Received October 8, 2018, accepted October 22, 2018, date of publication October 30, 2018,
date of current version October 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2878674

Community Detection Based on Local Information and Dynamic Expansion

YONGPING LUO, LI WANG, SHIWEN SUN, AND CHENGYI XIA
Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, China
Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin 300384, China

Corresponding authors: Li Wang (wltjut08@126.com) and Chengyi Xia (xialooking@163.com)

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 61773286, Grant 61374169, and Grant 61403280.

ABSTRACT Mining the community structure in the real-world networks has been a hot topic in the field of complex networks, and has emerged as a prominent research area and continues to grow with the introduction of requirements for personalized recommendation. However, most of the existing community detection algorithms are based on global information, fewer works are devoted to detecting the communities hidden in the network by using local information. To this end, in this paper, we propose two improved signed modularity functions to evaluate the community properties in complex network, and then we apply these indicators to identify the community structure by using the local information since it is difficult to obtain the global information in practice. During the dynamic expansion, each local community will absorb the neighboring node with the highest positive energy; in addition, a new local community is generated when all local communities cannot contain the neighboring node. Finally, the algorithm has been applied to unsigned networks and signed networks, respectively. The experimental results show that the division results given by our proposed algorithm are in line with the actual ones in artificial and real-life networks.

INDEX TERMS Community detection, global community, local modularity, local information, local community.

I. INTRODUCTION

As the information technology expands rapidly, the interactions on man-man, man-object, and object-object are becoming increasingly frequent and complicated. The information generated by different individuals and their interaction forms a complex system in which each individual plays a different role (e.g., bridge node, terminal node), and that individuals belong to the same cluster within a certain range of area (usually, there exists a core individual). Over the past two decades, researchers have studied the important characteristics, features or effects that are concealed inside the system by abstracting systems into complex networks or graphs, for instance, social networks [1]–[3], citation networks [4]–[6], transportation systems [7], international war networks [8] and so on. Although these characteristics have always existed on real life, they have not received sufficient attention. However, it is these characteristics that make the evolution of systems appear to become relatively diverse. Among then, several important topics have been extensively investigated: community structure [9] expounded that an individual is much more connected with members within the community than those outside the community; influential spreaders [10] explained that the more the interaction of an individual, owned the more influential is not necessarily correct, and that an influential spreaders can also be measured by other indicators; six degree of separation [11] indicated that any two individuals could be connected through at most five intermediaries; the butterfly effect [12] explicated that the complexity of the development of things, a small change can be drive the long-term and huge chain reaction of the whole dynamic system; cascade effect [13] interpreted that individual behavior will affect the system reliability and lead to a series of unexpected events, to name but a few.

In particular, the community structure has attracted a great deal of concern since the community hidden in the network — whether it is an unsigned network with positive edges or a signed network with positive edges and negative edges — represents the similarity of the internal members, and the similarity is of great significance for the prediction and the recommendation in the era of the big data [14]. So, researchers have proposed a series of algorithms that perform well for different networks (unsigned or signed, undirected or directed, and unweighted or weighted) in term of accuracy or efficiency. For instance, in unsigned networks, Newman and Girvan [9] proposed the GN algorithm based on the betweenness, Blondel *et al.* [15] offered the Louvain algorithm for optimizing the function of modularity which is proposed by Newman and Girvan [9], and Li *et al.* [16] studied the dynamical system and the general function of modularity to detect the community; in signed networks, Gómez *et al.* [17] presented an improved function of modularity used to detect the community structure, Traag and Bruggeman [18] combined the Potts model to improve the function of modularity, Li *et al.* [8] studied the hostile or alliance of relationship in the war network; Wen *et al.* [19] introduces the maximal-clique graph to study the overlapping community detection, etc.

Although there exists many effective community detection algorithms, they are based on the global information for community detection, that is to say, the global information must be considered during the operation of the algorithm since it will exert an influence on the formation of the local community. However, it is difficult or even impossible to obtain global information, which makes the above algorithms unable to perform efficiently and accurately in general. Hence, in order to solve the above problems, some researchers have proposed algorithms for local community detection by using local information, such as Clauset [20] proposed a detection algorithm based on local modularity, Luo *et al.* [21] proposed a new local modularity for community detection and so on.

Local community detection algorithms verified that local information can be used for the community detection, but these algorithms [20], [21] are mainly used to detect a local community rather than all communities. In fact, these algorithms start from an initial node to detect nodes that are in the same community with the initial node, meanwhile the maximum number of community members must be set up, it means that the problem of global community detection using local information has not been solved. In order to control community size, Eustance *et al.* [22], [23] presented a ratio function based on the local community neighborhood. A new detection method for overlapping communities is proposed in [24], which a structural center is mining by the local density and the relative distance of a node, and then expands community structure based on identified structural centers. In [25], the characterization of the short texts outperformed in the classification, which suggests that local information might improve its global characterization in large documents.

Thus, we propose a dynamic expansion algorithm that uses local information for community detection. The proposed algorithm is slightly similar to the random walk algorithm [26], while our algorithm is equivalent to multiple walkers moving around the network, at the same time, the walkers will choose and join the best community that are most beneficial to them; it is also similar to the label propagation algorithm [27], but only one node has a community label at the beginning of the proposed algorithm, and dynamically transfers community label or adds new community label with the increase of the detection area; the proposed algorithm is also more similar to the greedy algorithm [28] because only the most valuable community is retained when multiple individuals choose and try to join the community.

The main contributions can be summarized as follows:

1) The function of signed modularity is proposed to measure the division result of global community. The function takes into account the sign property of the signed network by the value of the function, and can be converted into the original function of modularity proposed by Newman *et al.*
2) The signed local modularity is proposed to measure the quality of the local community. Compared with the existing signed modularity, the information used in the new function is the local information.
3) A dynamic expansion algorithm for global community detection using local information is proposed. The information used by the algorithm is built on local community information or individual information, which means that the information is directly contacted by individuals or local communities. Meanwhile, the number of communities will expand gradually from one to the right number, and the operation of expansion depends on the energy value of the community.

The rest of this paper is structured as follows. Section II introduces existing community detection algorithms and measure indexes. Section III presents the improvement works we have done. Section IV describes the proposed algorithm and analyzes the time complexity. Section V shows the results of detection in different networks. Finally, some concluding remarks are summarized in Section VI.

II. MOTIVATION

A. RELATED WORKS

Most of the researches on community detection algorithm start with unsigned networks because there are fewer constraints to be considered, and these networks are also the basis for studying other types of networks.

In unsigned networks, Newmam and Girvan [9] proposed a community detection algorithm based on betweenness. The algorithm is divided into four steps: the first step is to calculate the betweenness for all edges in the network; the second step is to remove the edge with the highest betweenness; the third step is to recalculate the betweenness for edges that remaining in the network; the fourth step is to repeat the second step and the third step until there is no edge left. It can be known that the algorithm takes a lot of time to calculate the betweenness, that is, the time complexity is as high as $O(n^3)$, where n represents the number of nodes in the network, thus it can work well in the sparse network.

By optimizing the modularity, Blondel *et al.* [15] offered a fast detection algorithm, which can be divided into two steps: in the first one, each node can be considered as an independent community, and then the node moves into its best community — the node falling into the community will maximize the modularity of the community; in the second period, the nodes belonging to the same community are folded into new ones to rebuild the network, and then repeat the first stage until the value of modularity cannot be improved. For large-scale networks, the speed of the algorithm can be greatly accelerated during the execution of the algorithm.

By combining with vector distance and sparse spectral clustering, Mahmood and Small [29] presented a novel algorithm for community detection. First, the n nodes in the network are mapped into n-dimension real space such that each dimension represents a certain type of distance from a particular node to the i-th node. Then, using the Gaussian kernel function to convert the distance vector into a similarity score vector, we can know that for two nodes belonging to different communities, their distance is infinite and the similarity score is zero. Next, each column vector is selected and a remainder of it is used to find a matrix of symmetric linear coefficient in the similarity score vector. Finally, spectral clustering is performed on the coefficient matrix, and the detected segmentation results are indicated by the second least significant eigenvector of a symmetric Laplacian matrix.

The CDFR and the CDFR-U algorithm based on the fuzzy relations were proposed by Luo *et al.* [30], which the NGC node (Nearest node with Greater Centrality) plays an important role that determines the community label of each node. The centrality of each node is calculated according to the degree of each node and the degree of the direct neighbors, and then a special algorithm is used to find the NGC nodes and calculate the fuzzy relation between each node and its NGC node. To get a community partition scheme, a threshold needs to be set, which determines whether the node is a new community or the community in which belongs to its NGC node. Similarly, Moosavi *et al.* [31] implemented community mining of social networks in four steps: preprocessing on dataset, user frequent pattern mining to obtain harmonious groups, confirmation of harmonious groups as small communities, and expanding a small community.

