

Received November 28, 2020, accepted December 11, 2020, date of publication December 15, 2020, date of current version February 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3045085

TNS-LPA: An Improved Label Propagation Algorithm for Community Detection Based on Two-Level Neighbourhood Similarity

GUIQIONG XU¹, JIAWEN GUO¹, AND PINGLE YANG^{1,2,3}

¹Department of Information Management, School of Management, Shanghai University, Shanghai 200444, China

²Business School, University of Shanghai for Science and Technology, Shanghai 200093, China

³School of Electrical and Information Engineering, Jiangsu University of Science and Technology, Zhangjiagang 215600, China

Corresponding author: Pingle Yang (plyang@just.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 11871328 and in part by the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province under Grant 2020SJA2346.

ABSTRACT Community detection can not only help people understand organizational structure and function of complex networks, but also attributes to many potential applications including targeted advertising and customer relationship management. Due to the low time complexity, the label propagation algorithm is widely used, but there is still room to improve the community quality and the detection stability. Inspired by resource allocation and local path similarity, we first give a new two-level neighbourhood similarity measure called TNS, and on this basis we propose an improved label propagation algorithm for community detection. In this new algorithm, the minimum distance and local centrality index are considered to select the initial community centers, to ensure that they are both important and far away from each other. In the process of forming initial community, we employ the new similarity measure and an optimization strategy of asynchronously updating labels according to node importance. To further improve the accuracy of community division, we introduce the label influence based on the new similarity measure to further optimize the community division of networks. The experimental results on both the artificial network and ten real-world networks show that our proposed algorithm has better comprehensive performance than several existing algorithms in terms of modularity, normalized mutual information and adjusted rand index.


INDEX TERMS Complex networks, community detection, label propagation, community kernel, local similarity measure.

I. INTRODUCTION

Since Newman's original work [1], during the past two decades, community detection in complex networks has attracted considerable attention [2]–[8]. Mining the community structure in social networks can help us analyze the network topology and function, so as to understand, control and predict social networks. Most of social networks have obvious community structures. Community detection based on social media data can be employed in various applications, including epidemic control, crisis response, and predictive policing [9]–[13]. In Internet finance, community detection can be used in targeted advertising, customer relationship management and fraud detection [14], [15]. In biological

networks, community detection may contribute to recognize different functional modules of proteins [16]. For academia and scientometrics, community detection can use for recommendation system and data dimension reduction in pattern recognition [17], [18].

During the past years, scholars have expended a great deal of efforts studying community properties of networks and developed various community detection algorithms [19]–[22]. The GN algorithm [23] was firstly proposed by Givan and Newman, whose key idea lies at iterative removal of edges with high betweenness centrality. Due to its high computation complexity, Newman [24] shortly afterwards put forward a fast modularity maximization (FMM) algorithm, which optimized modularity through iteratively community merging and updating. Community detection is essentially the clustering of nodes in the network, thus various

The associate editor coordinating the review of this manuscript and approving it for publication was Yichuan Jiang .

clustering algorithms have been widely used to community division in networks. Clauset, Newman and Moore [25] proposed a hierarchical agglomerative algorithm to detect meaningful communities by employing greedy optimization strategy so that the proposed algorithm can be suitable for large scale networks, afterwards this algorithm is called CNM algorithm. Blondel *et al.* [26] proposed an efficient and heuristic algorithm called BGLL for finding high modularity partitions of large networks, which can also be used to weighted networks. Due to the limitations of the classical K-means algorithm, Rodriguez and Laio [27] proposed the density-based clustering algorithm for community division, which supposed that cluster centers are surrounded by neighbors with lower local density and that they are at a relatively large distance from any points with a higher local density. Recently, Bai *et al.* [28] generalized the K-prototypes-type clustering to community detection and proposed a new algorithm called ISCD+, which considered both the local importance of a node in a community and the “importance concentration” of the node in all the communities. The ISCD+ algorithm needs predefined parameters such as the number of communities. Considering the binary and triadic relations among vertices, Zhang *et al.* [29] provided a spectral k-way partition algorithm for discovering community structures, which has better performance than the normalized-cut graph partition algorithm.

The label propagation algorithm (LPA), firstly proposed by Raghavan *et al.*, is a fast and unsupervised learning algorithm for community detection [30], whose basic idea is to make use of the information of labelled nodes to predict the labels of the remained nodes. The best advantage of LPA is the near linear time complexity and thus it has attracted a great deal of attention of scholars. Barber and Clark reformulated the LPA as its equivalent optimization problem, and put forward the LPAm [31] algorithm by modifying the objective function, which can be applicable for both bipartite and unipartite networks. Based on this work, Liu *et al.* presented the LPAm+ [32] to avoid the local optimum problem using the multi-step greedy aggregative strategy. On the basis of the LPA algorithm, Liu *et al.* proposed a novel evolutionary clustering approach, which is adaptable to detect overlapping and non-overlapping communities in dynamic networks [33]. To overcome the community annexation problem, Li *et al.* [34] established the Stepping-LPA-S algorithm, a new evaluation function is introduced to guide merging the small communities in the process of optimizing community division. Ding *et al.* [35] proposed a novel community detection algorithm called DCN, where potential community centers are chosen by means of Chebyshev inequality and label propagation makes good use of neighbors of the node and adopts the multiple strategy of label propagation. Very recently, Wang *et al.* proposed a novel label propagation algorithm based on node importance [36], where the important nodes are determined by integrating neighborhood Jaccard distance, K-shell value and signal propagation amount.

However, there is still a lot of room to improve the original LPA algorithm as well as its various extensions, such as stability and accuracy. For this purpose, we present an improved label propagation algorithm for community detection based on two-level neighbourhood similarity measure. The rest of this article is organized as follows. In Section 2, based on resource allocation and local paths of networks, we propose a new local similarity measure between node pairs. In Section 3, an improved label propagation algorithm called TNS-LPA is proposed to optimize the stability and accuracy of community detection. Section 4 is devoted to the data sets and evaluation metrics. In Section 5, the performance of the proposed algorithm is tested on both artificial network and real-world networks with different scales. Finally, in Section 6, we discuss and summarize our results.

II. TNS: A NEW SIMILARITY MEASURE

For an unweighted and undirected network $G = (V, E)$, V represents the set of nodes, and E denotes the set of edges, where $|V| = N$, $|E| = M$. $A = (a_{ij})_{N \times N}$ stands for the adjacency matrix, $a_{ij} = 1$ if there exists a link between node v_i and node v_j and $a_{ij} = 0$ otherwise. $k(i) = \sum_{v_j \in V \setminus v_i} a_{ij}$ is the degree of node v_i .

According to the definition of community, the greater the similarity between two nodes, the more likely these two nodes belong to the same community. Therefore, how to define the similarity of nodes becomes a very essential and important issue. Many scholars have made great contributions to it [37]–[42]. From the viewpoint of network topology, Liben-Nowell and Kleinberg [38] proposed a number of node-based and path-based similarity measures. Among several local indices which only consider the neighbor information, the Adamic-Adar index [41] has the best performance. In Ref. [42], Zhou *et al.* compared nine known neighborhood-based similarity measures in link prediction on six different networks, which include Salton, Jaccard, Sørensen, Hub promoted index and so on.

