Research Article

Multilayer Social Network Overlapping Community Detection Algorithm Based on Trust Relationship

Junjie Jia,¹ Pengtao Liu¹, ,¹ Xiaojin Du,² and Yuchao Zhang¹

¹*School of Computer Science and Engineering, Northwest Normal University, China*

²*School of Numbers and Statistics, Northwest Normal University, China*

Correspondence should be addressed to Pengtao Liu; 2019221796@nwnu.edu.cn

Received 26 April 2021; Accepted 29 October 2021; Published 10 November 2021

Academic Editor: Hasan Khattak

Copyright © 2021 Junjie Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problem of the lack of user social attribute characteristics in the process of dividing overlapping communities in multilayer social networks, in this paper, we propose a multilayer social network overlapping community detection algorithm based on trust relationship. By combining structural trust and social attribute trust, we transform a complex multilayer social network into a single-layer trust network. We obtain the community structure according to the community discovery algorithm based on trust value and merge communities with higher overlap. The experimental comparison and analysis are carried out on the synthetic network and the real network, respectively. The experimental results show that the proposed algorithm has higher harmonic mean and modularity than other algorithms of the same type.

1. Introduction

As a representation model of user interaction, a social network is often used to analyze the complex social structure between people in real life. A social network is composed of a set of nodes and an edge set, where nodes represent users and edges represent interactions between users [1]. According to different types of interactions, different social network structures are formed. In recent years, a large number of social network applications have gradually catered to the needs of people in all aspects of daily life and have developed towards more humane, daily and intelligent aspects, such as Weibo, Twitter, and Facebook. According to research, more than 80% of users use at least one social network application [2]. Users use social network applications as a platform to post their emotions, attitudes, and values to their circle of friends. They can forward, comment, and like each other. As a result, various complex social relationship data are generated. Since different social platforms correspond to different social relationships, it is obviously not advisable to adopt a traditional single-layer social network to establish a user interaction model. Therefore, multilayer social networks are proposed to represent the

complex network of relationships between people in the real world. In a multilayer social network, the nodes are users, and each layer represents a type of user's social relationship [3–5].

Community detection, as an important tool for analyzing and understanding multilayered social networks, is manifested in close connections between nodes in the same community and sparse connections between different communities. It is widely used in various aspects of real life, such as public opinion detection [6], recommendation systems [7], and epidemic spread [8], and has become one of the key technologies for understanding and analyzing the structure of multilayer social networks. As shown in Figure 1(b), the research network consists of three layers: coauthoring layer, Twitter, and Facebook. At the coauthoring layer, if the authors have copublished papers, the nodes are connected to each other. In the Twitter and Facebook layers, the connection between nodes corresponds to the interaction between users. As shown in Figure 1(c), its community is a group of researchers with the same research direction. Through community discovery, users in the same community communicate with each other, answer questions, and recommend papers to better serve scientific research [9].

FIGURE 1: Schematic diagram of multilayer social network. (a) Multilayer social network (the research network), (b) social relations at every level, and (c) community structure.

For a long time in the past, many scholars mainly focused on the structural relationship of nodes in the network, that is, the association relationship between users, to detect nonoverlapping communities [10]. However, related studies have shown [11] that communities in the real world are not separated. For example, users have multiple roles in social networks and can be in multiple communities at the same time; that is, communities can overlap. Faced with such a complex multilayer social network, detecting overlapping communities has become a hot topic in the current research.

Currently, relevant researchers apply most overlapping community detection algorithms to single-layer social networks: for example, local optimization community detection algorithm [12, 13], improved label propagation community detection algorithm [14], and parallel multiobjective optimization community detection algorithm [15]. In addition, it also includes a nonnegative matrix factorization community detection algorithm [16]. Experimental results show that these algorithms have achieved good results in single-layer networks. However, a multilayer social network consists of multiple layers, and each layer corresponds to a social relationship between users. It can be seen that these algorithms cannot be directly adapted to multilayer networks. For this reason, many researchers adopt the most direct strategy to extend the single-layer overlapping community discovery method to multiple layers [17] or use a multiobjective optimization method [18] to divide communities. However, these methods show lower performance when the network is large.

In addition, some researchers are inspired by user interaction behavior in social networks to simulate social interaction by constructing trust between users and divide communities based on the trust relationship between users. Users with a stronger trust relationship establish similar interests and hobbies through mutual influence and finally form a relatively stable group [19]. Compared with other community division methods, the trust-based community division method is simple in principle and has strong practical significance. At the same time, the experimental results show that the trust-based community discovery method

greatly improves the accuracy of community detection [20, 21]. However, for multilayer social networks, these methods can only obtain nonoverlapping communities of the network, and in the process of building trust, they rely excessively on the structural characteristics of the network and ignore the characteristics of social attributes. Structure and social attributes are the basic characteristics of the network. The structural characteristics describe the user's social scope, while the social attributes are the inherent characteristics of the user and correspond to the user's social style.

Based on this, in this paper, we propose a multilayer social network overlapping community detection algorithm based on trust relationship (MOCBT). We integrate the characteristics of social attributes and transform the complex social information in the multilayer social network into a single-layer network based on trust value. We divide the community by using the local information of the network, thereby improving the accuracy of community detection.

The main contributions of this article are summarized as follows:

- (1) By integrating the characteristics of social attributes, we propose an evaluation model for evaluating the strength of trust between users
- (2) Based on the trust relationship between nodes, we propose a community detection algorithm based on trust value
- (3) We propose a multilayer network overlapping community detection algorithm based on trust relationship (MOCBT), which provides a general framework for multilayer network community discovery based on trust relationship. Experimental results show that the MOCBT algorithm is better than other community detection algorithms of the same type as a whole

The structure of the rest of the article is as follows: Section 2 summarizes the research status and the existing problems of multilayer social network community detection. Section 3 introduces the relevant theories of community

detection. Section 4 discusses the proposed algorithm in detail. Section 5 verifies the feasibility of the proposed algorithm through experimental analysis. Section 6 summarizes the full text and indicates the work to be done in the next step.