In addition to the above algorithms, many scholars have studied community detection in unsigned networks and proposed quite a few novel algorithms. For example, Khan *et al.* [3] studied virtual community detection algorithm based on the common interests of individuals in social networks; He *et al.* [32] used segmentation technology to divide the network into thousands of sub-networks to study scalable communities; Deng *et al.* [33] studied the community detection algorithm of directed networks with owner-member relationship based on clustering coefficients and classical probability graph model; Qi *et al.* [34] combined linkage behavior and edge content to present a novel edge-induced matrix factorization algorithm.

Although there are many mature algorithms for community detection in unsigned network, these algorithms cannot be always directly used in real-world networks since some networks in the real-world are signed. Therefore, a multitude of scholars have proposed many community detection algorithms that can be applied into signed networks after considering the sign property of signed networks.

Yang *et al.* [35] proposed a random walk algorithm in signed networks. Firstly, a node without a community tag will be selected as the starting node, then the node can conduct a certain number of walks to determine the set of node that it can reach. Next, a truncation function is proposed and used to determine which nodes are located in the same community with the initial node, in the meanwhile, the function of truncation considers the distribution of positive edges and negative edges between the community and the remaining part. As shown in [35], the result of the matrix presents the characteristics of the block matrix, and the matrix will be divided into different block matrices that characterize different communities by using the calculated truncation function to segment.

Cai *et al.* [36] improved a multi-objective discrete particle swarm optimization algorithm that can be used for multi-resolution network clustering. In the first step, the initialization phase generates the initial population information, including position, velocity, individual optimal solution, and neighborhood. The second step is the iterative update phase, first, randomly selects a particle from the neighborhood as the leader; then, calculates the new velocity and position of each particle, meanwhile adds a small disturbance to the new location; finally, the location after the disturbance is evaluated, the neighborhood information and optimal solution of the individual are updated according to the evaluation result.

Liu *et al.* [37] proposed a multi-objective evolutionary algorithm based on the similarity. The authors put forward two new indexes for taking into account the sign property of the signed network: one of the indexes is the signed similarity index based on the existing similarity index [38]; another one is the signed tightness index according to the existing tightness index [38]. During the execution of the algorithm, the node will be moving into another community when the movement makes the signed tightness of community increased, but the node will be independent as a new community if the move operation cannot increase the signed tightness of community. A special case is that multiple communities have the same nodes in the execution of the algorithm, so the two communities will be merged into one new community when they have more than half of the common nodes.

At the same time, Jiang [39] applied the Signed Stochastic Block-Model into community detection; Harakawa *et al.* [40] divided the web video by constructing a weighted signed network and maximizing the local modularity; Li *et al.* [41] proposed a new algorithm based on the improved function of the modularity and the density, while studied the performance of the proposed algorithm under various parameters.

Most of the algorithms introduced above are based on global information for community detection. However, with the increasing demand of personalized recommendation, which enables quite a few scholars have studied the community detection algorithm for specific individuals.

West *et al.* [5] proposed eigenfactor recommendation based on citation, which recommends similar articles for users from the hierarchical structure of citations. In the first step, a citation network is constructed based on the citation relationship among papers. In next step, the eigenfactor ranking is performed on each node in the citation network, which shows the best performance compared with the traditional PageRank algorithm since citation networks contain no loops. In the third step, the hierarchical mapping equation is used to cluster the network according to the previous results and then help to reveal the multi-level structure for modularity. The final step is to provide the corresponding expert, classic, and serendipity papers to specific user to achieve the purpose of the personalized recommendation.

According to the motion information of different individuals, Liu and Wang [42] classified the individuals in the network by a new similarity index, and implemented the targeted recommendation. The algorithm can be divided into four stages. In the first stage, four different dimensions (position, duration, proximity to other objects, speed of motion) are modeled, and the key features are extracted by applying the appropriate kernel into different dimensions in order to calculate the similarity. In the second phase, multi-modal diffusion process and weighted similarity measure are constructed. The third period uses dense sub-graph detection algorithm to divide individual trajectories by probabilistic clustering. In the fourth stage, the similarity is calculated according to the specific user, and the on-line recommendation is carried out by the memory-based collaborative filtering algorithm.

Athria and Thampi [43] studied the proximity of the posts through linguistic features, and then implemented post recommendations for members within the community. First, they clustered post from five aspects: sentiment, emotion, theme, stylistics and psycholinguistic. Then, according to the posts that the related users like or share, the clustering vector is constructed by the five aspects, and the user's community division is realized by dividing the clustering vector. Finally, the recommended posts and individuals are determined based on the acceptability function defined in the post and the popularity function defined in the post and the individual.

There are still many algorithms to be recommended through community detection. For example, Parimi and Caragea [44] introduced community similarity and sense of belonging; Rafailidis [45] combined non-negative matrix factorization and spectral clustering algorithms to implement community detection and recommendation in signed networks; Qiu *et al.* [46] implemented community detection by processing the rating matrix of the network instead of the entire network; Amancio [47] modeling texts as complex network to study features of the language that reflects particular choices made by individuals or groups.

B. MEASURE INDEXES

Newman and Girvan [9] proposed the community structure that is defined as a set of nodes in which the number of edges inside the same community is higher than the number of edges connecting the neighboring communities. Meanwhile, the index of modularity is also proposed for measuring the quality of the community, the greater the value, the better the community structure. The index is defined as follows:

$$Q = \sum_{i \in C}(e_{ii} - a_i^2) = \sum_{i \in C}[\frac{w_{i,in}}{w} - (\frac{w_i}{2w})^2], \tag{1}$$

where C is the division result of the communities in the network, for the unweighted (or weighted) network, e_{ii} describes the fraction of the total number (or weights) $w_{i,in}$ of edges within community i to the total number (or weights) w of edges in the network, a_i indicates the proportion of the total number (or weights) w_i of edges associated with the nodes of community i to double the total number (or weights) w of edges in the network.

However, the above function of modularity cannot be applied into the signed network directly since there exist negative edges. Using the existing function of modularity, Gómez *et al.* [17] put forward the function of modularity for signed networks. It can be formulated as:

$$Q = \frac{\sum_i \sum_j [w_{ij} - (\frac{w_i^+ w_j^+}{2w^+} - \frac{w_i^- w_j^-}{2w^-})]\delta(C_i, C_j)}{2w^+ + 2w^-}, \tag{2}$$

where Q means a weighted sum of the two parts of the value of modularity; w_i^+ or w_i^- describes the positive or negative weights of the edges of the node i; w_{ij} represents the weights of the edge between the node i and the node j; C_i denotes the community in which the node i is located; $\delta(C_i, C_j)$ is a piecewise function of variables C_i and C_j, which is one if the variables are equal, and zero otherwise.

In fact, we usually know the community partition scheme in advance when constructing the network. At this point, more desirable to give the similarity between the detection result given by the algorithm and the actual result, not just the quality of the division result. Therefore, Danon *et al.* [48] used related fundamentals of information theory to propose a numerical indicator of Normalized Mutual Information (NMI). It can be expressed as:

$$NMI = \frac{-2\sum_{i,j} N_{ij} ln(\frac{N_{ij} n}{N_{i.} N_{.j}})}{\sum_i N_{i.} ln \frac{N_{i.}}{n} + \sum_j N_{.j} ln \frac{N_{.j}}{n}}, \tag{3}$$

where N represents a difference matrix between the detection result and the actual partition; N_{ij} describes an element on row i and column j of matrix N, $N_{i.}$ indicates the sum of the elements on row i of matrix N, and $N_{.j}$ denotes the sum of the elements on column j of matrix N. Equation (3) measures the degree of correlation between the two partitioned results, and the range of value lies in [0, 1]. The larger the value, the more similar the two results; the smaller the value, the more dissimilar the two results.

Although the equations mentioned above can measure the quality of the community, they all use global information to measure the quality of the local community. Luo *et al.* [21], [49] offered a metric called local modularity M to measure the ratio of the edges within the local community to the edges outside the local community.

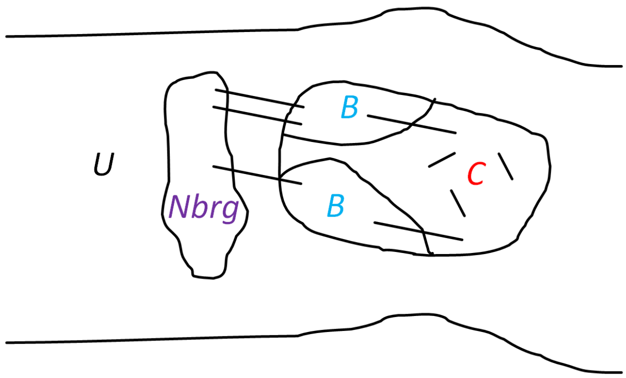

FIGURE 1. Local community: $U \bigcup C \bigcup B \bigcup Nbrg$ represents the network; $C \bigcup B$ is a local community; C is the core of local community; U is unknown area; B is the boundary of local community; $Nbrg$ is the neighbor of B.

The metric is given as follows:

$$M = \frac{e_{in}}{e_{out}}, \tag{4}$$

where e_{in} indicates the number of edges within the local community, e_{out} denotes the number of edges outside the local community.