Based on these classical indices, Zhou and Lü [42] proposed two new indices, namely, resource allocation (RA) index and local path (LP) index. The similarity S_{ij} between node v_i and v_j based on RA index can be defined as

$$S_{ij} = \sum_{v_z \in \Gamma_1(i) \cap \Gamma_1(j)} \frac{1}{k(z)}, \quad (1)$$

where $\Gamma_1(i)$ and $\Gamma_1(j)$ represent the first order neighbor sets of node v_i and node v_j , and $k(z)$ is the degree of node v_z . RA is found to be efficient in community detections and applicable to weighted networks [43]. From the viewpoint of resource allocation, Li *et al.* established an improved LPA algorithm called “Stepping-LPA-S [34]. Besides, the LP index is given by [42]

$$S = A^2 + \rho A^3, \quad 0 < \rho < 1, \quad (2)$$

which only considers the two-step and three-step paths in networks, and ρ is a tunable parameter.

Both RA and LP are found to have better prediction ability than the known nine local indices including the Adamic-Adar index [42]. Moreover, the LP has comparable prediction ability with some global measures, such as Katz index [44], [45], especially for the networks with small average shortest path. The local index only depends on the local information of the network, and its computational cost is far less than those indices based on the global information especially for the sparse and large-scale networks. Christakis *et al.* argued that the behavior of a node is highly related to its first-order, second-order and up to the third-order neighbors for most of social networks, namely, the so-called three degree of separation [46]. Recently, Kovács *et al.* defined a new similarity on the basis of two-step and three-step path and established a new link prediction algorithm, which outperforms greatly than the existing link prediction algorithms in protein interaction networks [47].

Inspired by the above work, this article puts forward a new similarity measure, called two-level neighborhood similarity (TNS). The similarity index S_{ij} between node v_i and v_j is defined as

$$S_{ij} = \sum_{v_z \in \Gamma_1(i) \cap \Gamma_1(j)} \frac{1}{k(z)} + \sum_{v_x \in \Gamma_1(i), v_y \in \Gamma_1(j)} \frac{a_{xy}}{\sqrt{k(x)k(y)}}, \quad (3)$$

where $\Gamma_1(i)$ and $\Gamma_1(j)$ represent the first order neighbor sets of node v_i and node v_j , and $k(x)$, $k(y)$, $k(z)$ represent the degrees of nodes v_x , v_y and v_z . The new index consists of two parts: the former evaluates the contribution of network paths of length two, and the latter measures the contribution of network paths of length three.

This new similarity measure is illustrated by the karate network shown in Fig. 1(a). The components related to nodes 1 and 6 are extracted and shown in Fig. 1(b). Obviously, there are two two-step paths and three three-step paths between node 1 and node 6, thus its similarity is calculated as follows,

$$S_{16} = \left(\frac{1}{3} + \frac{1}{4} \right) + \left(\frac{1}{\sqrt{3 * 4}} + \frac{1}{\sqrt{3 * 3}} + \frac{1}{\sqrt{4 * 2}} \right) = 1.5589.$$

III. PROPOSED ALGORITHM

In this article, we propose an improved label propagation algorithm called TNS-LPA, which is consisted of three phases. First, the initial centers of communities are chosen based on the local centrality and minimum distance. Second, we generate the prototype of community partition based on the proposed TNS index. Finally, the community partition is further optimized by combining nodes' label influence with traditional label propagation. The process of community detection model is schematically shown in Fig. 2.

A. SELECTING INITIAL COMMUNITY KERNELS

The limitation of traditional LPA is that node update order is random and thus it will produce unstable partition results

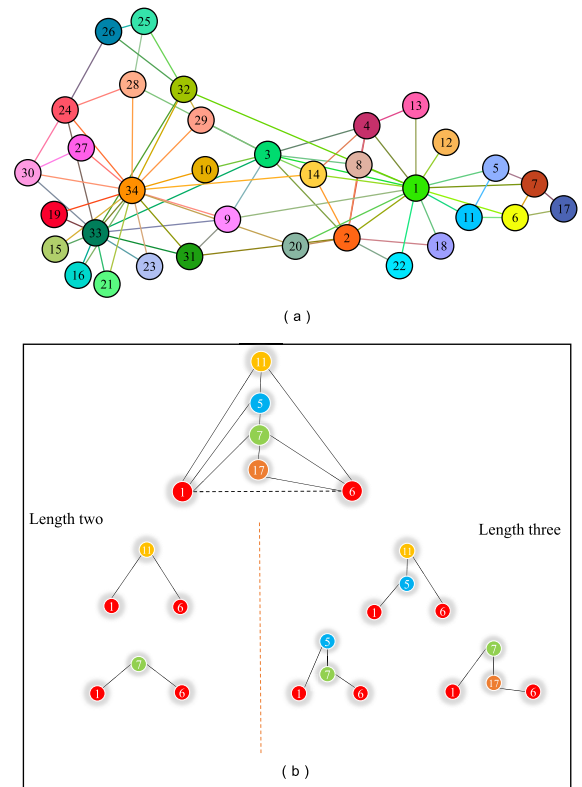


FIGURE 1. (a) Karate network. (b) The paths of length two and length three between node 1 and node 6.

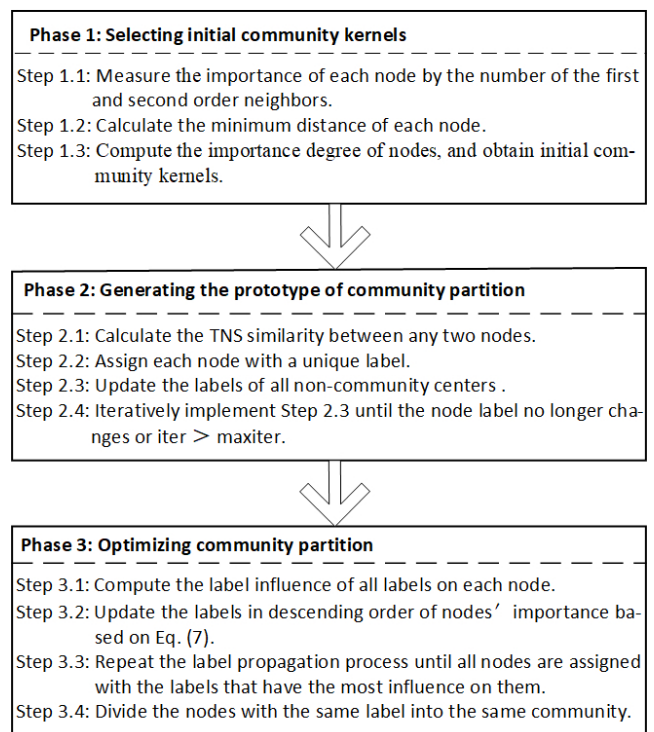


FIGURE 2. The framework of the TNS-LPA algorithm.

of networks. Inspired by DPC [27] and FCC [48], we propose a new strategy to select the initial community kernels,

which takes into account the distance and importance simultaneously [49]. It is well known to us that if centers are dispersed as much as possible, they may have less overlapping effects and can get more stable community structure. So, we try to find a relatively large distance among community centers. Rodriguez and Laio proposed an innovative index called minimum distance for measuring the proximity of nodes to more important nodes [27]. The importance of a node is measured by the number of the first and second order neighbors.

