

Multiscale community estimation based on temporary local balancing strategy

Qiang Zhou*, Shi-Min Cai^{*,†,‡,¶} and Yi-Cheng Zhang^{†,§}

**School of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu 611731, P. R. China*

*†Institute of Fundamental and Frontier Science
University of Electronic Science and Technology of China
Chengdu 611731, P. R. China*

*‡Big Data Research Center,
University of Electronic Science and Technology of China
Chengdu 611731, P. R. China*

*§Department of Physics, University of Fribourg
1700 Fribourg, Switzerland*

¶shimincai@uestc.edu.cn

Community division in complex networks has become one of the hot topics in the field of network science. Most of the methods developed based on network topology ignore the dynamic characteristics underlying the structure. By exploring the diffusion process in the network based on random walk, this paper sums up the general rule with temporal characteristics as a temporary local balancing strategy which can be used in the community division. The strategy divides the network into different communities according to the duration of a stable local balancing state in the diffusion process. The longer the duration, the more stable the structure of the community in that state. Applying the strategy to computer-generated and real-world networks, respectively, it is proved that these temporary local balancing states existing in the diffusion process can reveal the internal community structure of the network. In addition, the modular structure appears at different time scales of diffusion process, similar to the hierarchical organization, and also provides a new perspective for multiscale network community detection.

Keywords: Community estimation; dynamic process; multiscale time state; random walk; temporary local balance.

PACS Nos.: 05.40.Fb, 89.75.Fb.

[¶]Corresponding author.

1. Introduction

Complex network is an abstract manifestation of the relationship between various entities in nature, so as to facilitate people to study a series of laws and principles related to nature. With the increasing trend of globalization, it can be found in many contexts. Twitter, for instance, is a large online virtual interaction system that includes hundreds of millions of users.¹ Amazon, a global chain of shopping site, is based on a business-to-business operation that provides a new model for the development of social networks.² As the phenomena of small-world³ and scale-free⁴ existing in complex networks are uncovered, network science theory has become an inter-discipline across physics, computer science and sociology, as well as engineering.^{5–11}

Along with lots of researches focused on the properties of the network topology, the most interesting is the community structure.^{12–15} In reality, a large number of complex networks have been unveiled with community modules, that is, the connections between the nodes within the community are more intensive than those among communities. An important feature of the community is its potential structural functional units, which play an important role in the link prediction and node classification. Community detection is to establish an algorithm model which conforms to the objective fact to excavate the hidden modules in the network. Although this is interesting, it is a remarkably challenging and sensitive task. A broad extent of different methods have been developed, but the most widely used at present is the modularity maximization, which classifies the communities that get higher modularity scores as a basic criterion. Unfortunately, calculating modularity maximization over all partitions in an exhaustive way has been proved to be an NP-hard problem.¹⁶ Using the maximum likelihood similarity¹⁷ based on statistical inference to solve this problem is also a computationally hard one. So, a series of heuristic algorithms that maintain the polynomial time are introduced to seek the approximate maximization such as greedy algorithms,^{18–20} extremal optimization^{21,22} and evolutionary algorithms.²³

Because of the dynamic nature of the network topology, its structure shows different local or global characteristics at different time scales.^{24,25} Many efforts have been paid to understand the dynamic process that occurs in the underlying network. Earlier, Arenas *et al.* studied the relationship between dynamic time scales and topological scales in complex networks,²⁶ which illustrates that synchronization can detect community structures at different topological scales. Following that, the synchronized clusters in modular networks are depicted by similar nodes that are shown by a series of works,^{27–30} and the result can be used to guide the detection of underlying communities. Besides, based on spin dynamics, the Hamiltonian features of the ferromagnetic and antiferromagnetic interactions were introduced to dynamically divide the community.³¹ Based on diffusion dynamics, random walk, a method based on information theory, is widely used to dynamically analyze the network community structure.^{14,32} For example, in the popular method *Walktrap*,

the similarity of nodes i and j used to detect the community structure is calculated based on the probability of randomly walking from i to j at a fixed number of steps.³³ Lambiotte *et al.* also introduced a systematic dynamical framework to analyze some of the stability of mass function in community detection based on random walk.³⁴