III. OUR WORKS

A. LOCAL COMMUNITY

The definition of community structure has been mentioned above, but using the previous definition violates the premise of using local information because local information means the part that the current individual or community can reach, and the part that can be reached through the intermediary is an unknown area for the current individual or community. Thus, the concept of local community needs to be defined.

Definition 1 (Local Community): In the network, there exists a node set C where the nodes are with the same community label, and there are a large number of edges with the same sign within C, while the sign of edges outside C are different from the sign of edges inside C, or the sign of a small number of edges outside C are same as the sign of inner edges of C.

Fig.1 shows one of the local communities in the network. An edge is represented by a short line with an endpoint (or two endpoints) in $Nbrg$ or B or C; $C \bigcup B$ is a local community; U describes the area that is unknown to the local community; C represents the core of the local community — the two nodes of the edges belong to the local community; B denotes the boundary of the local community — one node of the edges belong to the local community, while the other one may be considered as an independent local community or located in another local community; $Nbrg$ indicates the neighbor of the local community — one node of the edges must belong to the local community, while the other one may be considered as an independent local community or located in another local community.

B. IMPROVED SIGNED MODULARITY

After analysing the existing signed modularity function proposed by Gómez *et al.* [17], it can be found that it is weighted by the modularity of the positive and negative communities, and the modularity value does not fully express the characteristics of signed networks. Compared with unsigned networks, signed networks have two kinds of sign at the same time. If the sign property of the network can be expressed in the detection results, it will give more meaning to the detection results, thus how to show it has become a meaningful research content.

According to the modularity function proposed by Newman and Girvan [9], if we divide the network into two sub-networks, one sub-network holding only positive edges, and the other one including only negative edges. Then when the modularity of the two sub-networks is calculated, their results actually differ by a sign. At the same time, the community detection problem is generally directed at the positive community, that is, to ensure that there are a large number of positive edges within the community, and there are a large number of negative edges between the communities. So we can use the modularity value of positive network subtracts the modularity value of negative network, and then the final value of the modularity is used as the modularity value of the entire network.

Therefore, we propose a function of signed modularity that takes into account the sign property of signed networks. It can be expressed as follows:

$$SQ = \sum_{i \in C} \{[e_{ii}^{+} - (a_i^{+})^2] - [e_{ii}^{-} - (a_i^{-})^2]\}, \tag{5}$$

where C is the division result of the communities in the network; e_{ii}^{+} and e_{ii}^{-} represent the proportion of the total number (or weights) of positive edges and negative edges within the community i to the total number (or weights) of edges of the whole network, respectively; a_i^{+} or a_i^{-} denotes the ratio of the total number (or weights) of positive or negative edges that connected to the internal nodes of the community i to double the total number (or weights) of edges of the whole network.

C. SIGNED LOCAL MODULARITY

The existing modularity function and signed modularity function mentioned above are all based on the global information for community detection, but the cost of obtaining global information is enormous when the number of nodes reach one million or more, while the number of edges in the network will also be larger. Therefore, we propose a signed local modularity function for evaluating the quality of the local community using local information, which can be expressed as follows:

$$LQ(i) = \frac{m_{i,in}^{+}}{m_{i,in}^{+} + m_{i,out}^{+}} - \frac{m_{i,in}^{-}}{m_{i,in}^{-} + m_{i,out}^{-}}, \tag{6}$$

where $m_{i,in}^{+}$ or $m_{i,in}^{-}$ represents the number (or weights) of positive or negative edges within the local community i;

$m^+_{i,out}$ or $m^-_{i,out}$ denotes the number (or weights) of positive or negative edges outside the local community i.

D. METHOD

Using local information to detect the global community is actually a process of dynamic expansion because the local community or node cannot get any information related to the whole system or network in the whole detection process, and it can only get information about nodes or local communities that are in direct contact with it. For a local community, it evaluates the energy of all the nodes that it can reach, then adds the node with the maximum energy to its own interior; for a node, it chooses to join the local community with the greatest increase in energy.

In the network $G(V, E)$, let $C(v_i, t)$ represents the local community which contains v_i at time t, $B(C(v_i, t))$ denotes the boundary of the local community, $Nbrg(v_i, t)$ indicates the neighbor of the node v_i. For each local community, we define a Hamiltonian function to represent the energy of local community based on the singed local modularity function:

$$\mathcal{H}(\{C(v_i,t)\}) = \frac{\mathcal{W}^+_{C(v_i,t),in}}{\mathcal{W}^+_{C(v_i,t),in} + \mathcal{W}^+_{C(v_i,t),out}} - \frac{\mathcal{W}^-_{C(v_i,t),in}}{\mathcal{W}^-_{C(v_i,t),in} + \mathcal{W}^-_{C(v_i,t),out}}, \quad (7)$$

where $\mathcal{W}^+_{C(v_i,t),in}$ or $\mathcal{W}^-_{C(v_i,t),in}$ represents the positive or negative energies inside the local community $C(v_i, t)$, $\mathcal{W}^+_{C(v_i,t),out}$ or $\mathcal{W}^-_{C(v_i,t),out}$ indicates the positive or negative energies outside the local community $C(v_i, t)$.

Meanwhile, for each node $v_j \in Nbrg(B(C(v_i, t)))$, there will definitely be an equation $\mathcal{W}_{v_j} = \mathcal{W}_{v_j,B(C(v_i,t))} + \mathcal{W}_{v_j,Nbrg(v_j,t)-B(C(v_i,t))}$. When trying to add node v_j to the local community where v_i is located, the energy of the local community will change as follows:

$$\mathcal{W}^\pm_{C(v_i,t+1),in} = \mathcal{W}^\pm_{C(v_i,t),in} + \mathcal{W}_{v_j,B(C(v_i,t))}, \quad (8)$$

$$\mathcal{W}^\pm_{C(v_i,t+1),out} = \mathcal{W}^\pm_{C(v_i,t),out} - \mathcal{W}_{v_j,B(C(v_i,t))} + W_{v_j,Nbrg(v_j,t)-B(C(v_i,t))}, \quad (9)$$

For a community detection algorithm, its goal is to optimize the Hamiltonian function using the global information to detect communities that hidden in the network. Thus the detection result has the largest energy and corresponds to the best division of communities. In the same way, in the case of using local information, for each local community, it is necessary to ensure that its energy reaches a maximum at time t, that is, the following equation is required:

$$\mathcal{H}(\{C(v_i,t)\}) = max_{t' \in [1,t)}\{\mathcal{H}(\{C(v_i,t')\})\}. \quad (10)$$

Therefore, the essence of using local information to expand the community is that as the size of the network grows, the modularity value of the local community will continue to grow as much as possible. In other words, in the process of detecting the communities, it is necessary to ensure that the energy of the local community meets the Equation (10) since it can guarantee the detection result has the highest energy.

So far, it is known that the proposed algorithm depends on how much energy the nodes to be added can provide to the local community. This raises the question of whether the robustness can be guaranteed or not when the nodes are removed, similar to the one in [50], and the answer is No. First of all, the removal of nodes will have an impact on the energy of the network. Specifically, assuming that the removal of node is the bridge node, the network will be divided into two sub-networks after removal, and then resulting in a reduction in the energy value of the network. Secondly, for the core node (or hub node) of a community, when it is removed, it will produce a large number of small communities, which will greatly reduce the energy value of the network.

IV. NEW ALGORITHM

The new algorithm attempts to use local information to detect the global community, i.e., the information that can be utilized is localization. Therefore, in each selection process, it is necessary to find the best local community to be expanded, and then absorbs its neighbors into the local community in order to maximize the energy value of the local community. One of the biggest features of the present algorithm is that no additional control parameters are needed, and even the initial community nodes are not needed because the algorithm can randomly select a node as the initial node.

The algorithm can be divided into four parts as follows:

1) In the first stage, we randomly select one of all nodes as the initial node after initializing the network (the initial node can also be given by parameter), and then the community label of the initial node is set up, in addition, the core part of the local community is updated, the boundary part of the local community is updated, and the energy value of the local community is updated.
2) In the second phase, the neighbors of the local community are moved into the local community if they can increase the energy of the local community, then the core part of the local community is updated, the boundary part of the local community is updated, and the energy value of the local community is updated. At this phase, an individual can only see the local community around it and choose to join the local community that is most beneficial to it.
3) In the third period, there may be cases where there is no node that can be moved after the second phase. At this time, a new local community needs to be generated, the core of the new local community is updated, the border part of the new local community is updated, and the energy value of the new local community is updated. Note that, it is possible that a single-node community will appear in the final detection result since this period will generate a new community, which depends on the network is dynamic or static on the one hand, and on the other hand relies on the topology of network.

4) After completing the third period, the algorithm ends and the result of the division is output when all the nodes in the network have assigned the community label; otherwise, the second phase is repeated.

So, the specific steps of the algorithm are described as **Algorithm 1**.