The importance degree of node v_i is defined as

$$IP(i) = C_{Nei}(i) * \min_{v_j: C_{Nei}(j) > C_{Nei}(i)} d_{ij}, \quad (4)$$

where d_{ij} is the shortest path length between node v_i and v_j . $C_{Nei}(i)$ denotes the local centrality of node v_i , which is defined as the number of the nearest neighbors and the next nearest neighbors. $C_{Nei}(i)$ can be calculated as

$$\begin{aligned} C_{Nei}(i) &= |\Gamma_1(i)| + |\Gamma_2(i)|, \\ \Gamma_1(i) &= \{v_j | v_j \in V, (v_i, v_j) \in E\}, \\ \Gamma_2(i) &= \{v_j | v_j \in \Gamma_1(u), v_u \in \Gamma_1(i), v_j \neq v_i, \\ & \quad v_j \cap \Gamma_1(i) = \Phi \} \end{aligned} \quad (5)$$

where $\Gamma_1(i)$ and $\Gamma_2(i)$ represent the first and second order neighbor of node v_i . The larger $IP(i)$ is, the higher probability that it is the center of some community. Once the importance degree for each node is obtained, one may sort IP in descending order and the top- s nodes are chosen as initial centers of communities.

The number of initial communities can get from decision graph [27]. The transverse axis and longitudinal axis denote nodes' importance and minimum distance respectively. The kernels are at top right of the decision graph and far away from other nodes, which have both relatively high importance and large distance.

B. GENERATING THE PROTOTYPE OF COMMUNITY PARTITION

The goal of this stage is to generate the initial community partition. We know that the original LPA updates the label according to the majority of neighbors. Li *et al.* [34] uses the maximum similarity instead of the majority number of neighbors to update the label, which can output rough community partition. Motivated by the idea, we update the labels based on the proposed TNS similarity,

$$l_i = \arg \max_{v_j \in \Gamma_1(i)} S_{ij} \quad (6)$$

until all nodes have the same labels with the most similar neighbors. In Eq. (6), $\Gamma_1(i)$ is the first order neighbor set of node v_i , and S_{ij} is given by Eq. (3). When the label propagation process is completed, nodes with the same label are divided into the same community. Then, we can get the initial community division. Note that the labels of all initial community centers remain unchanged. In order to form a relatively stable community division and avoid the turbulence

phenomenon during the label propagation process, the update order of nodes' labels is determined by Eq. (4).

The initial prototype of community partition for the Karate network is shown in Fig. 3. It can be clearly seen that the initial division result seems unreasonable and needs to be improved.

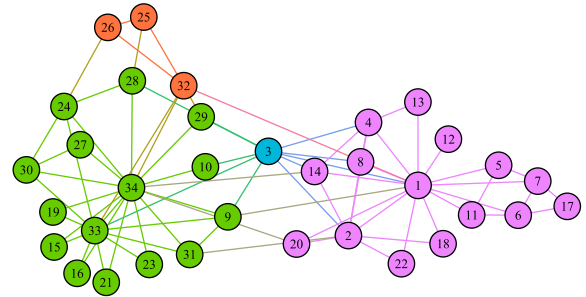


FIGURE 3. The initial community partition of Karate network.

C. OPTIMIZING COMMUNITY PARTITION

So far, our algorithm only outputs the prototype of community partition, it often occurs that the community is divided excessively and some nodes are divided incorrectly. In the above stage, each node label is determined by the most similar nodes in its neighborhood. In fact, nodes' labels are not only affected by their most similar nodes, but also by the remained nodes with lower similarity degree. Motivated by this idea and label influence [50], we introduce the label influence to further optimize the community division of networks. A node will be assigned to the label that has the greatest influence on it. Specifically, the influence of the label l^x on node v_i is defined as

$$IF^x(i) = \sum_{v_j \in \Gamma_1(i)} \delta_j^x * S_{ij}, \quad (7)$$

where $\Gamma_1(i)$ is the neighbor set of node v_i , S_{ij} denotes the similarity degree between node v_i and node v_j , which can be calculated by Eq. (3). $\Delta = (\delta_j^x)_{m \times N}$ is a confusion matrix. δ_j^x indicates whether the label of node v_j ($j = 1, 2, \dots, N$) is consistent with the label of the x -th category. If they are consistent, $\delta_j^x = 1$, otherwise it is 0. Here, $L = \{l^1, \dots, l^x, \dots, l^m\}$, is the label sets of the community partition, and $m = |L|$ denotes the number of communities. Note that the influence of the label on a node is measured by the sum of the similarity between a node and its neighbors with the same label.

With the label influence in hand, repeat the label propagation process until all nodes are assigned with the labels that have the most influence on them. Finally, the nodes with the same label are divided into the same community, and then the final community division are generated. The optimizing community partition for the Karate network is depicted in Fig. 4. It is easily seen that the community structure is more obvious and reasonable.

TABLE 1. The LFR Networks With Parameters: Node Number N , Average Degree $\langle k \rangle$, Maximum Degree k_{max} , Minimum Community Size $m(min)$, Maximum Community Size $m(max)$, the Ratio of External Degree of all Nodes μ .

Network	N	$\langle k \rangle$	k_{max}	$m(min)$	$m(max)$	μ
Net1	1000	20	50	20	100	[0.05, 0.65]
Net2	1000	20	50	50	150	[0.05, 0.65]
Net3	1000	20	50	20	100	[0.05, 0.75]
Net4	1000	20	50	50	150	[0.05, 0.75]

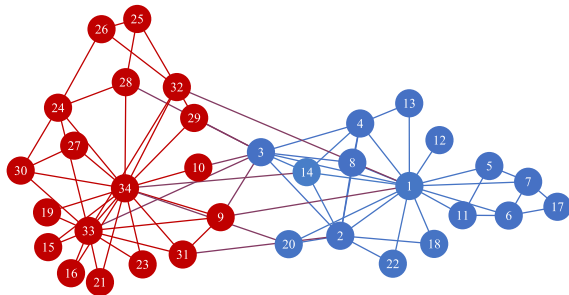


FIGURE 4. The final community division of Karate network.

D. THE PSEUDO CODE AND COMPLEXITY OF THE TNS-LPA ALGORITHM

The proposed algorithm consists of the above three phases, and the pseudo code is given in Algorithm 1.

For a given undirected and unweighted network $G = (V, E)$ with N nodes and M edges. The TNS-LPA algorithm have two key parts. The first part is the initial division of community structure. We need to compute the similarity TNS and minimum distance. The solution to TNS requires the nearest and the second nearest neighbors, with the complexity of $O(N * \langle k \rangle^2)$, where $\langle k \rangle$ denotes the average degree of the network. The minimum distance can be obtained through the shortest path, whose complexity can be $O(M + N * \log(N))$ [51]. The second part is to solve the problem of excessive community division in the previous stage, through LPA algorithm with the complexity of $O(M)$. To sum up, the complexity of TNS-LPA is $O(M + N * \log(N))$, far less than $O(N^2)$.

IV. EXPERIMENTAL SETUP

Before proceeding with the discussion of experimental results, we devote this section to the datasets and evaluation metrics.

A. DATASETS

1) ARTIFICIAL NETWORK

In order to test various community detection algorithms, Girvan and Newman firstly gave an artificial network [1], called GN benchmark. Due to its simple structure, most community detection algorithms perform very well on the GN benchmark. Subsequently, a new generalized Lancichinetti-Fortunato-Radicchi (LFR) benchmark was introduced, in which both the degree of nodes and the size of communities obey the power law distributions [52].