2. Related Work

Recalling the traditional community detection methods in multilayer social networks can be divided into two main strategies: single-layer community expansion and integration methods and other methods. The first strategy extends the single-layer network community detection algorithm to multilayer networks. Rodriguez [22] proposed a multilayer network community detection algorithm framework based on path algebra. They use tensors to represent multilayer networks and slice bidirectionally to generate a single relational path matrix. Chen et al. [23] fused the multilayer network into a weighted single-layer network by calculating the similarity between nodes and used an improved label propagation algorithm to obtain the community. However, these methods will cause the lack of network information in the process of integration and expansion, resulting in inaccurate community detection [24]. Besides, Ma et al. [24] proposed a semisupervised nonnegative matrix factorization algorithm for multilayer networks, by obtaining denser communities in the network in advance. Gao et al. [25] proposed a community detection model suitable for complex networks by combining a competitive learning mechanism and random deterministic roles. However, the communities obtained by the above algorithm are nonoverlapping. To this end, Tehrani and Magnani [17] used the first strategy to redefine the definition of faction filtering algorithms and proposed an extensible faction filtering overlapping community discovery algorithm. Based on the basic theory that users in the same community have dense connections and users in different communities have sparse connections, Shahmoradi et al. [18] obtained network overlapping communities by using multiobjective optimization. However, these methods only use the structural characteristics of the network in the process of community detection, and the divided communities are not ideal when the data scale is large.

As we all know, social networks are composed of users and the interaction between users, but the interaction between users is ultimately a trust relationship between each other. Based on this, Chen et al. [26] established a trust model by integrating direct trust and indirect trust between users. They proposed a single-layer social network-oriented nonoverlapping community discovery algorithm by combining edge fitness and community fitness. Ding et al. [27] established a trust evaluation model between users by integrating structural similarity and neighbor similarity and then obtained overlapping communities of single-layer networks by using coarse-grained K-medoids. In contrast, Li et al. [28] established a trust model between users by analyzing the homogeneity between nodes and the shortest path and proposed a multilayer network local community discovery algorithm based on trust relationships. However, in the process of establishing a trust model, these algorithms essentially only use the structural features of the network,

ignoring the social attributes of the network. Therefore, for multilayer social networks, it is necessary to design an overlapping community discovery algorithm based on trust relationships and integrating social attribute characteristics.

3. Related Theories

3.1. Multilayer Social Network. The user information we collect and extract is usually expressed as unstructured. To facilitate analysis and understanding, we use the ontology model [29] to structure the user's social information. The ontology model can be used as a standard modeling tool, while also ensuring interoperability between data.

After the data is structured, we define a multilayer social network as an undirected graph $G = (V, E, P, L)$ with attribute information.

$L = \{L_i\}_{i=1,\dots,s}$ represents the number of network layers. $V = \{v_i\}_{i=1,\dots,n}$ represents the set of n nodes in the graph. $E = \{(v_i, v_j, \omega) | v_i, v_j \in V\}$ s.t. $|E| = m$ represents the set of m edges in the graph. The edges represent the social relationships between nodes, and the social relationships at different levels are different. The edge weight ω represents the strength of the relationship between the nodes. $P = \{p_1, \dots, p_k\}$ represents the set of attributes of nodes in the network.

3.2. Community. We collected descriptions of communities from previous literature [7–18], which can be summarized as follows:

- (i) Nature 1: the nodes within the community are closely connected, and the nodes within the community are sparsely connected with nodes outside the community.

It can be seen that the definition of community in the past literature focuses on the structural characteristics of the network. However, this is not the case. Users with close structural relationships but different social attributes may not necessarily join the same community. This is mainly determined by the user's social status and value; that is, birds of a feather flock together. Structural characteristics and social attribute characteristics are the two basic attributes of the network. Social attribute characteristics include two aspects of personal status and value. Personal status refers to demographic characteristics, such as race, gender, age, education level, and occupation. Personal value refers to thoughts, attitudes, lifestyles, etc.

- (ii) Nature 2: structural features are only the prerequisites for the formation of communities, and what really matters is the characteristics of social attributes. In other words, nodes within the community are closely connected and have similar social attribute characteristics, while nodes with different social attribute characteristics and close connections may not be in the same community.

3.3. Degree of Community Overlap. When there are many common nodes between two communities, it means that they can be merged.

Definition 1 (degree of community overlap, β) [19]. β is used to judge whether the communities can be merged. When β is higher, it means that the communities can be merged.

$$\beta = \frac{|C_i \cap C_j|}{\min |C_i, C_j|}. \quad (1)$$

$|C_i \cap C_j|$ represents the number of nodes shared by communities C_i and C_j , and $\min |C_i, C_j|$ represents the minimum number of nodes owned by communities C_i and C_j .

3.4. Problem Description. Trust, as the core part of social interaction, is a concentrated expression of a person's mode of getting along in social networks, and it is also a bridge to maintain and establish interpersonal relationships. By building trust, people maintain a long-term and stable social connection. As time changes, users with stronger trust relationships gather together and share information with each other. Through mutual influence, similar interests and hobbies are established, and finally a relatively stable group is formed [19]. It can be seen that the strength of the trust relationship between users is the main factor that affects the formation of the community structure. Based on this relationship, in this article, we transform the complex information between users in a multilayer network into a trust relationship. After that, we proposed an overlapping community discovery algorithm based on trust relationship, which provides a new idea for the division of overlapping communities in multilayer social networks.

4. Algorithm Description

Aiming at the shortcomings of traditional community detection methods, in this article, we propose a multilayer network overlapping community detection algorithm based on trust relationships (MOCBT). By combining the social attribute characteristics of the network, we transform the complex information between nodes in the multilayer network into a single-layer network based on trust value. Based on the community detection algorithm of trust value, we obtain overlapping community structure.

As shown in Figure 2, our proposed community detection algorithm consists of three steps.