Here, we will analyze the time complexity of **Algorithm 1**. In the line 1, it takes $O(m)$ time to read the edge set of the network to construct an undirected network; line 2 to line 6 requires $O(|C|k)$ time to initialize the core, boundary, and energy of the local community; line 9 to line 15 requires $O(|C|k^2)$ time to find all neighbor nodes for all boundary nodes, which have no community label; line 21 to line 30 need to spend $O(k)$ time to find out the new energy value of the community after moving nodes into the local community; the worst-case complexity of line 32 to line 40 is $O(k)$, If accept the operation of the move that you will need $O(k)$ time to update the core, boundary, and energy of the local community, otherwise if you do not accept the operation of the move that only $O(1)$ time is required to record the node that needs to generate a new community label; line 18 to line 41 have a time complexity of $O(kN_{UC})$, N_{UC} denotes the number of neighbors — without the community label — of the local community's boundary nodes, and $N_{UC} < |C|k$; the time complexity of line 43 to line 46 is $O(k)$, which requires generate a new local community, and update the core, boundary, and energy of the new local community; the worst time complexity of the line 8 to line 47 is $O(|C|k^2+kN_{UC}+k)$. Here, k is the average degree.

During the execution process of the algorithm, the core, boundary, and energy of the community need to be preserved, and thus the worst space complexity is $O(n)$. In addition, each *while* loop generates the worst $O(n)$ space to save candidate nodes, in which n is the number of nodes.

As a result of the above analysis, if the number of times executed for the *while* operation is L times, the time complexity of the algorithm is $O(m + \sum_{l\in[1,L]}(|C^l|k^2 + kN_{UC}^l + k))$ or the upper limit $O(m + n^2)$; the space complexity is $O((L + 1)n)$, in fact, it can reduces the space complexity to $O(n)$ by using shared space principles in the *while* loop.

Algorithm 1 Local Information to Detect Global Community (LIDGC)

Input: undirected signed network $G(V, E)$
Output: detected communities indicated by C
1: read network $G(V, E)$ by list of edge
2: $C = B = \mathcal{W} = \emptyset$
3: randomly select starting node v_0 and C_{v_0}
4: $C = C \cup \{(v_0, C_{v_0})\}$
5: $B = B \cup \{v_0\}$
6: calculate $\mathcal{W}(C_{v_0})$ and add to $\mathcal{W}$
7: **while** there exists node without community label **do**
8:   $candiNbrg = \emptyset$
9:   **for all** $v_i \in B$ **do**
10:     **for all** $v_j \in Nbrg(v_i)$ **do**
11:       **if** $\{(v_j, C_{v_j})\} \notin C$ **then**
12:         $candiNbrg[C_{v_i}] = candiNbrg[C_{v_i}] \cup \{v_j\}$
13:       **end if**
14:     **end for**
15:   **end for**
16:   $C' = \emptyset$
17:   set *moved* to false
18:   **for all** $candiNbrg[C_{v_i}] \in candiNbrg$ **do**
19:     $best_{v_j} = \emptyset$
20:     **for all** $v_j \in candiNbrg[C_{v_i}]$ **do**
21:       $\mathcal{W}'(C_{v_i}) = \mathcal{W}(C_{v_i})$
22:       **for all** $v_k \in Nbrg(v_j)$ **do**
23:         **if** C_{v_k} equal to C_{v_i} **then**
24:           $\mathcal{W}'(C_{v_i}, in) = \mathcal{W}'(C_{v_i}, in) + E(v_j, v_k, w)$
25:           $\mathcal{W}'(C_{v_i}, out) = \mathcal{W}'(C_{v_i}, out) - E(v_j, v_k, w)$
26:         **else**
27:           $\mathcal{W}'(C_{v_i}, out) = \mathcal{W}'(C_{v_i}, out) + E(v_j, v_k, w)$
28:         **end if**
29:       **end for**
30:       record $best_{v_j}$
31:     **end for**
32:     **if** $best_{v_j}$ increases the energy of $C(v_i)$ **then**
33:       set *moved* to true
34:       $\mathcal{W}(C_{v_i}) = \mathcal{W}'(C_{v_i})$
35:       $B = B \cup \{v_j\}$
36:       remove the core of C_{v_i} from B
37:       $C = C \cup \{(v_j, C_{v_i})\}$
38:     **else**
39:       $C' = C' \cup \{v_j\}$
40:     **end if**
41:   **end for**
42:   **if** *moved* is false and C' is non-empty **then**
43:     randomly select node v_i from C'
44:     $C = C \cup \{(v_i, C_{v_i})\}$
45:     $B = B \cup \{v_i\}$
46:     calculate $\mathcal{W}(C_{v_i})$ and add to $\mathcal{W}$;
47:   **end if**
48: **end while**
49: **return** C

V. EXPERIMENTS

In order to verify the validity and accuracy of the proposed algorithm, we extensively test it on several random networks and actual networks, and then compare them with the detection results given by the existing community detection algorithms on the same measure index. Meanwhile, the results of each experiment will be analyzed and discussed in detail.

A. RANDOM NETWORKS

In the unsigned undirected network, the computer generated network $LFR(n, k, k_{max}, c_{min}, c_{max}, \mu, t_1, t_2, on, om)$ is a network with scale-free and community generated by the computer program proposed by Lancichinetti *et al.* [51], in which

n is the number of nodes, k is the average degree of the nodes, k_{max} is the maximum degree of the nodes, c_{min} is the minimum number of members in the community, c_{max} is the maximum number of members in the community, μ is mixing parameter that control the fuzziness of the community, t_1 is minus exponent for the degree sequence, t_2 is minus exponent for the community size distribution, on is the number of overlapping nodes, and om is the number of memberships of the overlapping nodes. By setting the same parameters as in [52], four small networks and two large networks are generated as follows:

- *LFR_S1*: $n = 100$, $k = 20$, $k_{max} = 20$, $c_{min} = 10$, $c_{max} = 20$, $\mu = 0.3$, $t_1 = 2$, $t_2 = 1$, $on = 0$, $om = 0$.
- *LFR_S2*: $n = 100$, $k = 5$, $k_{max} = 15$, $c_{min} = 10$, $c_{max} = 50$, $\mu = 0.1$, $t_1 = 2$, $t_2 = 1$, $on = 0$, $om = 0$.
- *LFR_S3*: $n = 1000$, $k = 15$, $k_{max} = 50$, $c_{min} = 10$, $c_{max} = 20$, $\mu = 0.4$, $t_1 = 2$, $t_2 = 1$, $on = 0$, $om = 0$.
- *LFR_S4*: $n = 1000$, $k = 7$, $k_{max} = 25$, $c_{min} = 10$, $c_{max} = 300$, $\mu = 0.1$, $t_1 = 2$, $t_2 = 1$, $on = 0$, $om = 0$.
- *LFR_B1*: $n = 10000$, $k = 7$, $k_{max} = 20$, $c_{min} = 10$, $c_{max} = 100$, $\mu = 0.1$, $t_1 = 2$, $t_2 = 1$, $on = 0$, $om = 0$.
- *LFR_B2*: $n = 20000$, $k = 15$, $k_{max} = 50$, $c_{min} = 10$, $c_{max} = 50$, $\mu = 0.4$, $t_1 = 2$, $t_2 = 1$, $on = 0$, $om = 0$.

Meanwhile, the specific measure indexes — *recall*, *precision*, and *fscore* — are used to evaluate the quality of the local community, which can be written as:

$$recall = \frac{|C_{Found} \cap C_{True}|}{|C_{True}|}, \tag{11}$$

$$precision = \frac{|C_{Found} \cap C_{True}|}{|C_{Found}|}, \tag{12}$$

$$fscore = \frac{2 * precision * recall}{precision + recall}, \tag{13}$$

where C_{Found} represents the local community that given by the algorithm, C_{True} indicates the actual community.

From the above equations, we can see that *recall* expresses the ratio of nodes that are detected and belonging to the actual local community, *precision* denotes the proportion of the correct nodes that are detected, *fscore* is a combination of *recall* and *precision* since *recall* and *precision* are mutually influential, in which the *recall* is high, the *precision* is low, and vice versa.

So far, we have followed up to the local community detection algorithms that include DMF_M1, DMF_M2, and DMF_M proposed in [52], and the M algorithm proposed in [21]. To compare the performance of these algorithms, experiments are carried out by using the aforementioned generated networks, and the comparison of the experimental results given by each algorithm are shown in Table 1. Meanwhile, we also compare the NMI index given by the global community detection algorithms that include Louvain [15], SA [53], EO [54], LPA [27], EM [55], CDFR-U [30] (it is necessary to set a *delta* threshold to enlarge the gap between the central node of the community and the non-central node), and LIDGC. We can see that our algorithm guarantees good performance in most cases, especially in the first two networks, the performance of the algorithm reaches the highest value. Although our algorithm uses local information to detect the communities, we can discover communities hidden in the network by gradually increasing the number of communities and the number of community members through dynamic expansion. More importantly, the algorithm can be suitable to the dynamic network by modifying the conditions of the *while* loop since it is based on local information and dynamic expansion.

