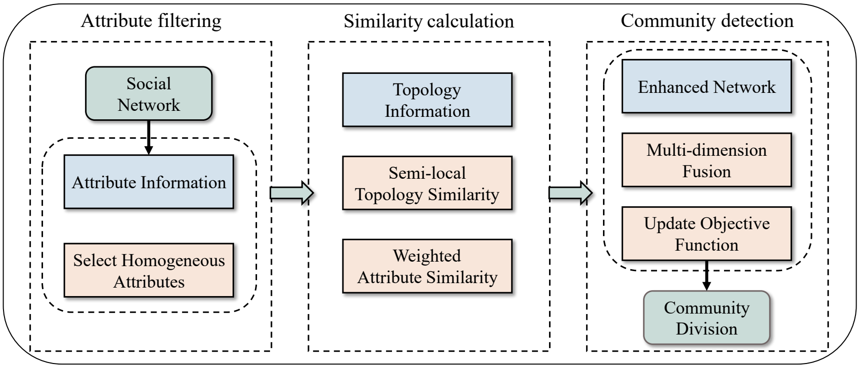


Figure 2. Overview of the proposed method

Stage 1. Use attributes to enhance the network to filter homogeneous attributes. In traditional community detection models that consider attributes, all attributes of nodes are typically included in the network structure. However, in this paper, we propose a different approach by suggesting that only influential attributes should be considered to enhance the effectiveness of community division. To achieve this, we introduce the concept of the attribute enhancement network, which allows us to filter out uninfluential attributes. The filtering process consists of two main parts:

1. Calculation of Structure Influence Degree and Information Entropy: In this

step, we calculate the influence degree and information entropy of each attribute within the network's structure. The influence degree quantifies the impact of an attribute on the division process, considering its contribution to the structural properties of the network. Information entropy measures the diversity and distribution of attribute values. These calculations provide valuable insights into the importance and variability of attributes within the network.

2. Selection of Influential Attributes: Based on predefined filtering rules, we select influential attributes from the calculated structure influence degrees and information entropies. These rules ensure that only attributes exhibiting a high influence on the division process and possessing sufficient diversity are considered influential. By filtering out less influential attributes, we focus on those attributes that truly contribute to the accuracy and effectiveness of community division.

Stage 2. Multi-dimensional feature fusion to obtain fusion similarity. In conventional methods, researchers often combine topology and attribute information iteratively to partition communities. However, in real-world scenarios, the importance of topology and attribute information may vary. Therefore, in this paper, we aim to strike a balance between topology and attribute information by conducting multi-dimensional fusion to obtain node similarity. To achieve this, we employ the following strategies:

1. Semi-local Topology Similarity: In this approach, we go beyond considering only common neighbors and incorporate second-order neighbors of nodes. By including second-order neighbors, we can more accurately capture the influence and connections of nodes within the network. This semi-local similarity strategy enhances the topological representation of nodes and provides a more comprehensive perspective on their relationships.
2. Weighted Attribute Similarity: For attribute information, we calculate the weights using information entropy. Information entropy helps us assess the diversity and significance of attributes. Based on these weights, we then obtain a weighted attribute similarity. This weighting mechanism allows us to emphasize attributes that carry more valuable and discriminative information, while de-emphasizing less informative attributes.

Stage 3. Optimize the objective function to maximize modularity. Firstly, we optimize the objective function to be specific to the attribute network. This optimization is achieved by utilizing the attribute enhancement network, which allows us to filter and select influential attributes. By aligning the objective function with the attribute network, we enhance the relevance and effectiveness of the community division process.

Next, in the process of multi-dimensional feature fusion, we incorporate attribute information by reflecting it on the edge weights. This fusion process combines both topology and attribute information, leveraging their complementary strengths. By updating the edge weights to incorporate attribute information, we achieve a more comprehensive and accurate representation of the network structure.

Once the optimized objective function is obtained, we proceed to divide the community based on this function. The Louvain algorithm is utilized to perform the community division, leveraging the optimized objective function. The effectiveness of the algorithm is then verified by evaluating the quality and accuracy of the resulting community divisions.

3.3 Attribute Enhanced Network

In this section, we focus on leveraging the attribute information of nodes to enhance the original topology and create an attribute enhancement network. However, we also acknowledge that in large networks, the abundance of attribute information can lead to information redundancy and potentially disrupt the integrity of the original topology.

To address this issue, we propose a filtering process after constructing the attribute enhancement network. This filtering step involves removing attribute virtual nodes that may introduce unnecessary redundancy and noise to the network. By filtering out these attribute virtual nodes, we aim to streamline the attribute information and preserve the integrity of the original topology.

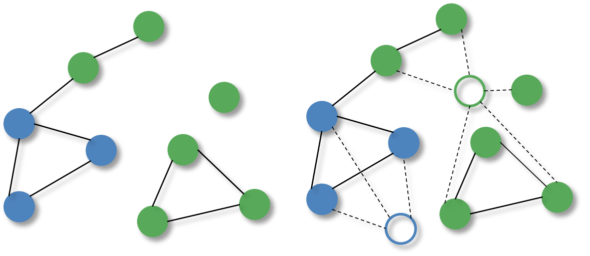
3.3.1 Network Construction

In real-world scenarios, many network structures exhibit sparsity, with numerous isolated nodes scattered throughout the network. These sparse structures and isolated points can significantly impact the division of community structures within the network. However, simply discarding these nodes may not be advisable since even outliers can have connections or share similar interests with other nodes.

To address this challenge, we leverage node attributes to enhance the network structure and enrich the available feature information. The node attribute types considered in this paper are categorized as continuous, discrete, and text types. For the sake of modeling simplicity, we assume that only discrete attribute types are present. We convert attribute information into virtual nodes, which allows us to construct an attribute enhancement network that measures the influence of attributes on the connections between nodes.

The construction process is depicted in Figure 3. The original topology network is illustrated in Figure 3 a), where blue and green nodes represent two distinct communities. Figure 3 b) shows the network enhanced by attributes. In this enhanced network, two hollow nodes represent different attribute values within the same attribute category. The dotted lines indicate that solid nodes possess corresponding attribute values.

By examining the structures of the two networks, we observe that the isolated nodes in Figure 3 a) and separate isolated communities have been connected through the inclusion of virtual attribute nodes in Figure 3 b). This enhancement process strengthens the relationships between these nodes, ultimately improving the relevance and connectivity within the network.



a) Topological network b) Enhanced network

Figure 3. Single attribute network

3.3.2 Attribute Selection

We acknowledge that certain large social networks contain a significant amount of attribute information. For instance, the CiteSeer dataset consists of 3 703 features, while the Cora dataset includes 1 433 features. If all attributes are directly used as virtual nodes to enhance the network, the resulting network structure would become excessively complex. This would significantly increase the time complexity of community detection algorithms and potentially obscure the original topology structure of the network.

Furthermore, some attributes may have a wide range of values, which may not effectively capture the similarity between nodes, as depicted in Figure 4. These factors can ultimately lead to a decrease in the accuracy of community detection algorithms. Therefore, it is essential to filter and control the attributes, limiting them to a manageable number, denoted as L, based on the size of the network. By carefully selecting and filtering attributes, we can strike a balance between capturing meaningful information and maintaining the computational efficiency and accuracy of the community detection process.

In this paper, the selection of attributes follows specific conditions to ensure their effectiveness in enhancing the network structure and promoting accurate community division.