There are several parameters involved in LFR benchmark, among them, N is the total number of nodes, $\langle k \rangle$ and k_{max} are the average degree and maximum degree, respectively. $m(min)$ and $m(max)$ denote the minimum and maximum community size. The parameter μ represents the ratio of the external degree of each node. Obviously, with the increase of the mixing parameter μ , the community structure of LFR network is more indistinct. The details of these parameters are listed in Table 1.

2) REAL NETWORKS

Here we choose ten real-world networks to test the performance of the TNS-LPA, the basic information of these networks is listed in Table 2. One can download the datasets from <http://vlado.fmf.uni-lj.si/pub/networks/data/>, <http://www-personal.umich.edu/~mejn/netdata/> and <http://vlado.fmf.uni-lj.si/pub/networks/data/dic/fa/FreeAssoc.htm>.

B. EVALUATION METRICS

In this work, we compare the proposed algorithm with five popular algorithms according to three evaluation metrics, including modularity, normalized mutual information (NMI) and adjusted rand index (ARI).

- Modularity: For a given un-weighted network $G(V, E)$ with M edges, the modularity of the partition can be defined as [23]

$$Q = \frac{1}{2M} \sum_{i \neq j} \left(a_{ij} - \frac{k_i k_j}{2M} \right) \delta(C_i, C_j), \tag{8}$$

where a_{ij} is the element of the adjacency matrix, and k_i, k_j represent the degrees of nodes v_i and v_j , respectively. The Kronecker function $\delta(C_i, C_j)$ has the value 1 if its arguments are equal and 0 otherwise. Generally speaking, the greater the value of modularity Q is, the network has more obvious community structure.

- Normalized mutual information (NMI): For a given network $G(V, E)$ with N nodes, the NMI value between two divisions $X = \{X_1, X_2, \dots, X_{m(X)}\}$ and $Y = \{Y_1, Y_2, \dots, Y_{m(Y)}\}$ can be defined as [53]:

$$NMI = \frac{-2 \sum_{i=1}^{m(X)} \sum_{j=1}^{m(Y)} n_{ij} \log \left(\frac{n_{ij} \cdot N}{n_{X_i} \cdot n_{Y_j}} \right)}{\sum_{i=1}^{m(X)} n_{X_i} \cdot \log \left(\frac{n_{X_i}}{N} \right) + \sum_{j=1}^{m(Y)} n_{Y_j} \cdot \log \left(\frac{n_{Y_j}}{N} \right)}. \tag{9}$$

Algorithm 1 The TNS-LPA algorithm

Input: $A = (a_{ij})_{N \times N}$, adjacency matrix of a given network;
 $maxiter$, the maximum number of iterations

Begin

1: Initialization

2: for v_i ($i = 1 \rightarrow N$) in G do

3: $C_{Nei}(i) = |\Gamma_1(i)| + |\Gamma_2(i)|$, \leftarrow according to Eq. (5)

4: $\theta(i) = \min_{v_j: C_{Nei}(j) > C_{Nei}(i)} d_{ij}$

5: $IP(i) = C_{Nei}(i) * \theta(i)$ // Calculate nodes' importance using Eq. (4)

6: end for

7: $U = \{v_1^*, v_2^*, \dots, v_N^*\}$ // sort IP in descending order.

8: $U_1 = \{v_1^*, v_2^*, \dots, v_s^*\} \leftarrow U_1$ is the set of initial centers, and s is the number of initial centers.

9: $U_2 = \{v_{s+1}^*, v_{s+2}^*, \dots, v_N^*\}$

10: Calculate the similarity S_{ij} using Eq. (3) // $i = 1 \rightarrow N - 1, j = i + 1 \rightarrow N$

11: assign node v_i ($i = 1, \dots, N$) with a unique label l_i

12: do

13: for v_i^* ($i = 1 \rightarrow N - s$) in U_2 do

14: $l_i = \arg \max_{v_j \in \Gamma_1(i)} S_{ij}$ // update the labels of all non-community centers

15: end for

16: while (I) there exists node whose label still changes and (II) $iter \leq maxiter$

17: $L = \{l^1, l^2, \dots, l^m\}$ // L is the label sets of initial community partition, and $m = |L|$ denotes the number of communities.

18: Compute the label influence $IF^x(i)$ using Eq. (7) // $x = 1 \rightarrow m, i = 1 \rightarrow N$

19: do

20: for v_i^* ($i = 1 \rightarrow N$) in U do

21: $l_i = \arg \max \{IF^x(i) \mid x = 1, \dots, m\}$ // update the labels based on nodes' importance

22: end for

23: while (I) there exists node whose label still changes and (II) $iter \leq maxiter$

Output: return the final community partition $C = \{C_1, C_2, \dots, C_m\}$

End

TABLE 2. The Statistical Properties of the Ten Real-World Networks, Which Include: Node Number (N), Edge Number (M), Average Degree ($\langle k \rangle$), Clustering Coefficient ($\langle c \rangle$), Average Path Length ($\langle d \rangle$) And Community Number (m).

Network	N	M	$\langle k \rangle$	$\langle c \rangle$	$\langle d \rangle$	m
Karate	34	78	4.558	0.588	2.408	2
Polbooks	105	441	8.400	0.448	3.079	3
Football	115	613	10.661	0.403	2.508	12
Plobogs	1490	16718	22.440	0.361	2.738	2
Lesmis	77	254	6.597	0.736	2.641	—
Netscience	379	914	4.823	0.798	6.042	—
Email	1133	5451	9.622	0.254	3.606	—
Yeast	2375	11693	9.847	0.388	5.096	—
PairsFSG	10617	72168	12.016	0.187	4.004	—
Cond2003	31163	120029	7.703	0.723	5.767	—

In the above equation, $m(X)$ and $m(Y)$ denote the community numbers of partitions X and Y , respectively, n_{ij} is the number of common nodes in communities X_i and Y_j . For the variables $W = \{n_{X_1}, n_{X_2}, \dots, n_{X_{m(X)}}\}$ and $Z = \{n_{Y_1}, n_{Y_2}, \dots, n_{Y_{m(Y)}}\}$, n_{X_i} and n_{Y_j} represent

the numbers of nodes in X_i and Y_j . The denominator of NMI is just the sum of the entropies of W and Z . Note that the value of NMI is in the range $[0,1]$ and equals 1 only when two community divisions are exactly consistent.

- Adjusted rand index (ARI) based on pair counting is computed as follows [54]

$$\text{ARI} = \frac{\sum_{i=1}^{m(X)} \sum_{j=1}^{m(Y)} \binom{n_{ij}}{2} - \Omega}{\left[\sum_{i=1}^{m(X)} \binom{n_{X_i}}{2} + \sum_{j=1}^{m(Y)} \binom{n_{Y_j}}{2} \right] / 2 - \Omega} \quad (10)$$

where Ω is given by

$$\Omega = \sum_{i=1}^{m(X)} \binom{n_{X_i}}{2} \cdot \sum_{j=1}^{m(Y)} \binom{n_{Y_j}}{2} / \binom{N}{2}$$

Modularity reflects the closeness of the internal connection of the community through the difference between the strength of the connected edges in the actual community and the strength of the connected edges in the network under random division. NMI and ARI indicate the accuracy of community detection mainly by comparing the consistency between the results of community detection and the “true” community division. The larger the NMI and ARI values, the better the effect of community detection is.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the proposed algorithm will be compared with five popular algorithms on both artificial network and ten real networks. These comparative algorithms include FMM [24], ISCD+ [28], LPA [30], Stepping-LPA-S [34] and NI-LPA [36]. In what follows we will analyze the results in artificial networks and ten real networks given in Section 4.1.