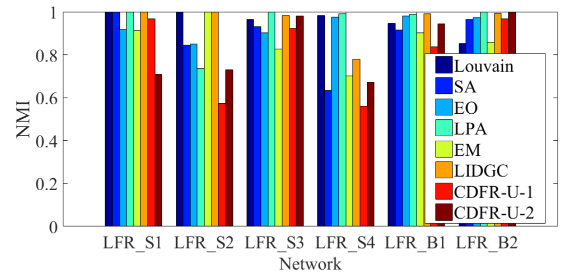

FIGURE 2. Comparison of NMI given by algorithms — Louvain, SA, EO, LPA, EM, CDFR-U (the *delta* threshold is set to 0.40 in the CDFR-U-1 and the *delta* threshold is set to 0.15 in the CDFR-U-2), and LIDGC on LFR datasets.

B. ACTUAL NETWORKS

1) THE ZACHARY KARATE CLUB NETWORK

The Zachary Karate Club network was established by Zachary [56] after investigating the social relationships of a Karate Club at an American university. The network is an unsigned undirected network and often used as a benchmark for testing the performance of community detection algorithm, which exist 34 nodes and 78 edges. The club split up into two sub clubs because of differences in fees and charges.

In Fig.3, we show the detection result of our algorithm in which the network is successfully divided into four communities, but not in accordance with the actual number of communities since our algorithm can reach the maximum value of the modularity. Although there are many classical algorithms for optimizing the modularity that can divide the network into four communities by using global information. However, the proposed algorithm only uses local information to detect the global communities, which gradually expands the number of communities and the number of community members through an initial node with a community label. First of all, one of the nodes is selected arbitrarily, and marks the selected node as the initial node with the community label. Therefore, a local community containing only one node will appear in the network. Next, we detect neighboring nodes of local communities, where these neighboring nodes do not have community label, and then add the neighboring nodes with maximum energy increment to local communities, so that it will be to achieve the goal that the gradual expansion of the number of community members through the initial nodes. In addition, it is naturally that the maximum energy increment may also be negative, which means that adding the

TABLE 1. Comparison of results given by algorithms — M, DMF_M1, DMF_M2, DMF_M, and LIDGC on LFR datasets .

Dataset	Index	M	DMF_M1	DMF_M2	DMF_M	LIDGC
LFR_S1	*recall*	0.7605±0.4064	0.9814±0.1307	0.8164±0.3608	0.9814±0.1307	**1.0000**±0.0000
	precision	0.7500±0.4017	0.9814±0.1312	0.7679±0.3645	0.9813±0.1316	**1.0000**±0.0000
	fscore	0.7549±0.4039	0.9814±0.1310	0.7859±0.3588	0.9813±0.1312	**1.0000**±0.0000
LFR_S2	*recall*	0.2044±0.2278	0.8453±0.3228	0.2055±0.2285	0.8453±0.3228	**1.0000**±0.0000
	precision	0.8270±0.2312	0.9333±0.1887	0.8272±0.2313	0.8595±0.2845	**1.0000**±0.0000
	fscore	0.2894±0.2273	0.8595±0.2844	0.2906±0.2285	0.8595±0.2845	**1.0000**±0.0000
LFR_S3	*recall*	0.4706±0.4719	0.9980±0.0438	0.4744±0.4700	0.9980±0.0438	**0.9981**±0.0035
	precision	0.5010±0.4254	0.9943±0.0459	0.5046±0.4224	0.9941±0.0464	**0.9983**±0.0031
	fscore	0.4722±0.4560	0.9961±0.0444	0.4757±0.4535	0.9960±0.0446	**0.9982**±0.0032
LFR_S4	*recall*	0.1459±0.2750	0.6920±0.4318	0.1470±0.2754	0.6925±0.4311	**0.7770**±0.2386
	precision	0.7949±0.2548	0.8271±0.3100	0.7940±0.2550	0.8275±0.3082	**0.9101**±0.0985
	fscore	0.1859±0.2718	0.7084±0.4125	0.1872±0.2725	0.7089±0.4116	**0.7950**±0.1405
LFR_B1	*recall*	0.5533±0.4456	0.9643±0.1734	0.5561±0.4450	0.9644±0.1731	**0.9650**±0.0577
	precision	0.8345±0.2879	0.9845±0.1094	0.8350±0.2862	0.9845±0.1091	**0.9951**±0.0092
	fscore	0.5880±0.4117	0.9672±0.1595	0.5901±0.4106	0.9673±0.1592	**0.9743**±0.0416
LFR_B2	*recall*	0.7112±0.4395	0.9992±0.0254	0.7220±0.4338	0.9992±0.0254	**0.9994**±0.0012
	precision	0.7078±0.4331	**1.0000**±0.0000	0.7108±0.4285	**1.0000**±0.0000	0.9988±0.0024
	fscore	0.7073±0.4378	**0.9993**±0.0211	0.7123±0.4320	**0.9993**±0.0211	0.9991±0.0018

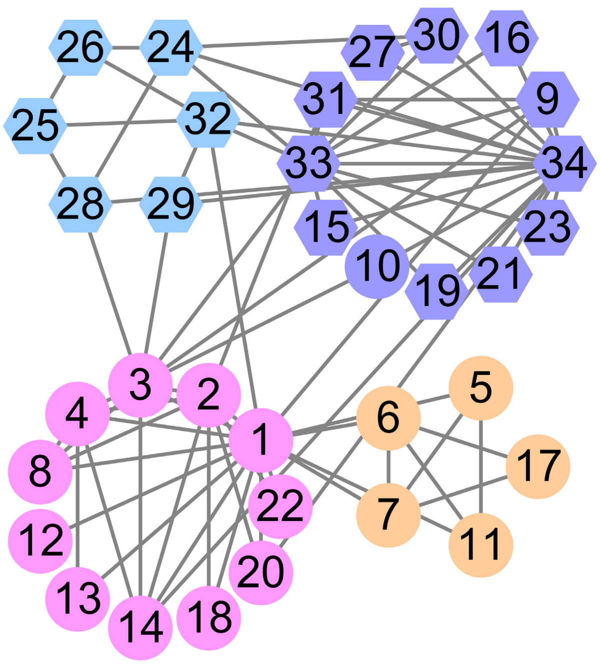

FIGURE 3. Detection results of the Zachary Karate Club network: the different detection communities are indicated by the different color, it can also distinguish the detected communities through different circles; the different actual communities are denoted by the different shapes, and the node set formed by the two circles in top constitutes an actual community, with the exception of node 10, because it forms another actual community with the node set formed by the two circles in bottom.

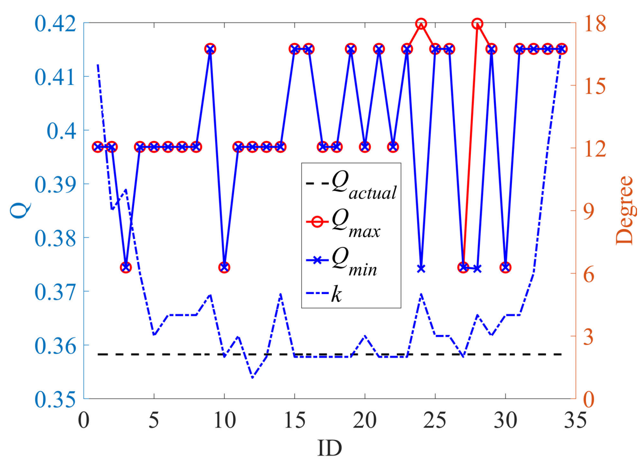

FIGURE 4. Comparison results of expansion methods: Q_{actual} represents the modularity value of actual results; Q_{max} denotes the modularity value of detection results by the maximum degree method; Q_{min} describes the modularity value of detection results by the minimum degree method; k indicates the degree of node; the horizontal axis "ID" means that the label of the initial node.

node into the local community will not increase the energy value of the whole network, but it will decrease. However, our goal is to maximize the energy of the network, so we will adopt different strategies to generate new local communities at this time. In the next paragraph, we will be discussing two strategies: the maximum degree and the minimum degree. According to the strategy used, we select a node as the new local community in the neighboring nodes of the local community. Hence, this step realizes the gradual expansion of the number of communities through the initial nodes.