Firstly, attributes should exhibit a high probability of nodes with compact structures sharing the same attribute value. This condition ensures that nodes with similar attributes are more likely to form cohesive communities within the network.

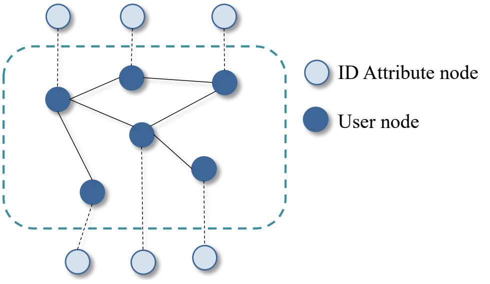


Figure 4. Network with ID attribute

Secondly, although nodes may not be closely connected in terms of topology, there should be a strong attribute contribution between nodes. This attribute contribution plays a crucial role in establishing connections between nodes within the network.

Lastly, the screening process aims to minimize the formation of new connections between different communities through attributes. This helps maintain the distinctiveness of communities and prevents excessive interconnectivity.

To achieve these goals, this section employs the concepts of structure influence degree and information entropy to quantify the importance of each attribute. The importance degrees are then reversed to select the first attribute as the virtual node for the attribute enhancement network. By filtering attributes based on their importance, we can strengthen the attribute structure and ultimately improve the accuracy of community division results.

In this paper, we define that K attributes have a structure attribute matrix $H^k \in \mathbb{R}^{N\times N}, k = 1, 2, \ldots$, and H^k represents the influence of attribute nodes on the relationship between topological nodes in the enhanced network. On the k^{th} attribute, if node v_i and v_j have the same attribute value, then $H_{ij}^k = 1$, otherwise $H_{ij}^k = 1$. At this time, the enhanced adjacency matrix of the network topology is defined as $A^k \in \mathbb{R}^{N\times N}$, and the enhanced neighbor matrix A^k is calculated from the structure attribute matrix H^k and the topology adjacency matrix A. If the sum of the two corresponding to the index (i, j) is greater than 0, then $A_{ij}^k = 1$, otherwise $A_{ij}^k = 1$. The specific formula is as follows:

$$A_{ij}^k = \begin{cases} 1, & \text{if } H_{ij}^k + A_{ij} > 0, \\ 0, & \text{if } H_{ij}^k + A_{ij} = 0. \end{cases} \tag{4}$$

In order to obtain the influence of attribute information on the topology network, based on the enhanced connection matrix, the structure attribute influence degree $\text{Affect}\left(A^k\right)$ of each attribute can be calculated by the following formula:

$$\text{Affect}\left(A^k\right) = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(A_{ij}^k - A_{ij}\right)}{\theta \cdot N}, \tag{5}$$

where A^k is the enhanced connectivity matrix, A is the topological adjacency matrix, 0 is a super parameter used to adjust the value difference.

However, it is not enough to rely solely on structure attribute influence degree Affect $\left(A^k\right)$, because it only uses topology. Therefore, we also consider the probability of the occurrence of the same attribute value. Information entropy can reflect the chaotic degree of attributes in the attribute enhancement network, that is, the amount of information contained. The calculation formula of information entropy of each attribute is:

$$E(k) = -\sum_{i=1}^{|A_k|} p_i \ln p_i, \tag{6}$$

where $|A^k|$ represents the number of discrete values of attribute k, and p_i represents the frequency of the i^{th} attribute value in the whole network:

$$p_i = \frac{\text{count}\left(a_{k,i}\right)}{N}, \tag{7}$$

where count $\left(a_{k,i}\right)$ represents the number of nodes in the social network whose attribute A_k is a_i, and N represents the number of nodes.

The smaller the entropy value, the more information the attribute carries. The more valuable this attribute is, the more accurate the judgment of community detection structure is. Because the entropy value range of different attributes is different, it is difficult to compare. So in order to make entropy comparable, we finally scale the value to $[0, 1]$ to represent information entropy.

$$E_w(k) = \frac{E(k) - E_{min}}{E_{max} - E_{min}}, \tag{8}$$

where E_{max} and E_{min} are the maximum and minimum of information entropies, respectively.

To begin, we calculate the information entropy of each attribute using Equation (8). Any attributes with an information entropy greater than the threshold value τ are eliminated. Next, we calculate the structure attribute influence degree using Equation (5). The attribute information entropy is then sorted from largest to smallest. Any attributes with a structure attribute influence degree greater than the threshold value μ are eliminated. Finally, we select the first L attributes.

3.4 Fusion of Topology and Attribute

In an attribute-enhanced network, both attribute similarity and topology similarity can be used to describe the correlation between nodes, providing effective assistance in improving community detection algorithms. In this section, we calculate both similarity degrees and fuse them to obtain a node similarity matrix that reflects the relationship between nodes in terms of both topology and attributes. This approach provides more comprehensive information about the nodes, which can be used to improve subsequent analyses.

3.4.1 Attribute Similarity

To construct an effective attribute similarity matrix, the uninfluential attributes are filtered out through attribute filtering, retaining only the influential attributes that actively contribute to the community detection process. The weighted attribute similarity, denoted as S_{att}, is then constructed using attribute weights derived from information entropy.

Considering that each attribute has a varying impact on the network topology and nodes, it is essential to capture their differences. To achieve this, the paper utilizes information entropy to calculate attribute weights. The first step is to obtain the information entropy redundancy:

$$e_i = 1 - E\left(A_i\right). \tag{9}$$

Then, we can normalize e_i to ensure that it falls within the range of $[0, 1]$:

$$\omega_i = \frac{e_i}{\sum_{i=1}^{l} e_i}. \tag{10}$$

Equation (10) provides us with weights that can be used to calculate edge weights. By using these weights, we can calculate the attribute similarity between nodes v_i and v_j. If the k^{th} attribute value of both nodes is the same, we can say that there is a certain attribute similarity between them. The formula for calculating the weighted similarity of node attributes is given below:

$$S_{att} = e^{-\omega^T \times \|f(v_j) - f(v_i)\|_2}, \tag{11}$$

where $\|f\left(v_j\right) - f\left(v_i\right)\|_2$ represents the two normal forms of attribute row vector difference between node v_i and v_j, and ω_i is the weight column vector of node attributes, thus obtaining the attribute similarity between two nodes.

3.4.2 Topological Similarity

To effectively detect the community structure, this section explores the scope of node interaction. It is evident that considering the entire scope of global interaction would be impractical and time-consuming, especially for large-scale networks. However, relying solely on neighboring nodes in local interaction leads to low accuracy in community division. To strike a balance between the advantages and disadvantages of global and local approaches, this paper adopts a semi-local strategy.

In real networks, besides direct friends, friends of friends also have a high probability of forming connections. The semi-local strategy employed in this research selects both direct and indirect neighbors of a node to interact with each other, utilizing their topology information to calculate similarity. The semi-local similarity method considered in this section takes into account not only common neighbors but also all the common nodes among the nearest and lower nearest neighbors of the node. This approach allows for a more accurate reflection of the node's influence.