- (Step 1) We combine node structure trust and social attribute trust to obtain a comprehensive trust value.
- (Step 2) We convert the trust value between nodes into trust distance and obtain the core node set according to the trust factor and the minimum trust distance and then expand the nodes that meet the conditions to the community where the core node is located.
- (Step 3) We merge communities with a high degree of overlap.

4.1. Establish Trust Value. In this section, we will describe in detail how to convert the complex information of nodes in a multilayer social network into a single-layer trust value. Since the structure and social attributes of nodes are the two basic characteristics of the network and trust value can be used to measure the intensity of interaction between user nodes in the network and the possibility of being in the same community, the structure and social attributes of nodes have become the main factors affecting trust.

4.1.1. The Impact of Structure on Trust

Definition 2 (trust value of the single-layer shortest path, $n_{(v_i, v_j)}$). The longer the shortest path between user nodes v_i and v_j , the smaller the value of $n_{(v_i, v_j)}$.

In the field of communication, the longer the path the information travels, the weaker the accuracy and completeness of the information. Similarly, in one of the layers of a multilayer social network, when two user nodes need to establish an interactive relationship, the more nodes they pass, the lower the trust between them. For example, if users A and B have no direct interaction relationship, but both have a direct connection with user C; the trust between users A and C or B and C is stronger than the trust between users A and B.

Definition 3 (trust value of the number of single-level paths, $D_{(v_i, v_j)}$). The more paths between user nodes v_i and v_j , the greater the value of $D_{(v_i, v_j)}$.

For example, users A and B are friends and have many friends in common, indicating that they have many common interests and show closeness.

Definition 4 (multilayer structure trust value, $D_trust_{(v_i, v_j)}$). Given a multilayer social network $G = (L, V, E, P)$, combined with Definitions 2 and 3, the multilayer structure trust value between nodes can be expressed as

$$D_trust_{(v_i, v_j)} = \sum_L \omega_{ij} \times D_{(v_i, v_j)} \times n_{(v_i, v_j)}, \quad (2)$$

where L represents the number of network layers and ω_{ij} represents the interaction weight, that is, the frequency of interaction between users. The larger the $D_trust_{(v_i, v_j)}$, the higher the user's trust.

4.1.2. The Impact of Social Attributes on Trust. Structural features mainly exist in the form of edges in multilayer social networks, while social attribute features are attached to the network and are inherent characteristics of users. Research shows [30] that the higher the similarity of social attributes between nodes in social networks, the stronger the homogeneity between users, and homogeneity can be described as the degree of trust between users [27].

FIGURE 2: Flow chart of MOCBT algorithm.

Definition 5 (social attribute trust value, $S_trust_{(v_i, v_j)}$). Given a multilayer social network $G = (L, V, E, P)$, the more social attributes shared between users, the stronger the social attribute trust value.

$$S_trust_{(v_i, v_j)} = \frac{1}{|P|} \sum_{p_1}^{p_k} \sum_{L_1}^{L_s} S_p(v_i, v_j) \lambda_L^p, \quad (3)$$

where $|P|$ represents the number of social attributes, λ_L^p represents the weight of the attribute p in the L layer, and $S_p(v_i, v_j)$ represents the trust relationship between nodes v_i and v_j for the attribute p . According to different types of attributes, the value of $S_p(v_i, v_j)$ can be divided into two situations. When the attribute p is a discrete attribute, such as occupation and diploma, if users v_i and v_j have the same attribute value, the value of $S_p(v_i, v_j)$ is 1; otherwise, it is 0. When the attribute p is the attribute of the text, such as search records and chat records, inspired by short text mining, we calculate the value of $S_p(v_i, v_j)$ based on the frequency of keywords in the text. Assuming that $\{k, k = 1, \dots, k\}$ is the keyword of the text attribute p , then $S_p(v_i, v_j) = \sum_{k=1}^k I_u(k) \times I_v(k) \times (1/(\lg D_{in}(k)))$. Where $D_{in}(k)$ represents the number of nodes containing the keyword k and $I_u(k)$ represents whether the attribute of node u contains k , if it does, $I_u(k) = 1$; otherwise, $I_u(k) = 0$.

4.1.3. Comprehensive Trust Value

Definition 6 (comprehensive trust value, $trust_{(v_i, v_j)}$). Given $G = (L, V, E, P)$, $D_trust_{(v_i, v_j)}$, and $S_trust_{(v_i, v_j)}$, the comprehensive trust value of user v_i to user v_j can be expressed as

$$trust_{(v_i, v_j)} = \frac{D_trust_{(v_i, v_j)}}{D_trust_{(v_i)}} \times \tau_{(v_i, v_j)}, \quad (4)$$

where $trust_{(v_i, v_j)}$ represents the one-way trust value of user v_i to v_j . When the value of $trust_{(v_i, v_j)}$ is closer to 1, the trust relationship between user v_i and v_j is stronger. $D_trust_{(v_i)}$ represents the sum of trust between user v_i and its neighbor node. $\tau_{(v_i, v_j)}$ represents an indicator function. When $S_trust_{(v_i, v_j)}$ is less than or equal to 0, $\tau_{(v_i, v_j)} = 0$; on the contrary, $\tau_{(v_i, v_j)} = S_trust_{(v_i, v_j)}$.

Affected by education, experience, and life background, the trust between users is not equivalent; that is, user v_i 's trust for v_j is not equivalent to user v_j 's trust for v_i . However, in the process of interaction, the establishment of a trust relationship between users often depends on the party with a lower trust value. For this reason, we choose $\min(trust_{(v_i, v_j)}, trust_{(v_j, v_i)})$ as the ultimate trust between users.

4.2. Discover the Community. In the previous section, we transformed the complex multilayer social network into a single-layer network based on trust value, which greatly

simplified the scale of the problem. In this section, we abstract the transformed single-layer trust network into a data domain based on the data field theory [27], where the data points represent users and the trust value is expressed as the distance between the data points, that is, the interaction strength of the nodes. In view of the network's small-world effect, scale-free nature, and the aggregation characteristics of the network node structure, the core nodes in the community often have strong local interactions. That is, the core node has strong interaction with neighbor nodes, but the interaction between core nodes in different communities is weak. We find the core node by calculating the trust factor and the minimum trust value, obtain the community where the core node is located according to the core value change rate, and propose a community discovery algorithm based on trust value (CAOT).