Since our algorithm is based on dynamic expansion method, there will be some problems for us to generate a new local community and add nodes into the local community. Furthermore, the different generating sequence and adding sequence will affect the result of the following detection results. We chose two different ways of community expansion to study their impact on the performance of the algorithm: the maximum degree, the minimum degree. The experimental results are shown in the Fig.4. It can be found that, in any

TABLE 2. Execution processes of LIDGC on the Zachary Karate Club Network (The starting node is the 28th node) .

t	({Node}, Community ID)	(Node, Incremental Energy)	Operation
1	({28},0)	(34,0.050),(24,0.125),(3,0.077),(25,**0.167**)	$C_{v_i} \cup \{25\}$
2	({28,25},0)	(26,**0.083**),(32,0.015),(34,-0.076),(24,0.033),(3,-0.033)	$C_{v_i} \cup \{26\}$
3	({28,25,26},0)	(32,0.083),(24,**0.114**),(34,-0.125),(3,-0.074)	$C_{v_i} \cup \{24\}$
4	({28,25,26,24},0)	(33,-0.136),(34,-0.133),(30,-0.006),(32,**0.036**),(3,-0.114)	$C_{v_i} \cup \{32\}$
5	({28,25,26,24,32},0)	(33,-0.080),(34,-0.090),(30,-0.011),(3,-0.108),(1,-0.167),(29,**0.012**)	$C_{v_i} \cup \{29\}$
6	**({28,25,26,24,32,29},0)**	(33,-0.078),(**34**,-0.045),(30,-0.012),(3,-0.052),(1,-0.162)	generating $(\{34\},1)$
7	({28,25,26,24,32,29},0) ({34},1)	(33,-0.078),(30,-0.012),(3,-0.052),(1,-0.162) (33,0.036),(23,**0.056**),(14,0.048),(15,**0.056**),(16,**0.056**),(27,**0.056**), (19,**0.056**),(9,0.048),(30,0.050),(20,0.053),(31,0.050),(10,**0.056**),(21,**0.056**)	- $C_{v_i} \cup \{23,15,16,$ $27,19,10,21\}$
8	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21},1)	(33,-0.078),(30,-0.012),(3,-0.052),(1,-0.162) (33,**0.142**),(14,-0.006),(9,-0.006),(30,0.054),(20,0.016),(31,0.005),(3,-0.049)	- $C_{v_i} \cup \{33\}$
9	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33},1)	(30,-0.012),(3,-0.052),(1,-0.162) (3,-0.039),(9,0.021),(30,**0.083**),(31,0.035),(14,-0.021),(20,0.004)	- $C_{v_i} \cup \{30\}$
10	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30},1)	(3,-0.052),(1,-0.162) (3,-0.055),(9,0.013),(31,**0.029**),(14,-0.030),(20,-0.001)	- $C_{v_i} \cup \{31\}$
11	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31},1)	(3,-0.052),(1,-0.162) (3,-0.058),(9,**0.055**),(14,-0.032),(20,-0.003),(2,-0.082)	- $C_{v_i} \cup \{9\}$
12	({28,25,26,24,32,29},0) **({34,23,15,16,27,19,10,21,33,30,31,9},1)**	(3,-0.052),(1,-0.162) (3,-0.029),(14,-0.036),(20,-0.005),(**1**,-0.160),(2,-0.088)	- generating $(\{1\},2)$
13	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1},2)	(3,-0.052) (3,-0.029),(14,-0.036),(20,-0.005),(2,-0.088) (11,0.056),(22,0.059),(12,**0.063**),(13,0.059),(14,0.050),(18,0.059),(2,0.042), (3,0.040),(4,0.048),(5,0.056),(6,0.053),(7,0.053),(8,0.053),(20,0.056)	- - $C_{v_i} \cup \{12\}$
14	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1,12},2)	(3,-0.052) (3,-0.029),(14,-0.036),(20,-0.005),(2,-0.088) (11,0.049),(22,**0.055**),(13,**0.055**),(14,0.038),(18,**0.055**),(2,0.021), (3,0.018),(4,0.033),(5,0.049),(6,0.043),(7,0.043),(8,0.043),(20,0.049)	- - $C_{v_i} \cup \{22,13,18\}$
15	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1,12,22,13,18},2)	(3,-0.052) (3,-0.029),(14,-0.036),(20,-0.005),(2,-0.088) (2,**0.069**),(4,0.050),(11,0.028),(14,0.007),(3,-0.032),(5,0.028),(6,0.017), (7,0.017),(8,0.017),(20,0.028)	- - $C_{v_i} \cup \{2\}$
16	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1,12,22,13,18,2},2)	(3,-0.052) (3,-0.029),(14,-0.036),(20,-0.005) (4,**0.077**),(11,0.016),(14,0.041),(3,-0.007),(5,0.016),(6,0.006),(7,0.006), (8,0.053),(20,0.066)	- - $C_{v_i} \cup \{4\}$
17	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1,12,22,13,18,2,4},2)	(3,-0.052) (3,-0.029),(14,-0.036),(20,-0.005) (11,0.010),(14,0.076),(3,0.014),(5,0.010),(6,-0.002),(7,-0.002),(8,**0.091**), (20,0.057)	- - $C_{v_i} \cup \{8\}$
18	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1,12,22,13,18,2,4,8},2)	(3,-0.052) (3,-0.029),(14,-0.036),(20,-0.005) (11,0.003),(14,**0.068**),(3,0.037),(5,0.003),(6,-0.011),(7,-0.011),(20,0.052)	- - $C_{v_i} \cup \{14\}$
19	({28,25,26,24,32,29},0) ({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1,12,22,13,18,2,4,8,14},2)	(3,-0.052) (3,-0.029),(20,-0.005) (3,**0.067**),(11,-0.001),(5,-0.001),(6,-0.016),(7,-0.016),(20,0.046)	- - $C_{v_i} \cup \{3\}$
20	({34,23,15,16,27,19,10,21,33,30,31,9},1) ({1,12,22,13,18,2,4,8,14,3},2)	(20,-0.005) (11,-0.004),(5,-0.004),(6,-0.019),(7,-0.019),(20,**0.038**)	- $C_{v_i} \cup \{20\}$
21	**({1,12,22,13,18,2,4,8,14,3,20},2)**	(11,-0.006),(5,-0.006),(**6**,-0.022),(7,-0.022)	generating $(\{6\},3)$
22	({1,12,22,13,18,2,4,8,14,3,20},2) ({6},3)	(11,-0.006),(5,-0.006),(7,-0.022) (11,0.167),(17,**0.200**),(7,0.143)	- $C_{v_i} \cup \{17\}$
23	({1,12,22,13,18,2,4,8,14,3,20},2) ({6,17},3)	(11,-0.006),(5,-0.006),(7,-0.022) (7,**0.229**),(11,0.086)	- $C_{v_i} \cup \{7\}$
24	({1,12,22,13,18,2,4,8,14,3,20},2) ({6,17,7},3)	(11,-0.006),(5,-0.006) (11,**0.016**),(5,**0.016**)	- $C_{v_i} \cup \{11,5\}$
25	**({6,17,7,11,5},3)**	-	-

case, modularity value of the partitioning scheme given by our algorithm is higher than the modularity value of the actual scheme. As we can see that in [57] the maximum value of modularity of this network is 0.4197, the modularity value is around 0.4 in [9], the modularity value is 0.416 in [16], while the peak value of modularity given by our method is 0.4197 and the detail of the expanding processes is shown in Table 2. In addition, the influence of the two community expansion methods on the experimental results is negligible because only two nodes are different. This demonstrates that our algorithm is not sensitive on the choice of nodes in this network, and it can give a reasonable solution, no matter which node it starts from.

One of the complete procedures of detection is shown in Table 2. The first column indicates the number of times of the *while* loop is executed; the second column represents the local community that has been detected, where the thickening indicates that the community has been completed, and lists the numbers of the community members and the label of community, respectively; the third column describes the neighboring nodes of the local communities and the energy increment obtained by adding nodes to the local community, and the bold value explains the maximum energy increment; the fourth column shows the operations performed by each local community.

2) THE DOLPHIN SOCIAL NETWORK

The Dolphin social network was established by Lusseau and Newman [58] after observing the frequency of contact between dolphins. The network is an unsigned undirected network in which consists of 62 nodes and 159 edges, and there are four different dolphin families.

The partition result of our algorithm is shown in Fig.5. Four different communities are successfully detected and only one node is partitioned incorrectly, which is remarked in yellow. Although our algorithm divides the network into four communities, it can be found that there are still many connected edges between different communities. In fact, the network can be divided into two communities which have shown in [9], so that the number of connecting edges between communities are reduced. However, at this time, the energy value or modularity value of the entire network cannot reach a higher value, that is, dividing the network into four communities can obtain the higher value in energy or modularity.

3) THE SLOVENE PARLIAMENTARY PARTIES NETWORK

The Slovene Parliamentary Parties Network depicts the relationship between 10 alliances. [59] The weight range of edges is from −3 to 3 means that the network is a signed undirected network, which in turn describes very dissimilar, relatively dissimilar, dissimilar, neither dissimilar nor similar, similar, relatively similar, and very similar.

The division result of our algorithm is shown in Fig.6, which the network is divided into two different communities, and the scheme is consistent with the actual communities.