From the semi-local strategy, we choose the direct and indirect neighbors to calculate node similarity based on topology. The similarity between nodes v_i and v_j in terms of direct neighbors indicates that the more common neighbors they have, the more similar their node structures are. The direct similarity, denoted as S_N, between node v_i and v_j, can be expressed as follows:

$$S_N\left(v_i, v_j\right) = \frac{N_{ij}}{N_i + N_j}, \tag{12}$$

where N_{ij} represents the set of common neighbors of node v_i and v_j, and N_i and N_j represent the number of direct neighbors of node v_i and v_j, respectively.

Additionally, the similarity among indirect neighbors captures the similarity between the adjacent network structures of nodes. Hence, the indirect neighbor similarity, denoted as S_{NN}, depends on the similarity of direct neighbors. In this study, the similarity of indirect neighbors is regarded as the superposition of similarities between neighbors and their neighbors. Therefore, the formula for calculating the similarity $S_{NN}(v_i, v_j)$ of indirect neighbors is as follows:

$$S_{NN}\left(v_i, v_k\right) = S_N\left(v_i, v_j\right) \times S_N\left(v_j, v_k\right), \tag{13}$$

where node v_j is the direct neighbor of node v_i, v_k is the indirect neighbor of v_i.

3.5 Fusion Node Similarity

In this paper, we acknowledge that attributes and topology are two distinct types of information in the community division process, and these two types of information are heterogeneous. The heterogeneous attributes of nodes can provide contradictory information about the network topology. Only by effectively leveraging both types of information can we improve the accuracy of community division. Therefore, we aim to balance them based on topological compactness and attribute homogeneity.

Topological compactness refers to the internal closeness and external sparsity of the network's connections. Attribute homogeneity implies that nodes within a community possess similar attributes, and these attributes themselves influence the network's topology.

By combining topology and homogeneous attributes, we can provide more comprehensive information for community detection. An effective fusion method can mitigate the impact of noise, compensate for the limitations of relying solely on topology, and ultimately achieve high-quality community partitioning results. In line with our proposed attribute-topology homogeneity, we seek to measure the corresponding strength of connection between attributes and topology. This is accomplished by combining topological similarity and attribute similarity:

$$S\left(v_i, v_j\right) = \alpha S_N + (1 - \alpha) S_{NN} + \beta S_{att}. \tag{14}$$

In this context, the parameters α and β are mixed parameters that govern the proportion of topological similarity and attribute similarity, respectively. These parameters have a value range of $[0, 1]$. By adjusting the values of α and β, we can control the contribution of topological similarity and attribute similarity to the overall similarity measurement.

To obtain the similarity matrix, we calculate the similarity value between any two nodes, v_i and v_j. The subsequent experiments allow us to determine the range of parameters that yields the best division result. Through empirical analysis, we identify the optimal values for α and β in order to achieve the most favorable community partitioning outcome.

3.6 Louvain-FTAS Community Detection Algorithm

After exploring the potential social relations between isolated points and other nodes, this paper introduces an enhanced version of the Louvain algorithm called Louvain-FTAS. Building upon the foundation of the original Louvain algorithm, this enhanced approach incorporates updates and improvements to the objective function for modular maximization. Through this iterative process, the algorithm ultimately generates the desired community division results. The specific process is shown in Figure 5.

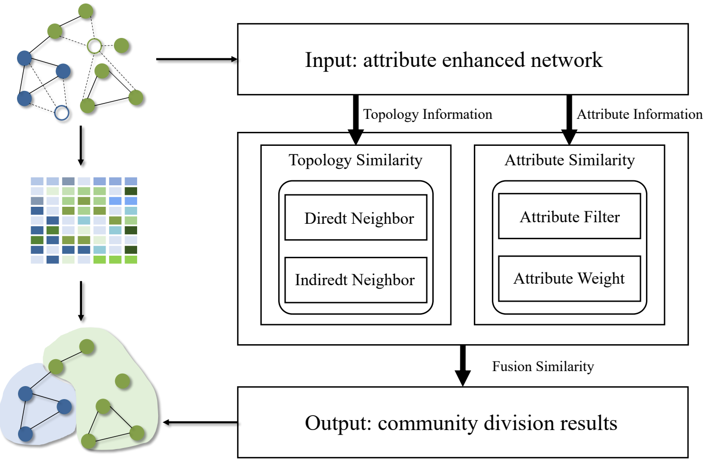


Figure 5. Specific process of community detection

The Louvain algorithm is a widely used and effective method for community partitioning, which leverages the concept of maximum modularity. The algorithm starts by assigning each node to its own separate community and then proceeds with iterative steps. During each iteration, the algorithm compares the allocation of each node with the community allocations of its neighboring nodes. The node is then assigned to the community that results in the largest increase in modularity. The formula used to maximize modularity is as follows:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{d_i d_j}{2m} \right] \delta \left(c_i, c_j \right), \tag{15}$$

where m represents the number of edges of the enhanced network, A_{ij} represents the adjacency matrix of the network, $\frac{d_i d_j}{2m}$ represents the connection probability of node v_i and v_j, d_i represents the degree of node v_i, and $\delta \left(c_i, c_j \right)$ represents whether node v_i and v_j currently belong to the same community.

Building upon the Louvain algorithm, this paper enhances the community partitioning process by integrating multi-dimensional topology information and attribute information in a synchronized manner. This integration ensures that both types of information are considered in a unified framework, leading to an improved modular maximization objective function. By incorporating these additional dimensions of data, the algorithm can achieve a more comprehensive and accurate assessment of community structures within the network:

$$Q = \frac{1}{m} \sum_{ij} \left[W_{ij}^{att} - \frac{k_i \times k_j}{2m} \right] S \left(v_i, v_j \right) \delta \left(c_i, c_j \right), \tag{16}$$

where W_{ij}^{att} is a weighted adjacency matrix. When node v_i and v_j have the same attribute value, attribute weights are assigned to edges.

The proposed Algorithm 1 is a multi-dimensional feature fusion community detection algorithm. In community detection tasks, it has been observed that focusing on high-influential attributes, which are closely associated with the community structure, can lead to improved community division results compared to using all attributes of nodes. To address this, our algorithm introduces a ranking mechanism that assesses the influence and information entropy of each attribute within the network structure. By calculating the influence of attributes, we can identify those that have a significant impact on the community structure. This influence can be measured by analyzing the attribute's contribution to the overall connectivity patterns and interplay between nodes within the network. Attributes with higher influence scores are considered more important in capturing the underlying community structure. In addition to influence, we also consider the information entropy of attributes. This metric captures the amount of uncertainty or randomness present in the attribute values. Attributes with low entropy indicate that they exhibit more consistent patterns within communities, making them valuable in community detection tasks.

Combining the influence and entropy measures can help rank the attributes and identify the high-influential ones that are likely to contribute significantly to the community division process. This allows us to filter out low-influential attributes, which may introduce noise or unnecessary complexity to the analysis. By selecting high-influential attributes and filtering out low-influential ones, we aim to obtain more accurate and reliable community division results in community detection tasks.

Algorithm 1: Louvain-FTAS

Input: Attribute enhanced network $G(V, E, A')$, S_N, S_{NN}, S_{att}
Output: Final community division results $G_{cm} = \{c_1, c_2, \ldots, c_n\}$

Initialize input parameters;
Calculate fusion similarity $S(v_i, v_j)$ according to Equation (14);
Calculate the attribute weight according to Equation (10) to obtain the weighted adjacency matrix W_{ij}^{att};
for *each* $v_i \in G$ **do**
    for *each* $c_i \in C$ **do**
        Calculate the modular degree ΔQ according to Equation (16);
        if $max\Delta > 0$ **then**
            $v_i \in c_i$
        end
        else
            $v_i \notin c_i$
        end
    end
end

---

4 EXPERIMENT

In this section, we will conduct experiments on two public datasets to assess the effectiveness of the Louvain-FTAS algorithm in detecting communities within social networks. By performing these experiments, we aim to validate and evaluate the performance of the proposed algorithm in comparison to other existing methods.