A. RESULTS IN ARTIFICIAL NETWORKS

The comparative experiments of six different algorithms have been conducted on LFR benchmark with two network sizes, $N = 1000$ and 4000 . and other four parameters are taken as $\langle k \rangle = 20$, $k_{max} = 50$, $\gamma = 2$, $\beta = 1$.

Figure 5 shows the comparison of the NMI values on LFR network. The larger the NMI value is, the closer the partition result is to the real community structure. As shown in Fig. 5, the NMI value of TNS-LPA is larger than those of other five algorithms, which indicates that TNS-LPA is completely superior to other five methods on this network. For the sake of simplicity, we take Fig. 5(a) as an example. The FMM algorithm performs not so well in the whole range of the mixing parameter μ . When $\mu > 0.35$, the performance of the ISCD+ and Stepping-LPA-S algorithms decreased significantly. With the increase of the parameter μ , especially when $\mu > 0.45$, the NMI value for LPA declined obviously, even to zero. To summarize, our proposed algorithm (TNS-LPA) in general has better comprehensive performance than those five contrast algorithms.

Figure 6 gives the ARI values under six different community detection algorithms. When μ is less than 0.5, the ARI values obtained by the TNS-LPA algorithm is very close to 1. With the increase of μ , our proposed algorithm is still superior to other comparative algorithms for four sets of

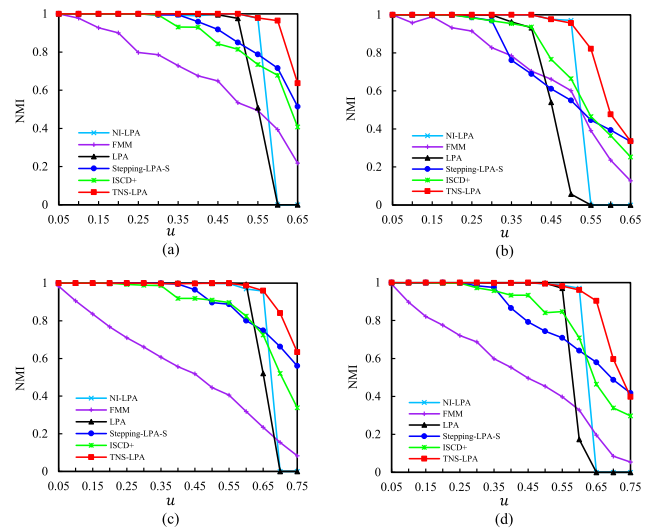


FIGURE 5. The comparison of normalized mutual information (NMI) between the proposed algorithm (square) and other six algorithms on LFR network with different parametric choices. (a) Net1, (b) Net2, (c) Net3, (d) Net4.

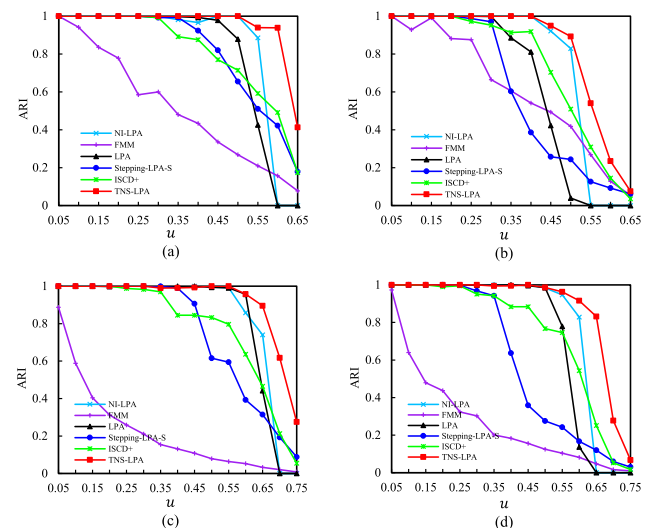


FIGURE 6. The comparison of adjusted Rand index (ARI) between the proposed algorithm (square) and other six algorithms on LFR network with different parametric choices. (a) Net1, (b) Net2, (c) Net3, (d) Net4.

parametric choices as depicted in Fig. 5. When the network scale increases up to 4000, the TNS-LPA algorithm is still better than other algorithms with respect to NMI and ARI. Therefore, the proposed algorithm shows good stability, reliability and scalability in artificial networks.

B. RESULTS IN REAL NETWORKS

In the following, we will compare the TNS-LPA with five algorithms on ten real networks. Among them, four networks including Karate, Polbooks, Football and Polblogs have known community division. Table 3 lists the modularity, NMI and ARI values of the six algorithms on these four real networks. In addition, Table 4 shows the comparison

TABLE 3. Comparison of the Modularity, NMI and ARI Values of TNS-LPA With Other Five Algorithms.

Network	Measure	FMM	LPA	Stepping-LPA-S	ISCD+	NI-LPA	TNS-LPA
Karate	Modularity	0.3807	0.3130	0.3715	0.3715	0.3490	0.3715
	NMI	0.6925	0.6446	0.9241	1	0.5801	1
	ARI	0.6803	0.6383	0.9393	1	0.5722	1
Polbooks	Modularity	0.5020	0.3023	0.4967	0.4974	0.4910	0.5187
	NMI	0.5308	0.3501	0.5712	0.5245	0.5754	0.5749
	ARI	0.6379	0.4051	0.6896	0.6390	0.6795	0.6967
Football	Modularity	0.5497	0.5874	0.5944	0.5972	0.6009	0.6010
	NMI	0.7078	0.8782	0.9035	0.9317	0.8685	0.9323
	ARI	0.4758	0.7436	0.8316	0.8920	0.6957	0.8946
Polblogs	Modularity	0.4269	0.0006	0.4187	0.4225	0.4259	0.4256
	NMI	0.3782	0.1161	0.3621	0.4574	0.3850	0.3887
	ARI	0.5285	0.0013	0.4984	0.5303	0.5357	0.5416

between the TNS-LPA algorithm and several popular contrast algorithms on six real networks without known community division. These real networks with different scales include Lesmis, Netscience, Email, Yeast, PairsFSG and Cond2003.

Figure 4 shows the community structure of the karate’s network which is obtained by the TNS-LPA algorithm. From Table 3, it can be found that the original LPA algorithm performs the worst according to three evaluating metrics. The modularity value of TNS-LPA is little worse than that of FMM. On the other hand, the NMI value of TNS-LPA is much larger than that of FMM. The ISCD+ has the same performance as our proposed algorithm on the Karate’s network. Therefore, among these five algorithms, the ISCD+ and TNS-LPA are the best algorithms to detect communities of Karate’s network.

Figure 7 depicts the community structure of the Polbooks network detected by the TNS-LPA algorithm. The Polbooks network contains three types of books related to American politics being purchased by users. If different books are purchased by the same user, there will be an edge between the corresponding nodes of books. Because of the implicit buyer of the “middle group”, the structure of the middle group community is not obvious, so the NMI and ARI values of five different algorithms are not so large. It is easily seen that the TNS-LPA algorithm has better performance because the detected community is more consistent with the original division. In addition, small communities are not merged into large communities due to the new update strategy in the TNS-LPA.