Since the trust value between nodes decreases with the increase of the distance, we use the Gaussian function to objectively measure the changing trend of the trust value of the nodes with the distance.

Definition 7 (trust distance, $d_{(v_i, v_j)}$). Given a multilayer social network $G = (L, V, E, P)$, the trust distance of node $v_i \in V$ to node $v_j \in V$ is expressed as

$$d_{(v_i, v_j)} = \exp \left(\frac{\text{trust}(v_i, v_j)}{2\sigma^2} \right). \quad (5)$$

Definition 8 (trust factor, ρ_{v_i}). The trust factor is described as the sum of the trust value of node v_i and its neighbor node $\Gamma(v_i)$. The trust factor ρ_{v_i} of node v_i is expressed as

$$\rho_{v_i} = \sum_{v_j \in \Gamma(v_i)} \text{trust}(v_i, v_j). \quad (6)$$

Definition 9 (minimum trust distance, δ_{v_i}). Due to the large trust distance between core nodes in different communities, δ_{v_i} is described as the minimum trust distance from node v_i to node v_j with a higher trust factor. According to the value of the trust factor, δ_{v_i} is divided into two types of values, defined as

$$\delta_{v_i} = \begin{cases} \max_{v_j \in V} (d_{(v_i, v_j)}), & \text{if } \rho_{v_i} = \max (\rho_{v_i}), \\ \min_{v_j: \rho_{v_j} > \rho_{v_i}} (d_{(v_i, v_j)}), & \text{otherwise.} \end{cases} \quad (7)$$

When the trust factor ρ_{v_i} of the node v_i is the largest, δ_{v_i} is the maximum value of $d_{(v_i, v_j)}$ of node v_i and other nodes v_j . On the contrary, when the trust factor of node v_i is not the maximum, δ_{v_i} is the minimum value of $d_{(v_i, v_j)}$ of the node v_i and other nodes v_j , and the trust factor of node v_j is greater than the trust factor of node v_i .

Since core nodes have a higher trust factor and a larger minimum trust distance than noncore nodes, we sort the core value $\omega_{v_i} = \rho_{v_i} \times \delta_{v_i}$ of each node in an ascending order. There is a clear downward trend from the core value of the core node to the core value of the noncore node. Therefore, we use the trend of slope change to automatically select the core node and define the trend as $\text{rate}_{v_i} = (i-1)((\omega_{v_{i-1}} - \omega_{v_i})/(\omega_{v_i} - \omega_{v_{i+1}}))$, where $(i-1)$ is the weight value to prevent the rapid increase of rate_{v_i} , so as to improve the accuracy of core node selection. Finally, we denote the obtained core node set as $H = \{h_i\}_{i=1, \dots, z}$.

Definition 10 (the community where the core node is located, C_{h_i}). Given the core node $h_i \in H$ and the noncore node $v_j \in (V - H)$, if $(d_{(v_j, h_i)} < \epsilon)$, then $v_j \in C_{h_i}$.

The threshold ϵ determines whether node v_j can join community C_{h_i} . In order to determine the value of ϵ , we conducted a detailed experimental comparative analysis.

The CAOT algorithm first uses the Gaussian function to objectively measure the trend of the trust value $\text{trust}(v_i, v_j)$ between nodes with distance. By calculating the trust factor ρ_{v_i} and trust distance δ_{v_i} of each node, the algorithm automatically obtains the network core node set H . After that, the algorithm initializes the community set $\{C_h\}$ where each core node is located and adds noncore nodes v_j that meet the conditions to the current community. Finally, the community set where each core node is located is obtained, and the pseudocode of the algorithm is shown in Algorithm 1.

4.3. Community Merger. When there are more public nodes in any two communities obtained by the above steps, the communities may have excessive overlap. Therefore, it is necessary for us to further process the obtained communities to reduce redundant communities and improve the quality of community discovery. In this paper, we use formula (1) to calculate β the degree of overlap between two communities to determine whether the communities can be integrated.

4.4. Algorithm Description. Aiming at the problem of the lack of social attribute characteristics when constructing user trust relationships in multilayer social networks, we integrate social attribute characteristics to propose a new way of representing user trust values. Combining the community discovery algorithm CAOT based on trust value, we propose a multilayer social network overlapping community discovery algorithm based on trust relationships (MOCBT). The proposed algorithm supports two or more layers of social networks. The input of the algorithm is a multilayer social network G , and the output is a set of overlapping communities $\{C\}$. The algorithm first combines structure and attribute trust to obtain a comprehensive trust value $\text{trust}(v_i, v_j)$. After that, the community set $\{C_h\}$ where the core node is located is obtained according to the CAOT algorithm. Finally, the communities with a higher degree of overlap are merged according to the fusion threshold γ . The algorithm steps are in Algorithm 2.

Input: Multi-layer social network, $G = (V, E, P, L)$.

Output: The community collection where the core node is located, $\{C_h | C_h = (C_{h_i})_{i=1, \dots, z}\}$

```

1: Initialization,  $C_{h_i} = \emptyset$ 
2: for  $v_i$  in  $V$  :
3:   for  $v_j$  in  $V$  :
4:      $d_{(v_i, v_j)} = \exp(\text{trust}_{(v_i, v_j)} / 2\sigma^2)$ 
5:   end for
6: end for
7: for  $v_i$  in  $V$  :
8:    $\omega_{v_i} = \rho_{v_i} \times \delta_{v_i}$ 
9: end for
10: sort( $\omega_{v_i}$ )
11: According to  $\text{rate}_{v_i}$ , the core node set  $H$  is obtained
12:  $C_{h_i} = h_i$  // We initialize the community where each core node is located, that is, the core node represents an initial community.
13: for  $h_i$  in  $\{C_h\}$ :
14:   for  $v_j$  in  $(V - H)$ :
15:     if  $(d_{(v_j, h_i)} < \varepsilon)$ 
16:        $C_{h_i} \leftarrow v_i$  // Node  $v_i$  is added to community  $C_{h_i}$ .
17:     end if
18:   end for
19: end for
20: return  $\{C_h\}$ 

```

ALGORITHM 1: CAOT algorithm.