4.1 Datasets

4.1.1 Real Network Datasets

PolBlogs [39]: This data set is a non confidential network composed of 1 490 Internet blogs (nodes) about American politics. There are 16 711 hyperlinks (edges) between online blogs, and each node has a "value" attribute to indicate their political orientation, that is, liberals or conservatives. However, the value attribute comes from the blog directory and is manually marked, which has certain errors. There are 7 communities in the network: Blog Catalog, Blog Pulse, Blog arama, Campaign Line, Labeled Manually, Left Directory, eTalkingHead, and some blogs come from two or more communities.

Cora [40]: This is a machine learning paper classification network, including 2 708 papers (nodes) and 5 429 references (edges). Each node has a 1 433 dimensional binary vector, representing the missing/existing words in the word dictionary collected from the paper corpus. The paper is divided into seven subcategories: case-based reasoning, genetic algorithms, neural networks, probabilistic methods, reinforcement learning, rule learning and theory.

Table 1 demonstrates the number of nodes, edges, communities and attributes of the two data sets.

Dataset	Node	Edge	Community	Attribute
PolBlogs	1 490	16 711	2	1
Cora	2 708	5 429	7	1 433

Table 1. The statistics of the real-world datasets

4.1.2 Composite Network Datasets

The LFR benchmark network [41] is also selected as a supplement to the attribute network. The degree of nodes and the size of communities in LFR networks obey power law distribution, so they are closer to real-life networks. The proposed community detection algorithm can be tested with real data sets.

The datasets generated by LFR datum network have different mixed parameters μ, different scales n, and different minimum community sizes c_{min}. The mixed parameter μ controls the degree of network mixing, with larger values indicating greater network mixing, making it more difficult to accurately detect communities. In each LFR datum network, r attribute vectors are added to each node to generate numeric attribute (*num*), binary attribute (*bin*) and absolute attribute (*cate*) networks, respectively. The probability of similarity of individual attributes within a population is p, and the subspace size of the population is t.

	n	μ	c_{min}	k
LFR1	1 000	0.1	10	10
LFR2	1 000	0.3	10	10
LFR3	1 000	0.5	10	20

Table 2. LFR benchmark network and parameters

4.2 Baselines

Louvain algorithm [24]: This algorithm is the initial model of the model in this paper. It is a greedy optimization algorithm, with the modular optimization as the objective function. Finally, modularity will not change any more, and hierarchical community structure can be found.

FastGreedy algorithm [27]: This algorithm is also an algorithm based on greedy optimization, which is used to detect the hierarchical aggregation algorithm of community structure and select the hierarchical partition with the largest Q value to obtain the final community structure.

K-means algorithm [33]: This algorithm only considers the method of attribute clustering on attribute graph, initializes their center points randomly, calculates the distance from each node to the center point, divides the nearest one into one category and obtains the final partition result after iteration.

SA Cluster algorithm [36]: Random walk distance topology information and node attribute information, and generate weight value of their distance into a weighted network. Then use K-Medoids to recalculate the matrix every time the attribute weight is updated, which is very complex.

MOEA-SA algorithm [38]: It is a multi-objective evolutionary algorithm. At the later stage of detection, it combines attribute similarity and modularity optimization formula, which is currently a more effective method for attribute community division.

4.3 Evaluation

In the past few decades, many evaluation criteria have been proposed to quantify the quality of a partition. For example, modularity, conductivity, NMI, ARI, Purity, and F-score. In this paper, three most commonly used evaluation indicators are selected to evaluate community performance: Normalized Mutual Information (NMI) [42], Overlapping Modularity (Q_{OV}) [43] and ARI [44].

Since the community structure in real networks is usually unknown, modularity can be used to measure the division result of unknown communities. This paper uses overlapping modular Q_{OV} to evaluate the impact of community detection algorithms. Generally, the larger the Q_{OV}, the better the result of community division.

$$Q_{OV} = \frac{1}{2m} \sum_{i} \sum_{v \in c_i, w \in c_i} \frac{1}{O_v O_w} \left[A_{vw} - \frac{d_v d_w}{2m} \right]. \tag{17}$$

When the number of communities in the network is known, this paper uses NMI and ARI to evaluate the performance of the algorithm. NMI is a similarity measure derived from information theory. It believes that if two partitions are similar, there is little need for additional information to infer a partition allocation from another partition. Its definition is as follows:

$$NMI = \frac{2I(X;Y)}{H(X) + H(Y)}, \tag{18}$$

where $I(X;Y)$ represents mutual information between partition X and Y, $H(X)$ represents entropy of X, and NMI ranges from 0 to 1. When the obtained partition is completely independent of the actual partition, it is $NMI = 0$. In contrast, when the obtained partition exactly matches the actual partition, it is $NMI = 1$.

In addition, this paper also selects ARI to quantify the accuracy of community detection. The ARI value range is $[-1, 1]$. The higher the value, the better the quality of community division, and the closer to the real situation.

Specific definitions are as follows:

$$ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}. \quad (19)$$

4.4 Experimental Settings

Among the five comparison algorithms mentioned, Louvain, fastGreedy, and K-means are not attribute-based community detection algorithms. They detect communities directly in the original topological network without considering attributes. On the other hand, the SA-Cluster and MOEA-SA algorithms take attributes into account and create an attribute enhancement map before performing community division.

The proposed Louvain-FTAS algorithm in this paper is a community detection algorithm based on multidimensional feature fusion. It involves four main steps: building an attribute-enhanced network, attribute filtering (not applicable for datasets with only one attribute, such as PolBlogs), topology-attribute fusion, and community detection.

During the community detection process, both the K-means algorithm and the SA-Cluster algorithm require specifying the number of iterations. This paper selects the better results obtained under different iterations. Additionally, there is a threshold involved in determining the fusion similarity. To determine an appropriate threshold, this paper treats the parameters α and β as independent variables and tests various threshold ranges during the experiments.

Based on the experimental results, it is found that when α ranges from 0.7 to 0.8, the modularity and NMI (Normalized Mutual Information) mean values are higher for the three artificial network datasets. Similarly, when β ranges from 0.4 to 0.7, the modularity and NMI fluctuations are relatively small. Therefore, the parameter values chosen in Section 3 are within this interval, as it yields favorable results.

4.5 Experimental Results and Analysis

4.5.1 Analysis of the Overall Experimental Results

In this section, simulation experiments are conducted on both real attribute network datasets and LFR attribute networks. The goal is to evaluate the performance of the

Louvain-FTAS algorithm. This is achieved by comparing the overall experimental results and assessing the quality of community partitioning, as well as analyzing the threshold parameters.

The validity of fused attributes is discussed by comparing the experimental results of attribute enhancement and topology structure. By examining the performance of attribute enhancement and its impact on the community detection process, the effectiveness and significance of incorporating attribute information into the algorithm are evaluated. The experimental results provide insights into the benefits and contributions of fused attributes in improving the accuracy and quality of community partitioning.