Figure 8 gives the community detection results of the football network obtained by the TNS-LPA algorithm. Compared with five contrast algorithms, the TNS-LPA can divide communities more accurately and get greater modularity value.

Figure 9 illustrates the community division results of Polblogs network detected by our proposed algorithm. The circle on the left represents liberals and the circle on the right represents conservatives. Because of the ambiguous political

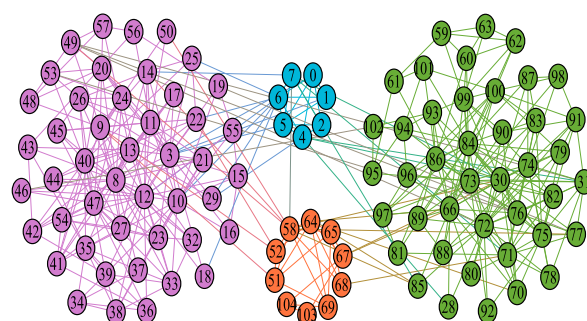


FIGURE 7. The community structure of polbooks network detected by TNS-LPA algorithm.

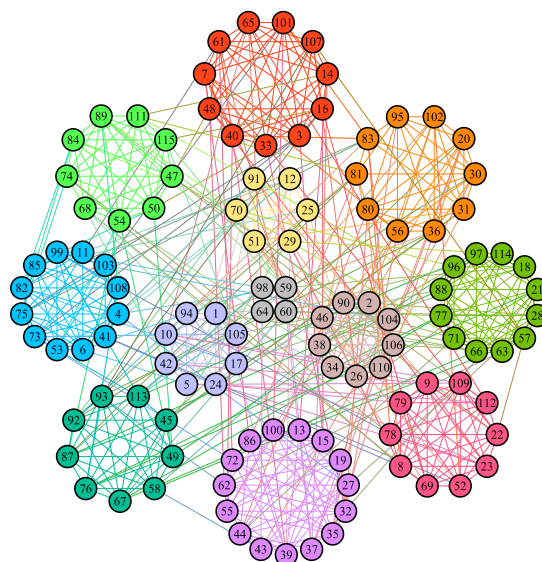


FIGURE 8. The community structure of football network detected by TNS-LPA algorithm.

attitude, some nodes are divided into opposing communities. For this network, the modularity value of the TNS-LPA is

TABLE 4. Comparison of the Modularity Values on the Six Real Networks Without Known Community Structures. “—” Means the Algorithm Run Out of Acceptable Computational Time.

Network	FMM	LPA	Stepping-LPA-S	ISCD+	NI-LPA	TNS-LPA
Lesmis	0.5006	0.3051	0.4677	0.4869	0.4839	0.5250
Netscience	0.8386	0.8016	0.8078	0.8083	0.8045	0.8262
Email	0.5037	0.0003	0.3640	0.5219	0.1160	0.5266
Yeast	0.7006	0.6948	0.5183	0.5356	0.6631	0.6916
PairsFSG	0.3797	0	0.3005	0.3601	0	0.4042
Cond2003	0.6788	0.6584	—	0.6903	0.6900	0.6947

TABLE 5. Comparison of 12 Different Similarity Indices Based on the Proposed Algorithm.

Network	Measure	CN	Salton	Jaccard	Sørensen	HPI	HDI	LHN	PA	AA	RA	LP	TNS
Karate	Modularity	0.3689	0.3781	0.0675	0.1540	0.3705	0.3566	0.3781	0.3715	0.3755	0.3736	0.3715	0.3715
	NMI	0.8986	0.5588	0.1079	0.3134	0.8575	0.5280	0.5588	1	0.7872	0.7829	1	1
	ARI	0.9188	0.3962	0.0720	0.2222	0.8720	0.3426	0.3962	1	0.8032	0.7836	1	1
Polbooks	Modularity	0.5160	0.5137	0.0058	0.4354	0.4954	0.5095	0.5005	0.4509	0.4991	0.5046	0.5114	0.5187
	NMI	0.5474	0.4674	0.0657	0.4455	0.4888	0.4414	0.4416	0.5442	0.5381	0.4845	0.5544	0.5749
	ARI	0.6608	0.4615	0.0378	0.5419	0.5373	0.4649	0.4097	0.6043	0.6529	0.5001	0.6721	0.6967
Football	Modularity	0.5295	0.5140	0.0362	0.2714	0.5284	0.5262	0.5152	0.4349	0.5331	0.5331	0.5331	0.6010
	NMI	0.8934	0.8871	0.2103	0.4582	0.8930	0.8920	0.8876	0.5593	0.8949	0.8949	0.8949	0.9323
	ARI	0.7595	0.7351	0.0304	0.1858	0.7577	0.7542	0.7371	0.2395	0.7650	0.7650	0.7650	0.8946
Polblogs	Modularity	0.4249	0.4246	0.0054	0.0034	0.4251	0.4236	0.4193	0.4246	0.4251	0.4250	0.4254	0.4256
	NMI	0.3775	0.3734	0.0839	0.1164	0.3788	0.3664	0.3512	0.3896	0.3797	0.3816	0.3845	0.3887
	ARI	0.0527	0.5181	0.0097	0.0029	0.5284	0.5037	0.4771	0.5437	0.5288	0.5306	0.5357	0.5416
Net1 ($\mu = 0.06$)	Modularity	0.2682	0.2561	0.1674	0.2184	0.2369	0.2546	0.2135	0.1892	0.2745	0.2633	0.2832	0.3364
	NMI	0.7841	0.7747	0.4426	0.5558	0.7525	0.7726	0.7256	0.5387	0.7992	0.7742	0.8403	0.9651
	ARI	0.5664	0.5141	0.1626	0.2635	0.4360	0.5050	0.3739	0.1015	0.5979	0.5445	0.6479	0.9382
Net3 ($\mu = 0.06$)	Modularity	0.2797	0.2453	0.2017	0.2531	0.2251	0.2414	0.2092	0.3433	0.2643	0.2492	0.2976	0.3711
	NMI	0.8400	0.8132	0.6002	0.6905	0.8028	0.8083	0.7831	0.8847	0.8291	0.8174	0.9036	0.9867
	ARI	0.5901	0.4745	0.2231	0.3402	0.4201	0.4578	0.3565	0.6702	0.5271	0.4811	0.6973	0.9575
PairsFSG	Modularity	0.3795	0.3033	0.1312	0.2156	0.3394	0.2937	0.2672	0.3874	0.3800	0.3560	0.3843	0.4042

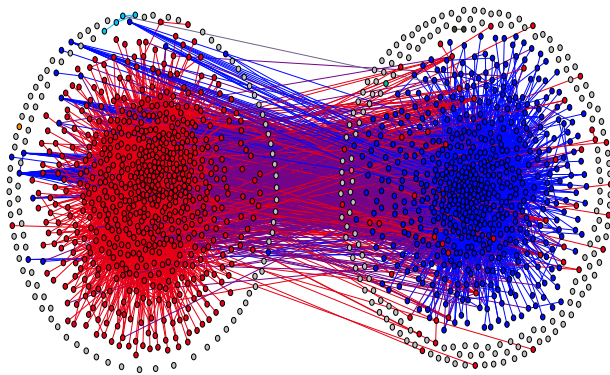


FIGURE 9. The community structure of polblogs network detected by TNS-LPA algorithm.

slightly less than that of FMM, but the ARI value is much larger than other comparative algorithms.