4.5. Algorithm Complexity Analysis. In order to verify the practicability of the algorithm, we analyzed the time and space complexity of Algorithms 1 and 2, respectively.

4.5.1. Time Complexity. In a multilayer network, n represents the number of nodes, and z represents the number of core nodes. In Section 4.1, we establish the trust value between user nodes by traversing each node in the network, and the time complexity is $O(n^2)$. In Section 4.2, we first convert the trust between nodes in the network into distance, and the time complexity is $O(n^2)$. Then, we get the core value of each node, and the time complexity is $O(n)$. We sort the core values to obtain core nodes, and the time complexity is $O(n \log_2 n)$. Finally, we obtain the community where each core node is located, and the time complexity is $O(z \times (n - z))$. In Section 4.3, the time complexity of merging communities is $O(z^2)$. Therefore, the total time complexity of Algorithm 1 is $O(n^2)$. The time complexity of Algorithm 2 is $O(n^2)$.

4.5.2. Space Complexity. Since we used the sorting method, the space complexity of Algorithms 1 and 2 is $O(1)$ in the best case, and the space complexity is $O(n \log_2 n)$ in the worst case.

5. Experimental Analysis

5.1. Data Set and Evaluation Indicators. We use the random model [31] to generate 8 multilayer synthetic social networks with different structure types according to the matching degree δ of the community and social attributes. The matching degree of the 8 synthetic networks is increased

from 0.3 to 1. The matching degree is described as the similarity between the social attribute and the community theme set. In other words, when $\delta = 0.3$, it indicates that the probability of the user's social attribute appearing in the community theme set is 0.3. The greater the probability, the higher the similarity. Each network is composed of 4 layers, and each layer adopts a different generation model. Each network is divided into two overlapping communities, including 100 nodes and 8 social attributes. As shown in Table 1, V represents the number of nodes and E represents the number of edges. P represents the number of social attribute features, and C represents the number of communities. L represents the number of network layers.

As shown in Table 2, we use four real multilayer social network data sets [28], such as Twitter data set, QQ Zone data set, Remote Sensing, and Bio GRID. The Twitter data set contains various topics such as entertainment, science, and sports and is divided into three levels: mention, repost, and reply. The QQ Zone data set collects the user's spatial data information, including the user's age and gender. The data set is divided into three levels: posting, comment, and like. The Remote Sensing data set is a network structure derived from satellite remote sensing images and contains five levels. Bio GRID contains human genetic data and various protein interaction data and consists of seven layers. The way the proteins interact in each layer is different.

- (1) Extended modularity, Q : since we considered the structure and attribute relationships of nodes, in order to measure the quality of overlapping communities, inspired by the literature [32], we rewrite the modularity as

Input: Multi-layer social network $G = (V, E, P, L)$.
Output: Overlapping community sets $\{C | C = (C_i)_{i=1, \dots, \varepsilon}\}$.

```

1: Initialization  $C_i = \emptyset$ 
2: for  $v_i$  in  $V$  :
3:   for  $v_j$  in  $V$  :
4:      $trust_{(v_i, v_j)}$ 
5:   end for
6: end for
7:  $trust_{(v_i, v_j)} = \min(trust_{(v_i, v_j)}, trust_{(v_j, v_i)})$ 
8: According to Algorithm 1, we obtain the community set  $\{C_h\}$  where the core node is located.
9:  $\{C\} = \{C_h\}$ 
10: for  $C_i$  in  $\{C\}$ :
11:   for  $C_j$  in  $\{C\}$ :
12:     if  $(|C_i \cap C_j| / \min |C_i, C_j| > \gamma)$ :
13:        $C_i = C_i \cap C_j$ 
14:     end if
15: end while
16: Return  $\{C\}$ 

```

ALGORITHM 2: MOCBT algorithm.

TABLE 1: Synthetic data set.

Data set	V	E	P	C	L	δ
G1	100	400	8	2	4	0.3
G2	100	400	8	2	4	0.4
G3	100	400	8	2	4	0.5
G4	100	400	8	2	4	0.6
G5	100	400	8	2	4	0.7
G6	100	400	8	2	4	0.8
G7	100	400	8	2	4	0.9
G8	100	400	8	2	4	1

TABLE 2: Real data set.

Data set	V	E	L
Twitter	438537	991854	3
QQ zone	562062	8055236	3
Bio GRID	38,936	342,599	7
Remote Sensing	642	4341	5

$$Q = \frac{1}{2m} \sum_{i=1}^{|C|} \sum_{p, q \in C_i} \left(A_{pq} - \frac{f_p f_q}{2m} \right) d_{(p, h_i)} d_{(q, h_i)}, \quad (8)$$

where $|C|$ represents the number of communities in the network, A_{pq} represents the weight between the node p and q , m represents the sum of the weights of all edges in the network, and $f_p f_q / 2m$ represents the possibility that the nodes p and q are connected in a random state. If the node p completely belongs to the community where the core node h_i is located; then $d_{(p, h_i)} = 1$. If node p does not completely belong to the community where core node h_i is located, the value range of $d_{(p, h_i)}$ satisfies $0 < d_{(p, h_i)} < 1$. Otherwise, $d_{(p, h_i)} = 0$.