Metrics	Modularity Q		NMI		ARI	
Network	PolBlogs	Cora	PolBlogs	Cora	PolBlogs	Cora
Louvain	0.4272	0.4853	0.4156	0.4562	0.7659	0.2664
fastGreedy	0.4270	0.3571	0.4272	0.4622	0.7821	0.2646
K-means	0.4263	0.4024	0.4453	0.4138	0.7891	0.0800
SA-Cluster	0.4382	0.4412	0.4521	<u>0.4790</u>	0.7953	**0.2973**
MOEA-SA	<u>0.4465</u>	<u>0.4967</u>	<u>0.4611</u>	0.4640	<u>0.8005</u>	0.2680
Louvain-FTAS	**0.4469**	**0.5027**	**0.4663**	**0.4837**	**0.8263**	<u>0.2855</u>

Table 3. Comparison of different algorithms on two real networks

Table 3 provides a comparison of modularity Q, NMI, and ARI on two real datasets. Here are some key observations:

1. Comparing the algorithms that consider only topological structure (Louvain, fastGreedy, and K-means) with those incorporating attributes, it is evident that adding node attributes improves the accuracy of community detection, as indicated by the higher modularity values. Among the attribute-based algorithms, the SA-Cluster algorithm maps all attributes to the network, but the excessive number of attributes can affect community boundaries.
2. NMI, which requires real community information, is used for comparison on the PolBlogs and Cora datasets. The results demonstrate that all three algorithms perform better after incorporating attributes compared to considering topology alone. This indicates that combining topology and attribute information leads to community detection results that are closer to the actual community structure. Additionally, the algorithm proposed in this paper outperforms the other comparison algorithms on the PolBlogs and Cora datasets, especially as the network complexity increases.
3. ARI is also compared among different algorithms. The results show that the community delimitation quality is better for the PolBlogs dataset due to the dataset's distinct network boundaries. On the other hand, the Cora dataset, with its 1 433 attributes, yields a lower ARI value for the overall algorithm. Overall, combining the detection of topology and attributes yields better accuracy in community detection compared to relying solely on topology.

These observations highlight the advantages of the Louvain-FTAS algorithm in accurately detecting communities by leveraging both topology and attribute information in real datasets.

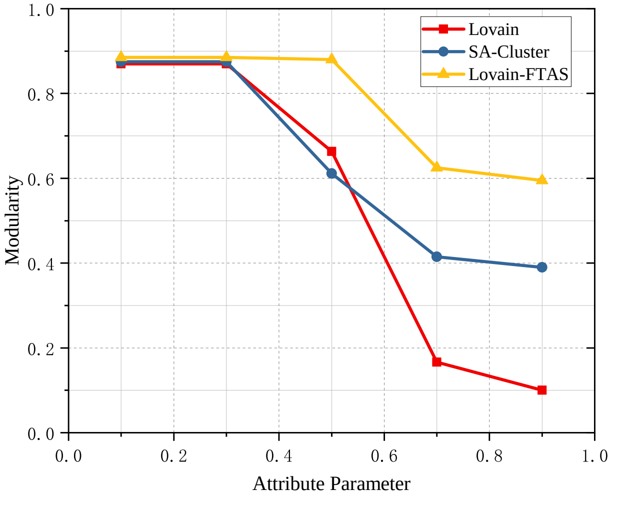
It is notable that in the case of the PolBlogs dataset where only one attribute is available, detecting overlapping communities solely based on attributes is challenging. However, the experimental results indicate that even when only a few attributes are fused with topology information, there is a significant improvement in the detection of overlapping communities. This suggests that the combination of attribute and topology information can enhance the overall effectiveness of community detection, even in datasets where attributes alone may not be sufficient to identify overlapping communities.

To provide a comprehensive comparison of algorithm performance, LFR networks are utilized to test the modularity and NMI values of community partitioning. This allows for further investigation into the impact of attribute parameters on algorithm performance. Attribute parameters play a crucial role in determining the attribute enhancement network. Thus, the main algorithms related to attribute enhancement, namely the Louvain algorithm, SA-Cluster algorithm, and the algorithm proposed in this paper, are selected for comparison.

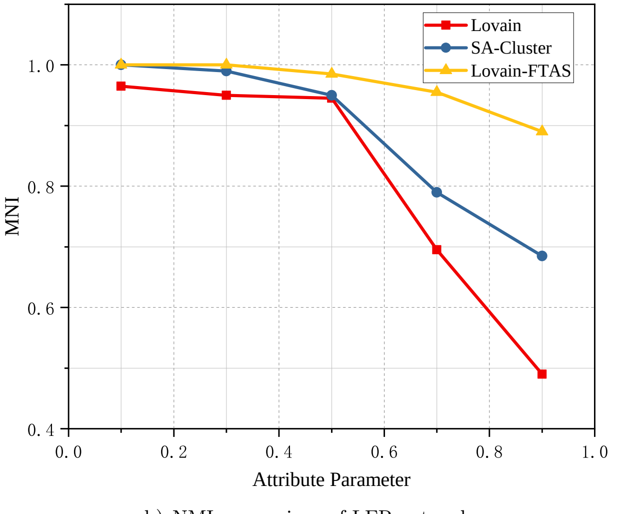
The range of attribute parameters considered in the experiments is $[0.1, 0.9]$. The experimental results, showcasing the modularity and NMI values, are presented in Figure 6. This graphical representation allows for a visual understanding of how varying attribute parameters influence algorithm performance. By analyzing the results, the paper gains insights into the optimal range of attribute parameters and their impact on community partitioning accuracy.

From Figure 6, several observations can be made regarding the impact of attribute mixing parameters on algorithm performance:

1. As the attribute parameters increase (greater than 0.5), the attribute enhancement network becomes more complex, and the network structure becomes less apparent. This leads to a decrease in both modularity and NMI for the Louvain algorithm, SA-Cluster algorithm, and the algorithm proposed in this paper.
2. In Figure 6 a), the Louvain algorithm and SA-Cluster algorithm exhibit a significant decline in modularity when the parameter exceeds 0.3. This suggests that the presence of heterogeneous attributes has a negative impact on the accuracy of community detection. However, the Louvain-FTAS algorithm maintains a modularity close to 0.6 even when the parameter approaches 0.9, indicating its strong performance in such scenarios.
3. In Figure 6 b), all three algorithms achieve higher NMI values when the parameter is below 0.6. However, as the degree of attribute chaos increases, the NMI values of the Louvain and SA-Cluster algorithms fluctuate. On the other hand, the algorithm proposed in this paper demonstrates relatively stable NMI values, indicating its ability to maintain good performance even in situations where the network structure is less evident and heterogeneous attributes are present.



a) Modularity comparison of LFR network



b) NMI comparison of LFR network

Figure 6. Attribute parameter change community detection algorithms comparison

This ensures that the partition results align closely with the actual community structure.

These findings highlight the advantages of the Louvain-FTAS algorithm in effectively handling complex attribute scenarios and maintaining accurate community partitioning results.

4.6 Analysis of Attribute Filtering Experiments

In this section, the PolBlogs dataset is not used due to its single attribute. Instead, the Carnegie University in Facebook 100's online social network [45] is chosen as an example. The method proposed in this paper is employed to calculate the structure-attribute influence and information entropy values of the network. These values are utilized to filter out homogeneous attributes for further feature fusion.