In order to further verify the TNS-LPA, we also consider six real networks without known community division, whose modularity values obtained six algorithms are listed in Table 4. The TNS-LPA algorithm can achieve maximum modularity in Lesmis, Email and PairsFSG networks. For the Netscience and Yeast network, the modularity value of TNS-LPA is slightly less than that of FMM. As shown in the comparative analysis of LFR network and four real networks with known divisions, the FMM can achieve larger modularity value but performs not very well in terms of NMI and ARI metrics. The traditional LPA performs worst on the Lesmis, Email and PairsFSG networks. Note that the TNS-LPA has good performance in large-scale networks such as Cond2003.

To sum up, the TNS-LPA algorithm performs competitively for the given networks. The proposed algorithm can not only produce the community partitions with a larger value of modularity, but also output a more accurate community which is more consistent with the original partitions. Thus the

proposed algorithm is the most suitable one among the six algorithms for community detections.

VI. DISCUSSION AND CONCLUSION

Label propagation is an efficient algorithm for community detection in complex networks due to its low time complexity. However, the uncertainty and randomness in the propagation of labels always affect its accuracy and stability. The node similarity measurement and the strategy for updating labels have a profound impact on the accuracy of community partition. For this purpose, this article proposes the TNS-LPA algorithm based on label propagation.

A. SIMILARITY COMPARISON

How to measure the similarity between nodes or links is a very essential and significant issue in community detection of complex networks. As mentioned in Section III, from the viewpoint of resource allocation and local topological structure, we present the TNS similarity measure by using the information of two-level neighborhood in networks. The new similarity is used in the latter two phases of the TNS-LPA. To illustrate the advantage of the TNS, we replace the similarity measure in the TNS-LPA with another 11 similarity indices [42], namely, common neighbours (CN), Salton index (Salton), Jaccard index (Jaccard), Sørensen index (Sørensen), Hub Promoted index (HPI), Hub Depressed index (HDI), Leicht-Holme-Newman index (LHN), Preferential Attachment (PA), Adamic-Adar index (AA), Resource Location (RA) and Local Path (LP). Table 5 gives the Q, NMI and ARI values on four real networks based on 12 different similarity measures, from which it is easily seen that the new TNS index performs the best on most of networks.

B. CONCLUSION

This article proposes an improved label propagation algorithm for community detection based on two-level neighborhood similarity (TNS-LPA), in which improves the LPA algorithm by using influence nodes and new community merging strategy. The TNS-LPA consists of three phases. First, we choose the initial community centers by measuring minimum distance and local centrality comprehensively. Then the label of each node is updated by employing a new label update strategy. Last, to avoid the excessive and inaccurate division, we introduce the label influence based on the proposed similarity to further optimize the community division of networks. The effectiveness of the TNS-LPA is illustrated through a series of experiments on both the artificial network and ten real networks. Compared with the five popular algorithms, namely, FMM, LPA, Stepping-LPA-S, ISCD+ and NI-LPA, our proposed algorithm has better comprehensive performance. In fact, there are complex and diverse community situations in real complex networks, such as a large number of small-scale communities, unbalanced distribution of community scale, less connections within small communities than between communities, and dynamic network structure, which brings great challenges to the research of

community detection. For the complex networks with these special community structure characteristics, further research on community detection is worthy in further works.

REFERENCES

- [1] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.
- [2] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, nos. 3–5, pp. 75–174, Feb. 2010.
- [3] K. Yang, Q. Guo, S.-N. Li, J.-T. Han, and J.-G. Liu, "Evolution properties of the community members for dynamic networks," *Phys. Lett. A*, vol. 381, no. 11, pp. 970–975, Mar. 2017.
- [4] N. Y. Chen, Y. Liu, and H. C. Chao, "Overlapping community detection using non-negative matrix factorization with orthogonal and sparseness constraints," *IEEE Access*, vol. 6, pp. 21266–21274, 2018.
- [5] C. Gui, R. Zhang, R. Hu, G. Huang, and J. Wei, "Overlapping communities detection based on spectral analysis of line graphs," *Phys. A, Stat. Mech. Appl.*, vol. 498, pp. 50–65, May 2018.
- [6] X. Pan, G. Xu, B. Wang, and T. Zhang, "A novel community detection algorithm based on local similarity of clustering coefficient in social networks," *IEEE Access*, vol. 7, pp. 121586–121598, Aug. 2019.
- [7] K. Berahmand, A. Bouyer, and M. Vasighi, "Community detection in complex networks by detecting and expanding core nodes through extended local similarity of nodes," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 4, pp. 1021–1033, Dec. 2018.
- [8] T. Wen and Y. Deng, "The vulnerability of communities in complex networks: An entropy approach," *Rel. Eng. Syst. Saf.*, vol. 196, Apr. 2020, Art. no. 106782.
- [9] J. Qadir, A. Ali, R. ur Rasool, A. Zwitter, A. Sathiseelan, and J. Crowcroft, "Crisis analytics: Big data-driven crisis response," *J. Int. Humanitarian Action*, vol. 1, no. 1, Aug. 2016, Art. no. 12.
- [10] A. Ali, "China pakistan economic corridor: Prospects and challenges for regional integration," *Arts Social Sci. J.*, vol. 7, no. 4, 2016, Art. no. 1000204.
- [11] H. Wang, J. Wu, W. Hu, and X. Wu, "Detecting and assessing anomalous evolutionary behaviors of nodes in evolving social networks," *ACM Trans. Knowl. Discovery Data*, vol. 13, no. 1, pp. 1–24, Jan. 2019, Art. no. 12.
- [12] F. Hao, Z. Pei, and L. T. Yang, "Diversified top-k maximal clique detection in social Internet of Things," *Future Gener. Comput. Syst.*, vol. 107, pp. 408–417, Jun. 2020.
- [13] 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.
- [14] K. Zhang, S. Bhattacharyya, and S. Ram, "Large-scale network analysis for online social brand advertising," *MIS Quart.*, vol. 40, no. 4, pp. 849–868, Apr. 2016.
- [15] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: A survey," *Data Mining Knowl. Discovery*, vol. 29, no. 3, pp. 626–688, May 2015.
- [16] B. Chen, W. Fan, J. Liu, and F.-X. Wu, "Identifying protein complexes and functional modules-from static PPI networks to dynamic PPI networks," *Briefings Bioinf.*, vol. 15, no. 2, pp. 177–194, Mar. 2014.
- [17] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Phys. Rep.*, vol. 519, no. 1, pp. 1–49, 2012.
- [18] P. Moradi and M. Rostami, "Integration of graph clustering with ant colony optimization for feature selection," *Knowl.-Based Syst.*, vol. 84, pp. 144–161, Aug. 2015.
- [19] K. Yang, Q. Guo, and J.-G. Liu, "Community detection via measuring the strength between nodes for dynamic networks," *Phys. A, Stat. Mech. Appl.*, vol. 509, pp. 256–264, Nov. 2018.
- [20] H. Jiang, L. J. Sun, J. Ran, J. X. Bai, and X. Y. Yang, "Community detection based on individual topics and network topology in social networks," *IEEE Access*, vol. 8, pp. 124414–124423, 2020.
- [21] 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, 2017.
- [22] F. Liu, J. Wu, S. Xue, C. Zhou, J. Yang, and Q. Sheng, "Detecting the evolving community structure in dynamic social networks," *World Wide Web*, vol. 23, no. 2, pp. 715–733, Mar. 2020.