In this paper, considering that communities can be identified by running dynamical processes of random walk (i.e. diffusion process) on the network, we have established a dynamic temporary balancing strategy to detect the communities in the network, which also makes up for the lack of dynamic performance of the above-mentioned algorithms. Through the detailed analysis of the whole dynamic processes of random walk on the network, we notice that in the temporary local stable state, it is not completely attenuated to the next state. That is, there will be an intermittent stagnation in the diffusion process, which is closely related to the number of communities in the network. These dynamic processes of random walk correspond to the local duration of the relatively stable state of the network during the diffusion, that is, different time intervals will exhibit different network structure changes. These changes determine the current community division of the network at different topological scales. Therefore, it can be used as an evaluation criterion for multiscale community division in the diffusion process. Also, the dynamic processes of random walk, in turn, reversely reflect the diffusion of information transfer among different communities. It is a very natural and easy transfer process within the network and is related to the initial topology of the network. Through testing experiments on computer-generated and real-world networks, it is confirmed that this temporary balancing strategy is able to achieve a relatively stable and non-attenuating zone (i.e. intermittent stagnation) in the diffusion process, and thus detect the presence of the community. Moreover, because the multiscale community structure is detected in the diffusion process, it can effectively avoid the problem of resolution in the modularity maximization and improve the accuracy of community division.³⁵

The remainder of this paper is organized as follows. In Sec. 2, the concept of random walk algorithm and its application in community detection are introduced. On this basis, in Sec. 3, the temporary local balancing strategy is described and some usage-oriented performance is briefly discussed. The experimental results of benchmark tests for computer-generated networks and real-world networks are presented and compared in Sec. 4. Finally, we conclude the work in Sec. 5.

2. Random Walk in Community Detection

Random walk dynamics is one of the most developed methods in community detection. Assuming that a community has a high internal edge density and can be clearly distinguished from other communities, the random walker would be trapped in the community for a stable time duration until it finds a path to move out of the community. In order to better understand the random walk in the network, let us first introduce the principle of the current flow in the electrical circuits.

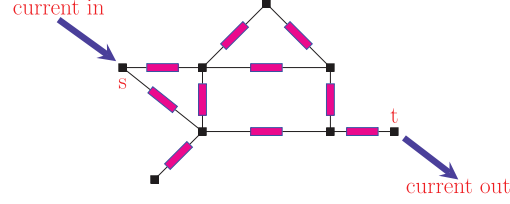


Fig. 1. (Color online) The schematic diagram of the current flow in an electrical circuit. As discussed in the main text, the edges in the circuit network have been replaced by a consistent unit resistor, and the unit current is injected into the circuit at the start node s and discharged at the target node t . According to Kirchhoff's law of conservation of currents, the current in or out of a node in a circuit is zero. Therefore, in the case of a known resistance, it is easy to calculate the voltage between two neighboring nodes.

First, we place the unit resistor on each edge of the network of interest to build the circuit. The unit current is injected into the circuit as the initial node at s , and it is retrieved at the target node t , as shown in Fig. 1. The voltage of node i in the network is represented by V_i and can be measured according to any convenient relative node. According to Kirchhoff's Law of current conservation,³⁶ the current entering or leaving a node in the circuit is zero, which indicates its voltage satisfying the following equation:

$$\sum_j A_{ij}(V_i - V_j) = \delta_{is} - \delta_{it}, \quad (1)$$

where δ_{ij} is the Kronecker factor, A_{ij} (for all i) is an element of the adjacency matrix and its form can be expressed as

$$A_{ij} = \begin{cases} 1, & \text{if node } i \text{ and } j \text{ are connected,} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

As $\sum_j A_{ij} = k_i$, where k_i is the degree of node i , Eq. (1) can be rewritten as a matrix

$$(\mathbf{D} - \mathbf{A}) \cdot \mathbf{V} = \mathbf{s}, \quad (3)$$

where \mathbf{D} is a diagonal matrix with $D_{ii} = k_i$, and the injected vector \mathbf{s} has elements

$$s_i = \begin{cases} +1, & i = s, \\ 0, & \text{otherwise,} \\ -1, & i = t. \end{cases} \quad (4)$$