The filtering process involves removing attributes with high information entropy and sorting the Affect(a_i) values. A smaller Affect(a_i) indicates a lesser contribution to node partitioning. The experimental results are presented in Table 4. Among the attributes, the "dorm" attribute had many data null values and was eliminated. As a result, four attributes, namely "student fac", "gender", "second major", and "year", were filtered out for subsequent community detection.

Attributes	$E(A_d)$	Affect(a_i)
student fac	0.3813	3 969
second major	0.5702	2 029
dorm	0.6640	1 394
year	0.6915	913
gender	0.8203	2 914
high school	0.8442	177
major index	0.8513	347

Table 4. Attribute contribution comparison table

This filtering process ensures that only relevant and informative attributes are retained, enhancing the accuracy of community detection in the subsequent analysis.

4.7 Attribute Enhanced Network Role Analysis

To demonstrate the effectiveness of utilizing attribute information as virtual nodes to enhance the network, this paper incorporates the Louvain algorithm, fastGreedy algorithm, and LPA algorithm in addition to the original experiments.

Comparisons are made between these algorithms under the initial topological network and the attribute-enhanced network. The experimental results are presented in Table 5. The findings indicate that community detection using the attribute-enhanced network yields results that are closer to the actual situation, thus improving the accuracy of the community detection algorithms.

By incorporating attribute information and leveraging it as virtual nodes, the algorithms can better capture the underlying structure and characteristics of the network, leading to more accurate community partitioning. This validates the effectiveness of attribute enhancement in enhancing the performance of community detection algorithms.

	Louvain	LPA	FastGreedy
Initial Topological Network	12	278	277
Attribute Enhanced Network	4	2	8
Number of Real Communities	2	2	2

Table 5. Number comparison of community divisions

From Table 5, it can be observed that in the initial topological network, the LPA and FastGreedy algorithms show low accuracy when dividing more than 200 communities. This is primarily due to the large number of isolated points in the dataset, which are assigned to separate communities by these algorithms.

To ensure fairness in the overall experimental comparison, the algorithm based on topological structure eliminates outliers. Among the compared algorithms, the Louvain algorithm demonstrates relatively good partition results. Therefore, the Louvain algorithm is chosen in this paper for further improvement.

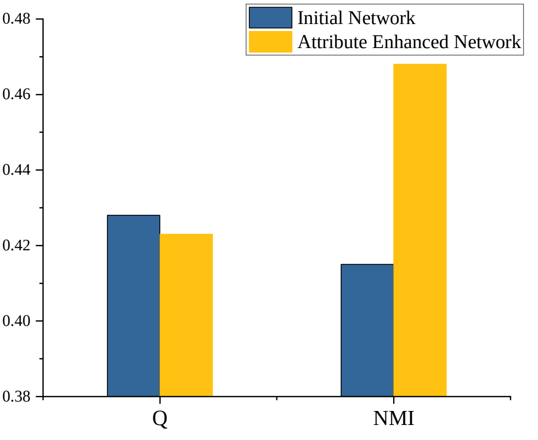
Furthermore, upon incorporating attributes, it can be observed that the results of all three algorithms are closer to the actual number of communities. This signifies that the inclusion of attribute information improves the accuracy of community detection, enabling the algorithms to better capture the underlying community structure.

From Figure 7, it is evident that the NMI values of the two community detection algorithms improve in the attribute-enhanced network compared to the network with a single attribute. This indicates that the accuracy of community division is enhanced by incorporating the node's own attributes.

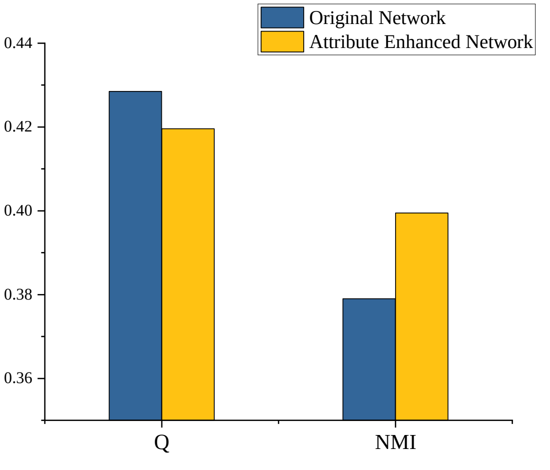
However, the modularity of both algorithms decreases in the attribute-enhanced network. This can be attributed to the introduction of attribute virtual nodes, which create new links that may blur the boundaries between communities. As a result, in scenarios with only one attribute, the division of communities can become less distinct.

In real-life networks, nodes often possess multiple attributes. It is impractical to directly incorporate all attributes as virtual nodes in large-scale networks. Therefore, the filtering of homogeneous attributes becomes essential. This observation further supports the rationale behind filtering homogeneous attributes, as it helps to maintain the clarity and accuracy of community detection in complex networks.

In Figure 8, we can observe the visualization of partitions in the PolBooks dataset, which represents a small-scale network with a real community of 3.



a) Comparison of Louvain algorithm result



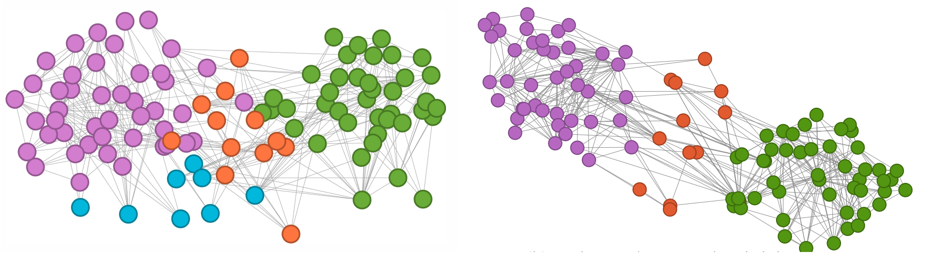
b) Comparison of FastGreedy algorithm result

Figure 7. Single attribute comparison

When using only topological information, the network is divided into four communities. However, when attribute and topological information are combined, the result is updated to three communities, which aligns more closely with the actual community structure.

5 CONCLUSION

The multidimensional feature fusion community detection model (Louvain-FTAS) proposed in this paper combines network topology and node attributes to improve



a) Original network division b) Enhanced network division

Figure 8. Comparison of Polbooks dataset community partitioning results

the accuracy of community division. The algorithm addresses the challenges posed by heterogeneous attributes and incorporates attribute filtering to enhance the impact of relevant attributes on the network structure. By considering the complementary nature of attributes and topology, Louvain-FTAS achieves better community detection results.

Future research directions include exploring improved methods for measuring the convergence of structural and attribute information and identifying additional potential communities to further enhance the accuracy of community division. Extending the Louvain-FTAS algorithm to handle dynamic networks could be another interesting area of research. By incorporating temporal aspects, such as evolving attribute values and changing network structure over time, the algorithm could adapt to communities that change over different time intervals. This could involve considering attribute and structural information in a time-dependent manner and developing mechanisms to identify and track communities as they evolve.

Acknowledgement

This work was supported in part by the National Key Research and Development Program of China under Grant No. 2022YFB4501704, in part by the National Natural Science Foundation of China under Grants No. 62302308, 62372300, 61702333 and U2142206, and in part by the Shanghai Sailing Program under Grant No. 21YF1432900.