- (2) Harmonic mean value, F -score [33]: F -score is a comprehensive measure of precision and recall. The closer F -score is to 1, the closer the community divided by the algorithm is to the real community

$$F\text{-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (9)$$

5.2. Experimental Results and Analysis. In order to verify the effectiveness of our proposed algorithm, the following five experiments are designed:

- (1) By changing the value of parameter ε , we observe the changes in the evaluation index in the synthetic data set to determine the appropriate value of ε
- (2) We selected 11 representative community discovery algorithms to compare with the proposed algorithm to verify the community overlap performance of the proposed algorithm
- (3) Through the analysis of the community detection results of the synthetic network and the real network, we verify the correctness of the proposed community nature
- (4) In the synthetic network, we compare and analyze the proposed algorithm with the contrast algorithm
- (5) In the real network, we compare and analyze the proposed algorithm with the contrast algorithm

Throughout the experiment, according to a previous literature, we set the parameter $\sigma = 1.866$. β reflects the degree of community overlap. The higher the value of β , the more nodes shared between communities. In order to reduce community redundancy, we set a fusion threshold γ . That is,

FIGURE 3: Experimental results of different ε values in the synthetic network.

when $\beta > \gamma$, it indicates that communities with many nodes in common can be merged into one community. The higher the value of γ , the smaller the number of final communities. In this article, we set $\gamma = 0.75$ [27].

5.2.1. The Value of Parameter ε . In this section, we discuss the influence of parameter ε on the proposed algorithm. In the 8 synthetic networks, we use three evaluation indicators: accuracy, recall, and harmonic mean, to observe the impact of the value of ε on the division of communities. The experimental results are shown in Figure 3. In each subgraph, as the value of ε increases, the original nodes in the community are deleted because they do not meet the conditions, which leads to an increase in the precision and a decrease in the recall. This satisfies the mutual restriction of precision and recall. The value of F -score increases first and then gradually decreases. Since we convert the trust value between users into a distance value, when the trust value between users is

greater, the distance is smaller, and the possibility of being in the same community is greater. When users have a small trust value, they have a large distance between them, and the possibility of being in the same community is relatively small. When the value of F -score reaches the maximum, in each subgraph, the value of corresponding ε gradually increases from 0.3 to 0.4. At this time, the quality of community division is the best. It can be seen that the value of ε can be divided into two situations. When the social attribute has a high similarity with the community theme, $\varepsilon = 0.4$. On the contrary, when the similarity between the social attribute and the community theme is weak, $\varepsilon = 0.3$.

5.2.2. Community Overlap Analysis. In this paper, based on the synthetic data set, we compare the proposed algorithm with the single-layer network community detection algorithm and the multilayer network community detection algorithm, respectively, to verify the community overlap performance of the proposed algorithm.

FIGURE 4: The proportion of each algorithm correctly identifying the classification node. (a) The experimental comparison between the algorithm proposed in this article and the single-layer network community discovery algorithm. (b) The experimental comparison between the algorithm proposed in this paper and the algorithm of multilayer network community discovery.

We chose some classic single-layer network community detection algorithms, such as local optimization algorithm (LCDNN) [12], improved label propagation algorithm (TNS-LPA) [14], multiobjective optimization algorithm (PMOEA) [15], and nonnegative matrix factorization community detection algorithm (MOCBT) [16].

We also choose some high-performance multilayer network community discovery algorithms, such as multilayer network community detection algorithm based on semisupervised nonnegative matrix factorization (S2-JNMF) [27], multiobjective optimization-based multilayer network overlap community detection algorithm (OMCS) [18], and the MTLCD [28] algorithm that expresses the trust relationship in the form of a tensor.

In this paper, we adopt the idea of expansion and merging [17] and apply the single-layer network community detection algorithm to the multilayer network. It can be seen from Figure 4 that as the matching degree δ gradually increases, the ratio of the proposed algorithm to correctly identify nodes is significantly higher than that of the comparison algorithm. δ reflects the degree of similarity between the social attributes and the community theme set. The smaller the value of δ , the more chaotic the node partition. As the value of δ increases, node partitions become clearer, and the accuracy of overlapping nodes falling into different communities gradually improves. In general, the proposed algorithm has better performance in detecting overlapping communities.

5.2.3. Verify Nature 2. We conduct experimental analysis on the synthetic network and the real network, respectively, to verify the correctness of *nature 2*.

By adjusting the matching degree δ , we observe the changes in the community. The value of δ is 0.3-1, corresponding to 8 synthetic networks. The experimental results

FIGURE 5: Schematic diagram of community detection of synthetic multilayer social network.

are shown in Figure 5. As the value of δ increases, the quality of community detection is better. When $\delta = 1$, the community detection basically tends to be stable, and the detection result is the best. The experimental results show that the structural relationship of user nodes is only a prerequisite for the formation of communities, and what really matters is the social attributes of user nodes.

We split the proposed algorithm according to social attribute features and structural features and divide it into a community detection algorithm that only considers social attribute features (PMOCBT), a community detection algorithm that only considers structural features (SMOCBT), and the proposed algorithm (MOCBT). The experimental results are shown in Figure 6. In the four real data sets, the

FIGURE 6: The Q value of each algorithm in the real network.

TABLE 3: Experimental results of the proposed algorithm and each single-layer community detection algorithm in the synthetic network.

Data set	Evaluation index	LCDNN	TNS-LPA	PMOEA	MDNMF	MOCBT
G1	Q	0.503	0.516	0.572	0.581	0.624
	F-score	0.529	0.570	0.593	0.599	0.617
G2	Q	0.586	0.594	0.601	0.611	0.688
	F-score	0.614	0.607	0.626	0.637	0.736
G3	Q	0.630	0.624	0.617	0.649	0.690
	F-score	0.634	0.649	0.640	0.671	0.708
G4	Q	0.661	0.638	0.640	0.661	0.704
	F-score	0.691	0.662	0.699	0.672	0.838
G5	Q	0.682	0.649	0.680	0.694	0.715
	F-score	0.716	0.698	0.724	0.715	0.775
G6	Q	0.738	0.704	0.715	0.728	0.826
	F-score	0.801	0.736	0.747	0.783	0.866
G7	Q	0.783	0.780	0.746	0.772	0.904
	F-score	0.824	0.816	0.793	0.824	0.941
G8	Q	0.867	0.842	0.834	0.882	0.954
	F-score	0.890	0.903	0.906	0.898	0.971

Q value of the SMOCBT algorithm is significantly lower than that of the PMOCBT and MOCBT algorithms. At the same time, the Q value of the PMOCBT algorithm is closer to that of MOCBT. The experimental results further show the correctness of the proposed *nature 2*.