It can be seen from Eq. (3) that because the Laplace matrix $\mathbf{D} - \mathbf{A}$ is singular, it cannot be inverted directly to obtain \mathbf{V} . As the voltage here is arbitrary within an additive constant, the vector $\mathbf{V} = (1, 1, 1, \dots)$ is always an eigenvector with eigenvalue zero. The decision item on the left side of Eq. (3) is the product form, so the result is always zero. In physics, this means that the electric current in the circuit is conserved, and mathematically, it suggests that for given N equations, there is at least one that is superfluous, so it has no effect on the solution. The solution to fix the

problem is also very simple. To remove one from the N equation, we can get the inverted matrix. Therefore, according to the voltage measured relative to a certain node v , the corresponding v th equation can be removed, which implies removing the v th row of $\mathbf{D} - \mathbf{A}$. Considering $V_v = 0$, the v th column can also be removed, so that an $(N - 1) \times (N - 1)$ matrix can be represented as $\mathbf{D}_v - \mathbf{A}_v$. Then according to Eq. (3), we find

$$\mathbf{V} = (\mathbf{D}_v - \mathbf{A}_v)^{-1} \cdot \mathbf{s}. \quad (5)$$

By definition, the voltage of the removed node v is zero. When performing the matrix representation, it is considered to re-add a v th row and column, and set its value to zero, which gives a new result matrix \mathbf{T} . Then, combined with Eq. (4), for source node s and target node t , the voltage at node i is given by

$$V_i^{(st)} = T_{is} - T_{it}. \quad (6)$$

The current flow through the i th node can be calculated by half the sum of the absolute values of the current flow along the edge incident on that node

$$\begin{aligned} I_i^{(st)} &= \frac{1}{2} \sum_j A_{ij} |V_i^{(st)} - V_j^{(st)}|, \\ &= \frac{1}{2} \sum_j A_{ij} |T_{is} - T_{it} - T_{js} + T_{jt}|, \quad (i \neq s, t). \end{aligned} \quad (7)$$

It is worth noting that Eq. (7) has no effect on the source node and the target node (the situation must consider the current injection and removal conditions), but these nodes have an exact current one unit, $I_s^{(st)} = I_t^{(st)} = 1$. If the traditional method is followed that does not include the ending node of the path, then both currents equal to zero instead of one. Here, only the case of one component in the network is described. For other cases, this process will be repeated for each component in the network. The model of the current flow seems intuitive and reasonable. The current will flow from the source node along all the paths to the target node, but it will choose a shorter path for the flow than the long path because the short path will generally have less resistance than the long path. However, there is no more sufficient reason to prove that the current flow in an electric circuit has anything to do with processes in nonpower networks, for instance, the social network. Although the principle of the current flow cannot be fully extended to large-scale nonpower networks, the current flow is similar to the random walk dynamics in the network, so, next, we will explore more friendly and intuitive random walks to various networks.

The aforementioned principle of the current flow is only used as a comparison of principles to facilitate us to better understand the random walk. The random walker is analogous to the current flow from the source node in the circuit. With a certain resistance, the voltage between the two endpoints determines the magnitude of the passed current. The greater the current, the closer the connection between the two endpoints. In other words, the higher the similarity between the two nodes in the

network, the greater the probability that the random walker will enter the community where the node is located. Therefore, with the notion of current flow in mind, it will become more natural when the principle of random walk dynamics is introduced.

Given a network G , we define a dynamic process of random walk on G , in which a walker moves from one node to another along the edges of the network. Suppose that $K = \{K_t, t \geq 0\}$ represents the current location of the walker, and $P\{K_t = j, 1 \leq j \leq n\}$ represents the probability that the walker will reach node j after t steps. And $P\{K_t | K_0, K_1, \dots, K_{t-1}\} = P\{K_t | K_{t-1}\}$ When $t > 0$. It means that the next state of the walker depends entirely on its previous state, which is called the Markov probability attribute.³⁷ Therefore, the dynamic process of random walk is a discrete Markov chain, and its state space is Z . Noting that $P\{K_t = j | K_{t-1} = i\} = p_{ij}$, so K_t is homogeneous, where p_{ij} is the transition probability from node i to j . According to Eq. (2), the adjacency matrix of the network G can be expressed as $A = (a_{ij})_{n \times n}$, we find

$$p_{ij} = \frac{a_{ij}}{\sum_r a_{ir}}. \quad (8)$$

For the aforementioned dynamic process, the position of source node s is specified. Let $\alpha_s^l(i)$ represent the probability that the walker starts from the node s through l steps and eventually reach an arbitrary target node i , then we can estimate iteratively by

$$\alpha_s^l(i) = \sum_{r=1}^n \alpha_s^{l-r}(r) \cdot p_{ri}, \quad (9)$$

α_s^l is the l step transition probability distribution. The sum of the probability of reaching all the nodes from the source node s is one. When $l = 0$, which means that the walker has been stuck in node s , so $\alpha_s^0(s) = 1$ and $\alpha_s^0(i) = 0$ for each $s \neq i$.