REFERENCES

[1] Fortunato, S.: Community Detection in Graphs. Physics Reports, Vol. 486, 2010, No. 3-5, pp. 75–174, doi: 10.1016/j.physrep.2009.11.002.

[2] Newman, M. E. J.: Modularity and Community Structure in Networks. Proceedings of the National Academy of Sciences, Vol. 103, 2006, No. 23, pp. 8577–8582, doi: 10.1073/pnas.0601602103.

[3] Blondel, V. D.—Guillaume, J. L.—Lambiotte, R.—Lefebvre, E.: Fast Unfolding of Communities in Large Networks. Journal of Statistical Mechanics: Theory and Experiment, Vol. 2008, 2008, No. 10, Art. No. P10008, doi: 10.1088/1742-5468/2008/10/P10008.

[4] Zhao, Q.—Wang, C.—Jiang, C.: HSim: A Novel Method on Similarity Computation by Hybrid Measure. 2015 6th International Conference on Information and Communication Systems (ICICS), IEEE, 2015, pp. 160–165, doi: 10.1109/IACS.2015.7103220.

[5] Zhao, Q.—He, Y.—Jiang, C.—Wang, P.—Qi, M.—Li, M.: Integration of Link and Semantic Relations for Information Recommendation. Computing and Informatics, Vol. 35, 2016, No. 1, pp. 30–54, `https://www.cai.sk/ojs/index.php/cai/article/view/2600`.

[6] Zhao, Q.—Wang, C.—Wang, P.—Zhou, M. C.—Jiang, C.: A Novel Method on Information Recommendation via Hybrid Similarity. IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 48, 2016, No. 3, pp. 448–459, doi: 10.1109/TSMC.2016.2633573.

[7] Zhao, Q.—Huang, J.—Liu, G.—Miao, Y.—Wang, P.: A Multiinterest and Social Interest-Field Framework for Financial Security. IEEE Transactions on Computational Social Systems, 2023, doi: 10.1109/TCSS.2023.3252611.

[8] Fu, N.—Zhao, Q.—Miao, Y.—Zhang, B.—Wang, D.: Representation Learning Method of Graph Convolutional Network Based on Structure Enhancement. Computing and Informatics, Vol. 41, 2022, No. 6, pp. 1563–1588, doi: 10.31577/cai_2022_6_1563.

[9] Pan, J.—Li, H.—Teng, J.—Zhao, Q.—Li, M.: Dynamic Network Representation Learning Method Based on Improved GRU Network. Computing and Informatics, Vol. 41, 2022, No. 6, pp. 1491–1509, doi: 10.31577/cai_2022_6_1491.

[10] Fan, S.—Liu, G.—Li, J.: A Heterogeneous Graph Neural Network with Attribute Enhancement and Structure-Aware Attention. IEEE Transactions on Computational Social Systems, 2023, doi: 10.1109/TCSS.2023.3239034.

[11] Lou, X.—Liu, G.—Li, J.: ASIAM-HGNN: Automatic Selection and Interpretable Aggregation of Meta-Path Instances for Heterogeneous Graph Neural Network. Computing and Informatics, Vol. 42, 2023, No. 2, pp. 257–279, doi: 10.31577/cai_2023_2_257.

[12] Simon, H. A.: The Sciences of the Artificial. M.I.T. Press, 1969.

[13] Watts, D. J.—Strogatz, S. H.: Collective Dynamics of 'Small-World' Networks. Nature, Vol. 393, 1998, No. 6684, pp. 440–442, doi: 10.1038/30918.

[14] Leskovec, J.—Lang, K. J.—Mahoney, M.: Empirical Comparison of Algorithms for Network Community Detection. Proceedings of the 19th International Conference on World Wide Web (WWW '10), 2010, pp. 631–640, doi: 10.1145/1772690.1772755.

[15] Du, N.—Wu, B.—Pei, X.—Wang, B.—Xu, L.: Community Detection in Large-Scale Social Networks. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis (WebKDD/SNA-KDD '07), 2007, pp. 16–25, doi: 10.1145/1348549.1348552.

[16] Duch, J.—Arenas, A.: Community Detection in Complex Networks Using Ex-

tremal Optimization. Physical Review E, Vol. 72, 2005, No. 2, Art. No. 027104, doi: 10.1103/PhysRevE.72.027104.

[17] Zhou, Z.—Amini, A. A.: Analysis of Spectral Clustering Algorithms for Community Detection: The General Bipartite Setting. The Journal of Machine Learning Research, Vol. 20, 2019, No. 1, pp. 1774–1820, doi: 10.48550/arXiv.1803.04547.

[18] Gui, C.—Zhang, R.—Hu, R.—Huang, G.—Wei, J.: Overlapping Communities Detection Based on Spectral Analysis of Line Graphs. Physica A: Statistical Mechanics and Its Applications, Vol. 498, 2018, pp. 50–65, doi: 10.1016/j.physa.2018.01.025.

[19] Li, H. J.—Wang, Y.—Wu, L. Y.—Liu, Z. P.—Chen, L.—Zhang, X. S.: Community Structure Detection Based on Potts Model and Network's Spectral Characterization. Europhysics Letters (EPL), Vol. 97, 2012, No. 4, Art. No. 48005, doi: 10.1209/0295-5075/97/48005.

[20] Newman, M. E. J.: Spectral Methods for Community Detection and Graph Partitioning. Physical Review E, Vol. 88, 2013, No. 4, Art. No. 042822, doi: 10.1103/PhysRevE.88.042822.

[21] Li, Y.—He, K.—Kloster, K.—Bindel, D.—Hopcroft, J.: Local Spectral Clustering for Overlapping Community Detection. ACM Transactions on Knowledge Discovery from Data (TKDD), Vol. 12, 2018, No. 2, Art. No. 17, doi: 10.1145/3106370.

[22] Chen, M.—Kuzmin, K.—Szymanski, B. K.: Community Detection via Maximization of Modularity and Its Variants. IEEE Transactions on Computational Social Systems, Vol. 1, 2014, No. 1, pp. 46–65, doi: 10.1109/TCSS.2014.2307458.

[23] Yazdanparast, S.—Havens, T. C.—Jamalabdollahi, M.: Soft Overlapping Community Detection in Large-Scale Networks via Fast Fuzzy Modularity Maximization. IEEE Transactions on Fuzzy Systems, Vol. 29, 2020, No. 6, pp. 1533–1543, doi: 10.1109/TFUZZ.2020.2980502.

[24] Mohammadi, M.—Fazlali, M.—Hosseinzadeh, M.: Accelerating Louvain Community Detection Algorithm on Graphic Processing Unit. The Journal of Supercomputing, Vol. 77, 2021, No. 6, pp. 6056–6077, doi: 10.1007/s11227-020-03510-9.

[25] Que, X.—Checconi, F.—Petrini, F.—Gunnels, J. A.: Scalable Community Detection with the Louvain Algorithm. 2015 IEEE International Parallel and Distributed Processing Symposium, 2015, pp. 28–37, doi: 10.1109/IPDPS.2015.59.

[26] Fazlali, M.—Moradi, E.—Malazi, H. T.: Adaptive Community Detection Algorithm on Graphic Processing Platform. Microprocessors and Microsystems, Vol. 54, 2017, pp. 26–34, doi: 10.1016/j.micpro.2017.08.002.