5.2.4. Compared with Other Algorithms on the Synthetic Network. The experimental results are shown in Table 3. It can be seen from the table that the evaluation index of the proposed algorithm is significantly better than that of other

single-layer network community detection algorithms. Multilayer networks cover a variety of complex relationships between users and are a further extension of single-layer networks. The proposed algorithm not only considers the user's structural relationship but also combines the user's attribute characteristics, so that the community detection result is significantly better than the single-layer network.

It can be seen from Table 4 that the evaluation index of the MTLCD algorithm based on the trust relationship is significantly better than that of the S2-JNMF and OMCS

TABLE 4: Experimental results of the proposed algorithm and each multilayer community detection algorithm in the synthetic network.

Data set	Evaluation index	S2-JNMF	OMCS	MTLCD	MOCBT
G1	Q	0.545	0.575	0.607	0.623
	<i>F</i> -score	0.602	0.609	0.610	0.618
G2	Q	0.648	0.637	0.664	0.688
	<i>F</i> -score	0.632	0.646	0.711	0.736
G3	Q	0.641	0.636	0.648	0.691
	<i>F</i> -score	0.676	0.688	0.693	0.708
G4	Q	0.676	0.664	0.693	0.704
	<i>F</i> -score	0.736	0.771	0.801	0.838
G5	Q	0.702	0.700	0.709	0.715
	<i>F</i> -score	0.751	0.743	0.765	0.775
G6	Q	0.803	0.801	0.813	0.826
	<i>F</i> -score	0.849	0.831	0.857	0.866
G7	Q	0.881	0.849	0.899	0.904
	<i>F</i> -score	0.907	0.897	0.913	0.941
G8	Q	0.901	0.931	0.940	0.958
	<i>F</i> -score	0.930	0.924	0.955	0.972

algorithms based on the structure relationship, and the proposed algorithm has the best performance compared to other comparison algorithms. The MTLCD algorithm obtains local communities on the network by establishing a trust model. The experimental results show that using the trust value to divide the community has better performance. However, this trust model essentially only uses the structural characteristics of the node, while ignoring the attribute characteristics of the node. Our proposed algorithm combines the characteristics of social attributes to ingeniously transform the complex relationships of multilayer social networks into trust relationships between users. We use the local information of the core node to divide the community and discard the inaccurate community results brought by the global information of the network, and the proposed algorithm shows a better detection result. In the data set G1, since the matching degree is low, the values of *Q* and *F*-score are relatively low. In data set G8, since the matching degree is the largest, the values of *Q* and *F*-score are relatively high. This fits our description of the nature of the community.

5.2.5. Compared with Other Algorithms on the Real Network.

We use four real multilayer social network data sets, namely, Twitter number, QQ Zone, Bio GRI, and Remote Sensing data sets, to verify the feasibility and scalability of our proposed algorithm. Since the result of a community in a given real network is unknown, this paper uses modularity as a measure of community detection. It can be seen from Table 5 that the community results of the MTLCD algorithm based on the trust relationship are better than those of the S2-JNMF and OMCS algorithms based on the structural relationship. Compared with other comparison algorithms, our proposed algorithm has the best evaluation

TABLE 5: Experimental results of each algorithm in the real network.

Comparison algorithm	Twitter	QQ Zone	Bio GRID	Remote Sensing
PMOEA	0.635	0.621	0.548	0.579
MDNMF	0.625	0.599	0.586	0.610
TNS-LPA	0.665	0.619	0.601	0.622
LCDNN	0.652	0.625	0.599	0.617
S2-JNMF	0.729	0.688	0.645	0.680
OMCS	0.735	0.691	0.623	0.667
MTLCD	0.748	0.710	0.651	0.689
MOCBT	0.753	0.724	0.686	0.703

index. It can be seen that the trend of the real network experiment results is similar to that of the synthetic network. At the same time, in these two data sets, the modularity *Q* is greater than 0.5, which meets the standards of community testing [32].

6. Conclusion and Summary

Although traditional community detection algorithms based on trust relationships greatly improve the accuracy of community discovery, these methods rely excessively on the structural characteristics of the network. Therefore, to solve the problem of the lack of social attributes when constructing user trust relationships in multilayer social networks, we propose a multilayer social network overlapping community detection algorithm based on trust relationship. By integrating structural trust and social attribute trust, we transform a complex multilayer social network into a single-layer trust network. Based on the community

discovery algorithm of trust value, we obtain the overlapping community structure of multilayer social networks. We carried out experimental comparative analysis on the synthetic network and the real network, respectively. The experimental results show that the proposed algorithm has higher Q and F -score compared with other algorithms of the same type, which effectively reveals the community structure of the network. However, the user's trust relationship is not static, but changes over time. Therefore, in a future work, we will focus on how to establish a dynamic trust model and design a fast and efficient dynamic overlapping community discovery algorithm.

Data Availability

Data are available at <http://twitter.mpi-sws.org/data-icwsm2010.html>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61967013) and was also supported by Northwest Normal University Graduate Research Funding Project (2020).