For a community, the connection density of its internal edge is much higher than that between them. When selecting an appropriate step l , a random walker can choose more paths from the source node s within l steps to reach the community where a node i is located. In contrast, the probability of a walker escaping from its current community is reduced. In other words, there are obstacles to the random walker in the path of escaping from the existing community and reaching the other community. Therefore, in a broad sense, this condition satisfies Eq. (10) when l is appropriate

$$\forall_{i \in C_s} \forall_{j \in C_s} : \alpha_s^l(i) > \alpha_s^l(j), \quad (10)$$

where C_s denotes the community where the node s is assigned. Although Eq. (10) describes a situation in which a node is unable to escape from the current community because of some special reasons, it is another way of stating that the community is very attractive to this node. This is consistent with the fact that nodes with similar

attributes are divided into the same community. Therefore, the random walk dynamics depicted above is suitable for most networks.

3. Temporary Local Balancing Strategy

In the dynamic diffusion process, if there are high edge density distributions within the community in a network, the walkers will stay within the community for quite some time before the next state. Assuming that in time t , the walker moves randomly from node i to its neighbor node j , combining with Eq. (8), we can modify the transition probability p_{ij} to be

$$\frac{d\rho_i(t)}{dt} = -\gamma \sum_j L_{ij}^T \rho_j(t), \quad (11)$$

where $\rho_i(t)$ ($1 \leq i \leq n$) is the probability that walker stays at nodes i at time t , and γ is a rate control parameter of diffusion process. Here, L is the normalized Laplacian matrix in the form of $L = I - D^{-1}A$, where D is a diagonal matrix like Eq. (3) and I is the identity matrix.

When the diffusion process in Eq. (11) reaches a local balancing state at a certain time (that is, the probability that the random walker continuously passes through the nodes i and j equals the probability that the random walker continuously passes the nodes j and i), the so-called detailed balance condition (i.e. the basis of temporary local balancing strategy) is obtained.³⁹ Under this condition, for any node, the diffusion process will positively shift toward a local balancing state as time evolves. Instead of only studying the final equilibrium, we consider the entire diffusion process (i.e. the whole dynamic processes of random walk). According to the detailed balance condition, the diffusion process satisfies the principle of node priority with high cohesion, and then considers the nodes with low cohesion. In order to determine how close two nodes i and j are under the detailed balancing condition at time t , the measure $c_{ij}(t)$ is given by

$$c_{ij}(t) = \left\langle \frac{\rho_i(t)}{\sum_j A_{ij}} - \frac{\rho_j(t)}{\sum_i A_{ij}} \right\rangle. \quad (12)$$

The right side of Eq. (12) illustrates the average over different implementations of the diffusion process that randomly selects the start node. In the testing experiments, if a pair of nodes i and j satisfy the detailed balance condition, then c_{ij} should be less than a given threshold (depending on the specific experimental environment). We can infer that, compared with the final balancing state, the diffusion processes corresponding to a group of nodes satisfy the detailed balance condition. By using Eq. (12), it is possible to track the different temporary local balancing states in the diffusion process which in turn represent the number of communities.

In order to illustrate the temporary local balancing strategy more clearly, the dynamic diffusion process based on the H13-4 network is analyzed in Fig. 2. The