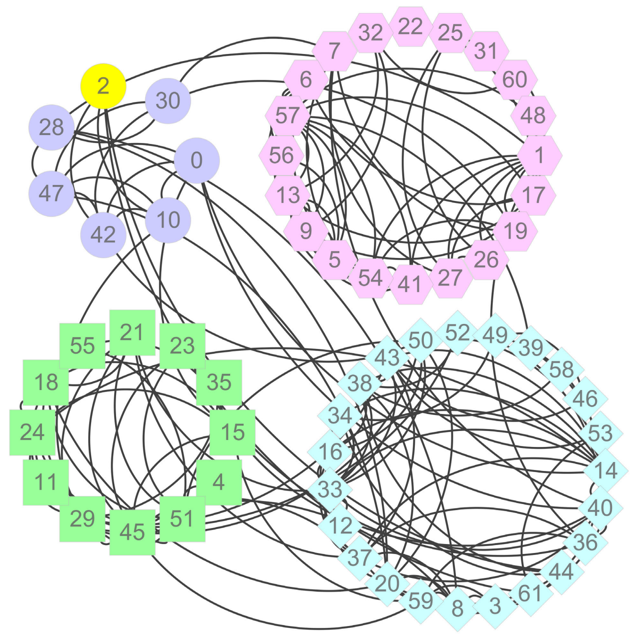

FIGURE 5. Detection results of the Dolphin social network: the node 2 is divided into the wrong community in the top left, and is marked with yellow; the different detection communities are indicated by the different color, it can also distinguish the detected communities through different circles; the different actual communities are denoted by the different shapes, that is, each circle represents an actual community, and the node 2 should belong to the community in the lower right.

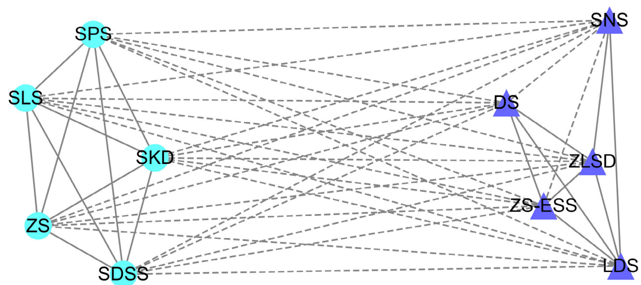

FIGURE 6. Detection results of the Slovene Parliamentary Parties network: the solid line represents a positive edge; the dashed line indicates a negative edge; the different color and shapes of node are used to distinguish different communities.

As expected, the edges inside the community are mostly linked by positive weights, while the edges between the communities are mainly linked by negative weights.

4) THE GAHUKU-GAMA SUBTRIBES NETWORK

The Gahuku-Gama Subtribes Network describes the relationship between tribes in the Eastern Central Highlands of New Guinea, which was created based on Read's investigation. [60] There are 16 nodes and 58 edges in the network, including 28 positive edges and 28 negative edges, which shows that the network is a signed undirected network.

As shown in the Fig.7, the network is divided into three different communities, and the result is also consistent with the actual result. Comparing the results of experiment given

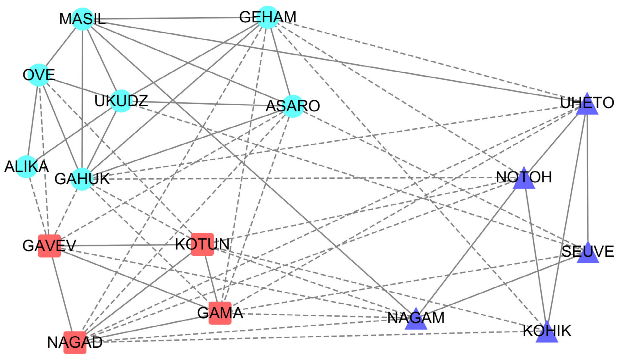

FIGURE 7. Detection results of the Gahuku-Gama Subtribes network: the solid line represents a positive edge; the dashed line indicates a negative edge; the different color and shapes of node are used to distinguish different communities.

TABLE 3. Detection results of the LIDGC algorithm on large-scale networks .

Network	N	E	C_{real}	C_{find}	Q
Amazon	0.33M	0.92M	75,149	32,077	0.707
DBLP	0.32M	1.05M	13,477	28,310	0.587
Youtube	1.13M	2.99M	8,385	15,961	0.631

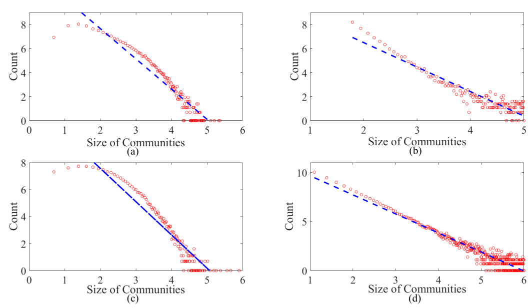

FIGURE 8. (a) The distribution of detection community sizes in the DBLP network. (b) The distribution of actual community sizes in the DBLP network. (c) The distribution of detection community sizes in the Amazon network. (d) The distribution of real community sizes in the Amazon network. The x-axis is the size of communities, and the y-axis is the count of the size.

by other algorithms, we found an interesting partition scheme in [36] in which the network will be divided into four communities in order to achieve balanced network topology, that is, {NAGAM, SEUVE} will be independent and become a new community.

5) THE LARGE-SCALE NETWORKS

To illustrate the performance of the proposed algorithm in large-scale networks, the Amazon network, the DBLP network and the Youtube network is selected and used to detect communities, which these networks are unsigned undirected network obtained in [61], meanwhile, the properties and the detection results of these networks are shown in Table 3. In Table 3, the N is the number of nodes, the E represents the number of edges, the C_{real} describes the number of the actual communities in the network, C_{find} indicates the number of the detection communities given by the LIDGC algorithm, and the Q is the modularity value of the detection result.

In addition, an important phenomenon is presented in Fig.8 where the count distribution of community sizes is in line with the scale-free distribution. As aforementioned, the number of community members is gradually expanded, which explains that a new local community will be generated when the current local community reaches a ''saturation state'' (the local communities cannot absorb nodes to increase energy). All of these explain that the real world contains a large number of small-scale communities, and these small communities make them similar to its neighboring communities as a whole in areas such as local community mining and recommendation [62], etc.

VI. CONCLUSION

In summary, we have proposed a novel community detection algorithm by using local information, meanwhile a more generalized signed modularity function and the signed local modularity function is proposed to evaluate the quality of the local community by using the local information based on the existing modularity function. Using the proposed modularity function, a dynamic expansion algorithm based on local information for global community detection is proposed. The algorithm gradually detects the community structure hidden inside the network by dynamically expanding the community size and adding nodes through an initial node that with community label. The results given by the proposed algorithm in different networks show that it can give a reasonable result in both unsigned and signed networks. The current results are also conductive for us to perform the personalized recommendation in electronic business website.

REFERENCES

[1] C. Pizzuti, ''A multiobjective genetic algorithm to find communities in complex networks,'' *IEEE Trans. Evol. Comput.*, vol. 16, no. 3, pp. 418–430, Jun. 2012.

[2] M. E. J. Newman, ''Fast algorithm for detecting community structure in networks,'' *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, p. 066133, 2004.

[3] M. S. Khan, A. W. A. Wahab, T. Herawan, G. Mujtaba, S. Danjuma, and M. A. Al-Garadi, ''Virtual community detection through the association between prime nodes in online social networks and its application to ranking algorithms,'' *IEEE Access*, vol. 4, pp. 9614–9624, 2016.

[4] S. Datta, P. Basuchowdhuri, S. Acharya, and S. Majumder, ''The habits of highly effective researchers: An empirical study,'' *IEEE Trans. Big Data*, vol. 3, no. 1, pp. 3–17, Mar. 2017.

[5] J. D. West, I. Wesley-Smith, and C. T. Bergstrom, ''A recommendation system based on hierarchical clustering of an article-level citation network,'' *IEEE Trans. Big Data*, vol. 2, no. 2, pp. 113–123, Jun. 2016.

[6] F. N. Silva, D. R. Amancio, M. Bardosova, L. F. da Costa, and O. N. Oliveira, ''Using network science and text analytics to produce surveys in a scientific topic,'' *J. Informetrics*, vol. 10, no. 2, pp. 487–502, May 2016.

[7] L. Wang and X. Li, ''Spatial epidemiology of networked metapopulation: An overview,'' *Sci. Bull.*, vol. 59, no. 28, pp. 3511–3522, Oct. 2014.

[8] H.-J. Li *et al.*, ''Evolving the attribute flow for dynamical clustering in signed networks,'' *Chaos Solitons Fractals*, vol. 110, pp. 20–27, May 2018.

[9] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, p. 026113, Feb. 2004.

[10] C. Li, L. Wang, S. Sun, and C. Xia, "Identification of influential spreaders based on classified neighbors in real-world complex networks," *Appl. Math. Comput.*, vol. 320, pp. 512–523, Mar. 2018.

[11] J. Guare, *Six Degrees of Separation: A Play*. New York, NY, USA: Vintage, 1990.

[12] F. Dantas-Torres, "Climate change, biodiversity, ticks and tick-borne diseases: The butterfly effect," *Int. J. Parasitol. Parasites Wildlife*, vol. 4, no. 3, pp. 452–461, Aug. 2015.

[13] J. Li, J. Wang, S. Sun, and C. Xia, "Cascading crashes induced by the individual heterogeneity in complex networks," *Appl. Math. Comput.*, vol. 323, pp. 182–192, Apr. 2018.

[14] C. Li and F. Xiong, "Social recommendation with multiple influence from direct user interactions," *IEEE Access*, vol. 5, pp. 16288–16296, Aug. 2017.