References

- [1] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 7821–7826, 2002.
- [2] B. Zhou, "Privacy issues in discovering communities in social networks," in *User Community Discovery*, Human-Computer Interaction Series, G. Paliouras, S. Papadopoulos, D. Vogiatzis, and Y. Kompatsiaris, Eds., pp. 135–155, Springer, Cham, 2015.
- [3] M. Magnani and L. Rossi, "The ML-model for multi-layer social networks," in *2011 International Conference on Advances in Social Networks Analysis and Mining*, pp. 5–12, Kaohsiung, Taiwan, 2011.
- [4] M. de Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Communications*, vol. 6, no. 1, article 6864, 2015.
- [5] Q. Jiang, A. Campbell, G. Tang, and J. Pei, "Multi-level relationship outlier detection," *International Journal of Business Intelligence and Data Mining*, vol. 7, no. 4, pp. 253–273, 2012.
- [6] X. Liu, F. Wang, Z. Li et al., "Topic detection based on similar networks," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 188–193, Banff, AB, Canada, 2017.
- [7] P. Moradi, S. Ahmadian, and F. Akhlaghian, "An effective trust-based recommendation method using a novel graph clustering algorithm," *Physica A*, vol. 436, pp. 462–481, 2015.
- [8] A. Bouyer and H. Roghani, "LSMD: a fast and robust local community detection starting from low degree nodes in social networks," *Future Generation Computer Systems*, vol. 113, pp. 41–57, 2020.
- [9] W. Rebhi, N. Ben Yahia, and N. B. B. Saoud, "Hybrid community detection approach in multilayer social network: scientific collaboration recommendation case study," in *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–8, Agadir, Morocco, 2016.
- [10] H. Li, D. Zhao, and H. Zhang, "Community detection method based on seed extension [J]," *Journal on Communications*, vol. 39, no. S2, pp. 138–143, 2018.
- [11] J. Ma and J. Fan, "Local optimization for clique-based overlapping community detection in complex networks," *IEEE Access*, vol. 8, pp. 5091–5103, 2020.
- [12] W. Luo, N. Lu, L. Ni, W. Zhu, and W. Ding, "Local community detection by the nearest nodes with greater centrality," *Information Sciences*, vol. 517, pp. 377–392, 2020.
- [13] F. Cheng, C. Wang, X. Zhang, and Y. Yang, "A local-neighborhood information based overlapping community detection algorithm for large-scale complex networks," *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 543–556, 2021.
- [14] G. Xu, J. Guo, and P. Yang, "TNS-LPA: an improved label propagation algorithm for community detection based on two-level neighbourhood similarity," *IEEE Access*, vol. 9, pp. 23526–23536, 2021.
- [15] Y. Su, K. Zhou, X. Zhang, R. Cheng, and C. Zheng, "A parallel multi-objective evolutionary algorithm for community detection in large-scale complex networks," *Information Sciences*, vol. 576, pp. 374–392, 2021.
- [16] J. Huang, T. Zhang, W. Yu, J. Zhu, and E. Cai, "Community detection based on modularized deep nonnegative matrix factorization," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 35, no. 2, article 2159006, 2021.
- [17] T. N. Afsarmanesh and M. Magnani, "Partial and overlapping community detection in multiplex social networks," in *Social Informatics*, S. Staab, O. Koltsova, and D. Ignatov, Eds., vol. 11186 of SocInfo 2018. Lecture Notes in Computer Science, pp. 15–28, Springer, Cham, 2018.
- [18] M. R. Shahmoradi, M. Ebrahimi, Z. Heshmati, and M. Salehi, "Multilayer overlapping community detection using multi-objective optimization," *Future Generation Computer Systems*, vol. 101, pp. 221–235, 2019.
- [19] L. Min and Y. Liang, "Construction of interpersonal trust in multi-ethnic communities," *Social work and management*, vol. 20, no. 6, pp. 60–69, 2020.
- [20] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *RecSys '10: Fourth ACM Conference on Recommender Systems*, pp. 135–142, Barcelona, Spain, 2010.
- [21] J. Wu and F. Chiclana, "A social network analysis trust-consensus based approach to group decision-making problems with interval-valued fuzzy reciprocal preference relations," *Knowledge-Based Systems*, vol. 59, pp. 97–107, 2014.
- [22] A. Marko, "Exposing multi-relational networks to single-relational network analysis algorithms," *Journal of Informetrics*, vol. 4, no. 1, pp. 29–41, 2010.
- [23] D. Chen, P. du, Q. Jiang, X. Huang, and D. Wang, "A feasible community detection algorithm for multilayer networks," *Symmetry*, vol. 12, no. 2, p. 223, 2020.
- [24] X. Ma, D. Dong, and Q. Wang, "Community detection in multi-layer networks using joint nonnegative matrix factorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 273–286, 2019.

- [25] X. Gao, Q. Zheng, V. Filipe, R. D. Rodrigues, and L. Zhao, "Particle competition for multilayer network community detection," in *ICMLC '19: 2019 11th International Conference on Machine Learning and Computing*, pp. 75–80, Zhuhai, China, 2019.
- [26] X. Chen, C. Xia, and J. Wang, "A novel trust-based community detection algorithm used in social networks," *Chaos, Solitons & Fractals*, vol. 108, pp. 57–65, 2018.
- [27] S. Ding, Z. Yue, S. Yang, F. Niu, and Y. Zhang, "A novel trust model based overlapping community detection algorithm for social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 11, pp. 2101–2114, 2020.
- [28] X. Li, Q. Tian, M. Tang, X. Chen, and X. Yang, "Local community detection for multi-layer mobile network based on the trust relation," *Wireless Networks*, vol. 26, no. 8, pp. 5503–5515, 2020.
- [29] F. Reid and N. Hurley, "Diffusion in networks with overlapping community structure," in *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 969–978, Vancouver, BC, Canada, 2011.
- [30] Q. Zeqi and Q. Tianyu, "Weak ties, strong ties, or the homophily based on social attribute?," *The Sociological Review*, vol. 6, no. 1, pp. 3–20, 2008.
- [31] M. Contisciani, E. A. Power, and C. De Bacco, "Community detection with node attributes in multilayer networks," *Nature Scientific Reports*, vol. 10, no. 1, article 15736, 2020.
- [32] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, article 026113, 2004.
- [33] J. Liu, D. Wang, S. Feng, and Y. Zhang, "Local community discovery approach based on fuzzy similarity relation [J]," *Journal of Software*, vol. 31, no. 11, pp. 3481–3491, 2020.