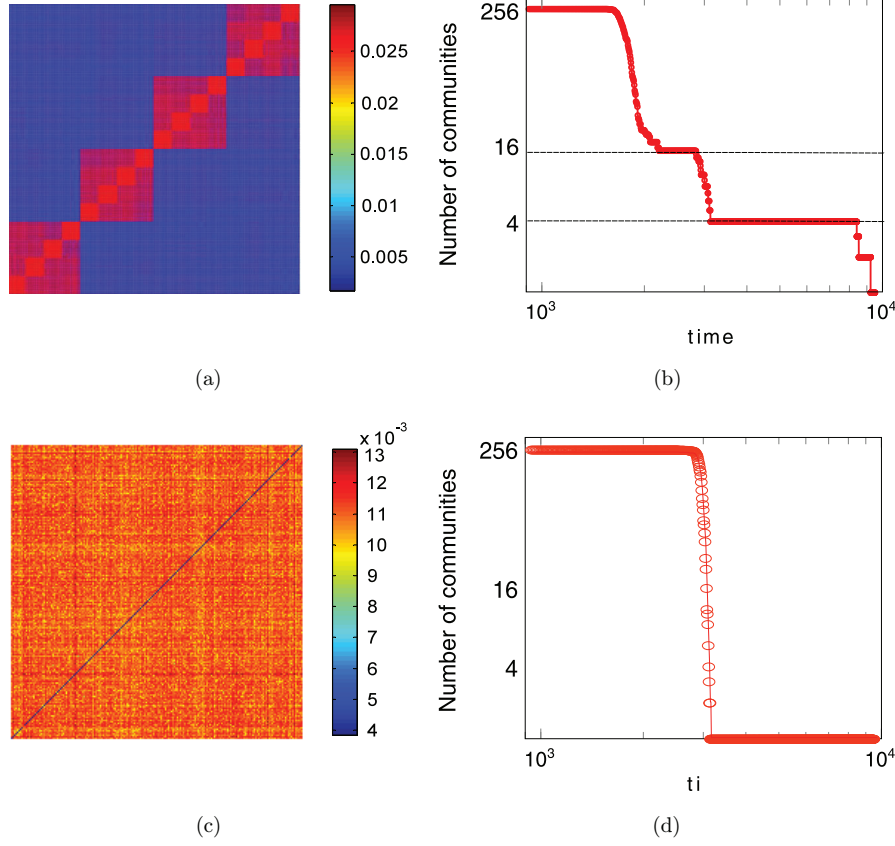


Fig. 2. (Color online) The verification of the temporary local balancing strategy by the dynamic diffusion process based on the H13-4 networks, (a) the matrix $c_{ij}(t)$ at two temporary local balancing states, which clearly shows that the H13-4 network is divided into two hierarchical levels and at each level the number of detected communities is 4 and 16; (b) the number of communities detected by the temporary local balancing strategy, which suggests that it is possible to determine whether the state transfer can be a new candidate for the number of communities by the temporary local balancing states in different transient states. The controlled trial is performed based on the randomized H13-4 network through randomly shuffling the edges for weakening the community structure, (c) the matrix $c_{ij}(t)$ at temporary initial and final balancing states, which clearly shows that the division satisfying the detailed balance condition is not clearly seen, (d) there is no local balancing states and the number of communities quickly changes from 256 to 1. Note that in the testing experiments, the rate control parameter of diffusion process is set $\gamma = 0.03$, and the threshold for $c_{ij}(t)$ is set with 1×10^{-3} . The value of each $c_{ij}(t)$ is the average over 1000 realizations of the diffusion process on a randomly selected start node.

computer-generated (or synthetic) H13-4 network contains two predefined hierarchies, the first level consists of four communities of 64 nodes and the second level consists of 16 communities of 16 nodes.²⁶ The average degree of node is $z_{in_1} + z_{in_2} + z_{out} = 18$, where z_{in_1} and z_{in_2} represent the internal degree of node at the first and second level, respectively, z_{out} indicates the external degree of node (e.g. for H13-4 network, $z_{in_1} = 13$, $z_{in_2} = 4$ and $z_{out} = 1$). According to Eq. (12), the two

temporary local balancing states associated with the diffusion process are shown in Fig. 2(a). The squares shown along the diagonal indicate a corresponding division that satisfies the detailed balance condition, which also conforms to the predefined hierarchy levels. In Fig. 2(b), we show the dynamic change of the number of communities as time evolves. Obviously, the divisions that satisfies the detailed balance condition is actually the community structures at two hierarchical levels obtained in a transient state. As an experimental contrast, we further study the dynamic diffusion process based on the randomized H13-4 network. Because the hierarchical community structure is weakened by randomly shuffling connections of the H13-4 network, the division that satisfies the detailed balance condition is not clearly seen in Fig. 2(c). The result is further verified in Fig. 2(d), in which there are no local balancing states and the number of communities quickly changes from 256 (each node is a community) to 1 (all nodes form a community).

The community detection based on temporary local balancing strategy is similar to the phenomena in the previous work.⁴⁰ That is, we adjust the network through the information in the oscillation circuit, so that the states between nodes are consistent. In addition, the strategy adopted in this paper determines that it can be tested at multiple scales because different time transients correspond to community structures at different topological scales. Compared with the method in the previous work,²⁶ the Laplacian matrix is used to calculate the eigenvectors corresponding to the network nodes, so the performance is very low in the face of heterogeneous networks. Herein, the normalized Laplacian matrix is used to effectively improve the distribution homogeneity of the temporary local balancing states, making it more stable and clear during the transient states in which the community exists.