- [23] 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, no. 2, Feb. 2004, Art. no. 026113.
- [24] 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, Jun. 2004, Art. no. 066133.
- [25] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 70, no. 6, Dec. 2004, Art. no. 066111.
- [26] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, Oct. 2008, Art. no. P10008.
- [27] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
- [28] L. Bai, X. Cheng, J. Liang, and Y. Guo, "Fast graph clustering with a new description model for community detection," *Inf. Sci.*, vols. 388–389, pp. 37–47, May 2017.
- [29] T. Zhang, L. Sun, and C. Bu, "Community detection via a triangle and edge combination conductance partitioning," *J. Stat. Mech., Theory Exp.*, vol. 2020, no. 7, Jul. 2020, Art. no. 073405.
- [30] 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, Sep. 2007, Art. no. 036106.
- [31] M. J. Barber and J. W. Clark, "Detecting network communities by propagating labels under constraints," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 2, Aug. 2009, Art. no. 026129.
- [32] X. Liu and T. Murata, "Advanced modularity-specialized label propagation algorithm for detecting communities in networks," *Phys. A, Stat. Mech. Appl.*, vol. 389, no. 7, pp. 1493–1500, Apr. 2010.
- [33] K. Liu, J. Huang, H. Sun, M. Wan, Y. Qi, and H. Li, "Label propagation based evolutionary clustering for detecting overlapping and non-overlapping communities in dynamic networks," *Knowl.-Based Syst.*, vol. 89, pp. 487–496, Nov. 2015.
- [34] W. Li, C. Huang, M. Wang, and X. Chen, "Stepping community detection algorithm based on label propagation and similarity," *Phys. A, Stat. Mech. Appl.*, vol. 472, pp. 145–155, Apr. 2017.
- [35] J. Ding, X. He, J. Yuan, Y. Chen, and B. Jiang, "Community detection by propagating the label of center," *Phys. A, Stat. Mech. Appl.*, vol. 503, pp. 675–686, Aug. 2018.
- [36] T. Wang, S. Chen, X. Wang, and J. Wang, "Label propagation algorithm based on node importance," *Phys. A, Stat. Mech. Appl.*, vol. 551, Aug. 2020, Art. no. 124137.
- [37] J.-G. Liu, L. Hou, X. Pan, Q. Guo, and T. Zhou, "Stability of similarity measurements for bipartite networks," *Sci. Rep.*, vol. 6, no. 1, Jan. 2016, Art. no. 18653.
- [38] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, May 2007.
- [39] K. Berahmand and A. Bouyer, "LP-LPA: A link influence-based label propagation algorithm for discovering community structures in networks," *Int. J. Mod. Phys. B*, vol. 32, no. 6, Mar. 2018, Art. no. 1850062.
- [40] Y. Zhang, Y. Liu, J. Zhu, C. Yang, W. Yang, and S. Zhai, "NALPA: A node ability based label propagation algorithm for community detection," *IEEE Access*, vol. 8, pp. 46642–46664, 2020.
- [41] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Social Netw.*, vol. 25, no. 3, pp. 211–230, Jul. 2003.
- [42] T. Zhou, L. Lü, and Y.-C. Zhang, "Predicting missing links via local information," *Eur. Phys. J. B*, vol. 71, no. 4, pp. 623–630, Oct. 2009.
- [43] Y. Pan, D.-H. Li, J.-G. Liu, and J.-Z. Liang, "Detecting community structure in complex networks via node similarity," *Phys. A, Stat. Mech. Appl.*, vol. 389, no. 14, pp. 2849–2857, Jul. 2010.
- [44] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, Mar. 1953.
- [45] P.-L. Yang, G.-Q. Xu, Q. Yu, and J.-W. Guo, "An adaptive heuristic clustering algorithm for influence maximization in complex networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 30, no. 9, Sep. 2020, Art. no. 093106.
- [46] N. A. Christakis and J. H. Fowler, "Social contagion theory: Examining dynamic social networks and human behaviors," *Stat. Med.*, vol. 32, no. 4, pp. 556–577, Feb. 2013.
- [47] I. A. Kovács, K. Luck, K. Spirohn, Y. Wang, C. Pollis, S. Schlabach, W. Bian, D.-K. Kim, N. Kishore, T. Hao, M. A. Calderwood, M. Vidal, and A.-L. Barabási, "Network-based prediction of protein interactions," *Nature Commun.*, vol. 10, Mar. 2019, Art. no. 1240.
- [48] Y. Chen, P. Zhao, P. Li, K. Zhang, and J. Zhang, "Finding communities by their centers," *Sci. Rep.*, vol. 6, no. 1, Apr. 2016, Art. no. 24017.
- [49] P. Yang, X. Liu, and G. Xu, "An extended clustering method using H-index and minimum distance for searching multiple key spreaders," *Int. J. Mod. Phys. C*, vol. 30, no. 07, Jul. 2019, Art. no. 1940008.
- [50] X.-K. Zhang, J. Ren, C. Song, J. Jia, and Q. Zhang, "Label propagation algorithm for community detection based on node importance and label influence," *Phys. Lett. A*, vol. 381, no. 33, pp. 2691–2698, Sep. 2017.
- [51] Y. Deng, Y. Chen, Y. Zhang, and S. Mahadevan, "Fuzzy dijkstra algorithm for shortest path problem under uncertain environment," *Appl. Soft Comput.*, vol. 12, no. 3, pp. 1231–1237, Mar. 2012.
- [52] 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, Oct. 2008, Art. no. 046110.
- [53] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.
- [54] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, Dec. 1971.



GUIQIONG XU received the B.S. and M.S. degrees from Lanzhou University, China, in 1995 and 1998, respectively, and the Ph.D. degree in computer software and theory from East China Normal University, in 2004. She is currently a Professor with the School of Management, Shanghai University, Shanghai, China, where she is also the Director of the Data Mining and Decision Support Research Center. She has also published more than 50 articles in academic journals indexed in well-reputed databases, such as Science Citation Index. Her current research interests include complex/social network modeling and analysis, data mining, decision theory and applications, and nonlinear systems and dynamics.



JIAWEN GUO received the B.S. degree in information system and information management from the Jiangsu University of Science and Technology, Soochow, China, in 2018. He is currently pursuing the M.S. degree in management science and engineering with Shanghai University, China. His current research interests include complex networks, machine learning, and business analytics.



PINGLE YANG received the B.S. and M.S. degrees in computer science and technology from the China University of Mining and Technology, China, in 2005 and 2008, respectively, and the Ph.D. degree in management science and engineering from Shanghai University, in 2020. He is currently an Associate Professor with the School of Electrical and Information Engineering, Jiangsu University of Science and Technology, Soochow, China, where he is also the Vice Director of the Artificial Intelligence and Data Processing Research Center. He is also engaging in the Postdoctoral Research with the University of Shanghai for Science and Technology. He has published more than ten articles in academic journals. His current research interests include complex/social network modeling and analysis, data mining, and public opinion management.

• • •