[15] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech.*, vol. 2008, no. 10, pp. 155–168, Apr. 2008.

[16] H.-J. Li, Z. Bu, A. Li, Z. Liu, and Y. Shi, "Fast and accurate mining the community structure: Integrating center locating and membership optimization," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 9, pp. 2349–2362, May 2016.

[17] S. Gómez, P. Jensen, and A. Arenas, "Analysis of community structure in networks of correlated data," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 2, p. 016114, Aug. 2009.

[18] V. A. Traag and J. Bruggeman, "Community detection in networks with positive and negative links," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 2, p. 036115, Sep. 2009.

[19] X. Wen *et al.*, "A maximal clique based multiobjective evolutionary algorithm for overlapping community detection," *IEEE Trans. Evol. Comput.*, vol. 21, no. 3, pp. 363–377, Jun. 2017.

[20] A. Clauset, "Finding local community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 72, no. 2, p. 026132, Sep. 2005.

[21] F. Luo, J. Z. Wang, and E. Promislow, "Exploring local community structures in large networks," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell.*, Hong Kong, China, Dec. 2006, pp. 233–239.

[22] J. Eustace, X. Wang, and Y. Cui, "Community detection using local neighborhood in complex networks," *Phys. A, Stat. Mech. Appl.*, vol. 436, pp. 665–677, May 2015.

[23] J. Eustace, X. Wang, and Y. Cui, "Overlapping community detection using neighborhood ratio matrix," *Phys. A, Stat. Mech. Appl.*, vol. 421, pp. 510–521, Mar. 2015.

[24] X. Wang, G. Liu, and J. Li, "Overlapping community detection based on structural centrality in complex networks," *IEEE Access*, vol. 5, pp. 25258–25269, Nov. 2017.

[25] D. R. Amancio, "Comparing the topological properties of real and artificially generated scientific manuscripts," *Scientometrics*, vol. 105, no. 3, pp. 1763–1779, Dec. 2015.

[26] G. S. Thakur, R. Tiwari, M. T. Thai, S.-S. Chen, and A. W. M. Dress, "Detection of local community structures in complex dynamic networks with random walks," *IET Syst. Biol.*, vol. 3, no. 4, pp. 266–278, Jul. 2009.

[27] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 76, no. 3, p. 036106, 2007.

[28] E. L. Martelot and C. Hankin, "Multi-scale community detection using stability optimisation within greedy algorithms," *Eprint Arxiv*, vol. 9, no. 3, pp. 323–348, Jan. 2012.

[29] A. Mahmood and M. Small, "Subspace based network community detection using sparse linear coding," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 801–812, Mar. 2016.

[30] W. Luo, Z. Yan, C. Bu, and D. Zhang, "Community detection by fuzzy relations," *IEEE Trans. Emerg. Topics Comput.*, to be published, doi: 10.1109/TETC.2017.2751101.

[31] S. A. Moosavi, M. Jalali, N. Misaghian, S. Shamshirband, and M. H. Anisi, "Community detection in social networks using user frequent pattern mining," *Knowl. Inf. Syst.*, vol. 51, no. 1, pp. 159–186, Apr. 2017.

[32] T. He, L. cai, T. Meng, L. Chen, Z. Deng, and Z. Cao, "Parallel community detection based on distance dynamics for large-scale network," *IEEE Access*, vol. 6, pp. 42775–42789, Jul. 2018.

[33] X. Deng, J. Zhai, T. Lv, and L. Yin, "Efficient vector influence clustering coefficient based directed community detection method," *IEEE Access*, vol. 5, pp. 17106–17116, Aug. 2017.

[34] G. J. Qi, C. C. Aggarwal, and T. Huang, "Community detection with edge content in social media networks," in *Proc. IEEE 28th Int. Conf. Data Eng.*, Apr. 2012, pp. 534–545.

[35] B. Yang, W. Cheung, and J. Liu, "Community mining from signed social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 10, pp. 1333–1348, Oct. 2007.

[36] Q. Cai, M. Gong, S. Ruan, Q. Miao, and H. Du, "Network structural balance based on evolutionary multiobjective optimization: A two-step approach," *IEEE Trans. Evol. Comput.*, vol. 19, no. 6, pp. 903–916, Dec. 2015.

[37] C. Liu, J. Liu, and Z. Jiang, "A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2274–2287, Dec. 2014.

[38] J. Huang, H. Sun, Y. Liu, Q. Song, and T. Weninger, "Towards online multiresolution community detection in large-scale networks," *PLoS ONE*, vol. 6, no. 8, p. e23829, Aug. 2011.

[39] J. Q. Jiang, "Stochastic block model and exploratory analysis in signed networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 91, no. 6, p. 062805, Jun. 2015.

[40] R. Harakawa, T. Ogawa, and M. Haseyama, "Extracting hierarchical structure of web video groups based on sentiment-aware signed network analysis," *IEEE Access*, vol. 5, pp. 16963–16973, Aug. 2017.

[41] Y. Li, J. Liu, and C. Liu, "A comparative analysis of evolutionary and memetic algorithms for community detection from signed social networks," *Soft Comput.*, vol. 18, no. 2, pp. 329–348, Feb. 2014.

[42] S. Liu and S. Wang, "Trajectory community discovery and recommendation by multi-source diffusion modeling," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 4, pp. 898–911, Apr. 2017.

[43] U. Athira and S. M. Thampi, "Linguistic feature based filtering mechanism for recommending posts in a social networking group," *IEEE Access*, vol. 6, pp. 4470–4484, Jan. 2018.

[44] R. Parimi and D. Caragea, "Community detection on large graph datasets for recommender systems," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Dec. 2014, pp. 589–596.

[45] D. Rafailidis, "Modeling trust and distrust information in recommender systems via joint matrix factorization with signed graphs," in *Proc. 31st Annu. ACM Symp. Appl. Comput.*, Pisa, Italy, Apr. 2016, pp. 1060–1065.

[46] H. H. Qiu, Y. Liu, Z. J. Zhang, and G. X. Luo, "An improved collaborative filtering recommendation algorithm for microblog based on community detection," in *Proc. 10th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Kitakyushu, Japan, vol. 17, Aug. 2014, pp. 876–879, no. 2.

[47] D. R. Amancio, "Probing the topological properties of complex networks modeling short written texts," *PLoS ONE*, vol. 10, no. 2, p. e0118394, Dec. 2015.

[48] L. Danon, J. Duch, and A. Díaz-Guilera, and A. Arenas, "Comparing community structure identification," *J. Stat. Mech.*, vol. 2005, no. 9, p. 09008, Sep. 2005.

[49] F. Luo, Y. Yang, C.-F. Chen, R. Chang, J. Zhou, and R. H. Scheuermann, "Modular organization of protein interaction networks," *Bioinformatics*, vol. 23, no. 2, pp. 207–214, Apr. 2007.

[50] D. R. Amancio, O. N. Oliveira, and L. F. da Costa, "Robustness of community structure to node removal," *J. Stat. Mech.*, vol. 2015, no. 3, p. P03003, Mar. 2015.

[51] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 78, no. 4, p. 046110, Nov. 2008.

[52] W. Luo, D. Zhang, H. Jiang, L. Ni, and Y. Hu, "Local community detection with the dynamic membership function," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3136–3150, Mar. 2018, doi: 10.1109/TFUZZ.2018.2812148.

[53] R. Guimerà and L. A. N. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, Feb. 2005.

[54] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 72, p. 027104, Aug. 2005.

[55] M. E. J. Newman, "Mixing patterns in networks: Empirical results and models," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 67, no. 2, p. 026126, Sep. 2002.

[56] W. W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol. Res.*, vol. 33, no. 4, pp. 452–473, Nov. 1976.

[57] G. Agarwal and D. Kempe, "Modularity-maximizing graph communities via mathematical programming," *Eur. Phys. J. B*, vol. 66, no. 3, pp. 409–418, Dec. 2008.

[58] D. Lusseau and M. E. J. Newman, "Identifying the role that animals play in their social networks," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 271, pp. 477–481, Dec. 2005.

[59] S. Kropivnik and A. Mrvar, *An Analysis of the Slovene Parliamentary Parties Network*. Ljubljana, Slovenia: Metodološki zvezki, Jan. 1996.

[60] K. E. Read, "Cultures of the central highlands, New Guinea," *Southwestern J. Anthropol.*, vol. 10, no. 1, pp. 1–43, Apr. 1954.

[61] J. Leskovec and A. Krevl. (Jun. 2014). *SNAP Datasets: Stanford Large Network Dataset Collection*. [Online]. Available: http://snap.stanford.edu/data

[62] Y. Huang, Y. Tang, C. Li, Z. Wu, and H. Dong, "A method for latent-friendship recommendation based on community detection in social network," in *Proc. 12th Web Inf. Syst. Appl. Conf. (WISA)*, Jinan, China, Feb. 2016, pp. 3–8, doi: 10.1109/WISA.2015.16.

Authors' photographs and biographies not available at the time of publication.