[27] Mu, C. H.—Xie, J.—Liu, Y.—Chen, F.—Liu, Y.—Jiao, L. C.: Memetic Algorithm with Simulated Annealing Strategy and Tightness Greedy Optimization for Community Detection in Networks. Applied Soft Computing, Vol. 34, 2015, pp. 485–501, doi: 10.1016/j.asoc.2015.05.034.

[28] Ghasemian, A.—Hosseinmardi, H.—Clauset, A.: Evaluating Overfit and Underfit in Models of Network Community Structure. IEEE Transactions on Knowledge and Data Engineering, Vol. 32, 2019, No. 9, pp. 1722–1735, doi: 10.1109/TKDE.2019.2911585.

[29] Zhang, X. K.—Ren, J.—Song, C.—Jia, J.—Zhang, Q.: Label Propaga-

tion Algorithm for Community Detection Based on Node Importance and Label Influence. Physics Letters A, Vol. 381, 2017, No. 33, pp. 2691–2698, doi: 10.1016/j.physleta.2017.06.018.

[30] Xie, J.—Szymanski, B. K.: Towards Linear Time Overlapping Community Detection in Social Networks. In: Tan, P. N., Chawla, S., Ho, C. K., Bailey, J. (Eds.): Advances in Knowledge Discovery and Data Mining (PAKDD 2012). Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 7302, 2012, pp. 25–36, doi: 10.1007/978-3-642-30220-6_3.

[31] Gregory, S.: Finding Overlapping Communities in Networks by Label Propagation. New Journal of Physics, Vol. 12, 2010, No. 10, Art. No. 103018, doi: 10.1088/1367-2630/12/10/103018.

[32] Rosvall, M.—Bergstrom, C. T.: Maps of Random Walks on Complex Networks Reveal Community Structure. Proceedings of the National Academy of Sciences, Vol. 105, 2008, No. 4, pp. 1118–1123, doi: 10.1073/pnas.0706851105.

[33] Hartigan, J. A.—Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 28, 1979, No. 1, pp. 100–108, doi: 10.2307/2346830.

[34] Kaufmann, L.—von Aster, M.: The Diagnosis and Management of Dyscalculia. Deutsches Ärzteblatt International, Vol. 109, 2012, No. 45, pp. 767–778, doi: 10.3238/arztebl.2012.0767.

[35] Li, Z.—Liu, J.—Wu, K.: A Multiobjective Evolutionary Algorithm Based on Structural and Attribute Similarities for Community Detection in Attributed Networks. IEEE Transactions on Cybernetics, Vol. 48, 2017, No. 7, pp. 1963–1976, doi: 10.1109/TCYB.2017.2720180.

[36] Cheng, H.—Zhou, Y.—Huang, X.—Yu, J. X.: Clustering Large Attributed Information Networks: An Efficient Incremental Computing Approach. Data Mining and Knowledge Discovery, Vol. 25, 2012, No. 3, pp. 450–477, doi: 10.1007/s10618-012-0263-0.

[37] Berahmand, K.—Haghani, S.—Rostami, M.—Li, Y.: A New Attributed Graph Clustering by Using Label Propagation in Complex Networks. Journal of King Saud University – Computer and Information Sciences, Vol. 34, 2022, No. 5, pp. 1869–1883, doi: 10.1016/j.jksuci.2020.08.013.

[38] Ma, H.—Liu, Z.—Zhang, X.—Zhang, L.—Jiang, H.: Balancing Topology Structure and Node Attribute in Evolutionary Multi-Objective Community Detection for Attributed Networks. Knowledge-Based Systems, Vol. 227, 2021, Art. No. 107169, doi: 10.1016/j.knosys.2021.107169.

[39] Rossi, R.—Ahmed, N.: The Network Data Repository with Interactive Graph Analytics and Visualization. Vol. 29, 2015, No. 1, pp. 4292–4293, doi: 10.1609/aaai.v29i1.9277.

[40] McCallum, A. K.—Nigam, K.—Rennie, J.—Seymore, K.: Automating the Construction of Internet Portals with Machine Learning. Information Retrieval, Vol. 3, 2000, pp. 127–163, doi: 10.1023/A:1009953814988.

[41] Lancichinetti, A.—Fortunato, S.—Radicchi, F.: Benchmark Graphs for Testing Community Detection Algorithms. Physical Review E, Vol. 78, 2008, No. 4,

Art. No. 046110, doi: 10.1103/PhysRevE.78.046110.

[42] Estévez, P. A.—Tesmer, M.—Perez, C. A.—Zurada, J. M.: Normalized Mutual Information Feature Selection. IEEE Transactions on Neural Networks, Vol. 20, 2009, No. 2, pp. 189–201, doi: 10.1109/TNN.2008.2005601.

[43] Pil, F. K.—Cohen, S. K.: Modularity: Implications for Imitation, Innovation, and Sustained Advantage. Academy of Management Review, Vol. 31, 2006, No. 4, pp. 995–1011, doi: 10.5465/amr.2006.22528166.

[44] Fujino, A.—Isozaki, H.—Suzuki, J.: Multi-Label Text Categorization with Model Combination Based on F1-Score Maximization. Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II, 2008, `https://aclanthology.org/I08-2116.pdf`.

[45] Traud, A. L.—Mucha, P. J.—Porter, M. A.: Social Structure of Facebook Networks. Physica A: Statistical Mechanics and Its Applications, Vol. 391, 2012, No. 16, pp. 4165–4180, doi: 10.1016/j.physa.2011.12.021.

Qin Zhao received his Ph.D. degree from the Department of Computer Science and Technology, Tongji University, in 2016. Currently, he is Associate Professor with the Department of Computer Science and Technology, Shanghai Normal University. He also serves as the Deputy Director of the Shanghai Engineering Research Center of Intelligent Education and Big Data. He is recognized as a Senior Member of both the IEEE and the China Computer Federation (CCF). He has published more than 30 papers in premier international journals and conferences, including IEEE Transactions on Systems, Man, and Cybernetics: Systems, IEEE Transactions on Computational Social Systems, etc. His research interests include but are not limited to new-generation artificial intelligence technologies, social network analysis, data mining, and intelligence in scientific computing.

Yaru Miao received her B.Eng. degree from the Department of Computer, Shanghai Normal University in 2021. She is currently pursuing her M.Sc. degree in the Department of Computer Science and Technology at Shanghai Normal University, China. Her research interests include community detection, natural language processing and sentiment analysis.

Jie Lian is Associate Professor in the Department of Computer Science and Technology at Shanghai Normal University where she has been a faculty member since 2017, and she obtained the "Sailing" Talent Program of China in 2019. She completed her doctor degree at the Towson University in 2017. Her research interests lie in the area of spatio-temporal data mining, deep learning and big data, ranging from theory to design to implementation.

Xiaoya Li received her Bachelor degree from the Department of Information and Computing Science, Qingdao Agricultural University, Qingdao, China, in 2019, and received her Master degree from the Department of Computer Science and Technology at Shanghai Normal University, Shanghai, China, in 2022. Currently, she is an Engineer with China United Network Communications Group Co., Ltd., Shanghai, China. Her research interests mainly include social network, complex network and data mining.

Dongdong An received her B.Sc. degree and her Ph.D. degree in software engineering from the East China Normal University, Shanghai, China, in 2013 and 2020, respectively. She is currently a Lecturer with the Department of Computer Science and Technology, Shanghai Normal University, Shanghai, China. Her current research interests include model-driven architecture, machine learning, formal methods and statistical model checking techniques.