4. Applications

In this section, we will give some applications for the proposed strategy to test its feasibility and accuracy. The data used include computer-generated (or synthetic) networks and real-world networks.

4.1. Synthetic networks

The use of computer-generated networks in testing can effectively detect the feasibility of the algorithm because such networks can be configured according to actual needs. First, similar to the aforementioned H13-4 network, we employ the H15-2 network by adjusting the internal degree of node at two levels (i.e. $z_{in_1} = 15$, $z_{in_2} = 2$ and $z_{out} = 1$). Obviously, the H15-2 network has more modularity in comparison with the H13-4 network. We perform the dynamic diffusion process based on the H15-2 network to verify the temporary local balancing strategy for such generalized stochastic block model.

Like the experimental results in Figs. 2(a) and 2(b), we illustrate the matrix $c_{ij}(t)$ at two temporary local balancing states and the number of communities as time

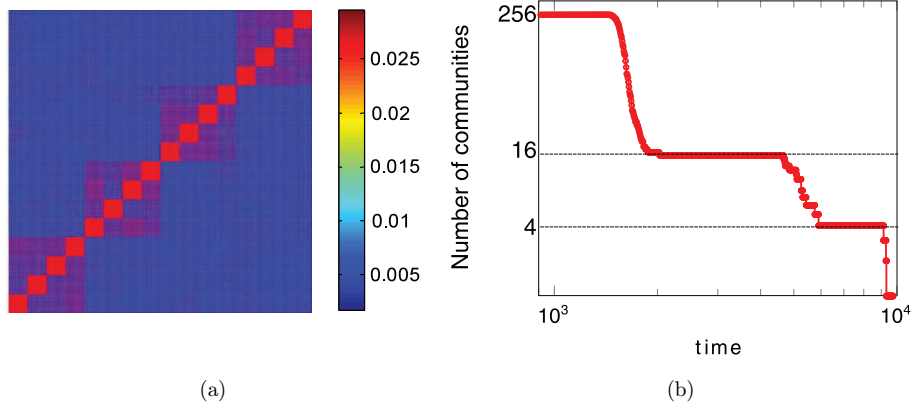


Fig. 3. (Color online) The verification of the temporary local balancing strategy by the dynamic diffusion process based on the H15-2 network, (a) the matrix $c_{ij}(t)$ at two temporary local balancing states, which clearly shows the prescribed hierarchical community structure of the H15-2 network; (b) the number of communities as a function of time, which indicates the effectiveness of the temporary local balancing strategy on multiscale community estimation. Note that in the testing experiments, the rate control parameter is set $\gamma = 0.03$, and the threshold for $c_{ij}(t)$ is set with 1×10^{-3} . The values of each $c_{ij}(t)$ is the average over 1000 realizations of the diffusion process at a random selected start node.

evolves, as shown in Fig. 3. First, in Fig. 3(a), it can be seen that the H15-2 network is accurately represented with two hierarchical levels according to $c_{ij}(t)$, and the first level contains 16 communities, while the second level consists of 4 communities. The intensity of $c_{ij}(t)$ at the first level is stronger than that at the second level, which implies that the temporary local balancing state at the first level is much more stable. Second, in Fig. 3(b), it shows the number of communities as a function of time, which suggests that the two temporary local balancing states are vividly reflected in the diffusion process based on the H15-2 network.

In order to make the network attribute be closer to the reality (i.e. the nodes have stronger degree heterogeneity), we also consider applying the alternative LFR model that generates synthetic networks with a power-law degree distribution and a clear community structure.⁴¹ Thus, when compared with the above-mentioned hierarchical community network, it has much more stringent requirements for the proposed algorithm. Herein, when the mixing factor is $\mu = 0.26$ (μ control the modularity of network), the synthetic network generated by LFR model (short of LFR network) includes 1000 nodes that construct 30 communities, and its average degree is set $\langle k \rangle = 20$. We also perform the diffusion process based on the LFR network, and show the experimental results of the matrix $c_{ij}(t)$ and the number of communities in Figs. 4(a) and 4(b), respectively. From them, we can find these similar results in comparison with those obtained from the above-mentioned hierarchical community network (i.e. the H14-3 network and H15-2 network), which verify the temporary local balancing strategy for detecting the community structure.

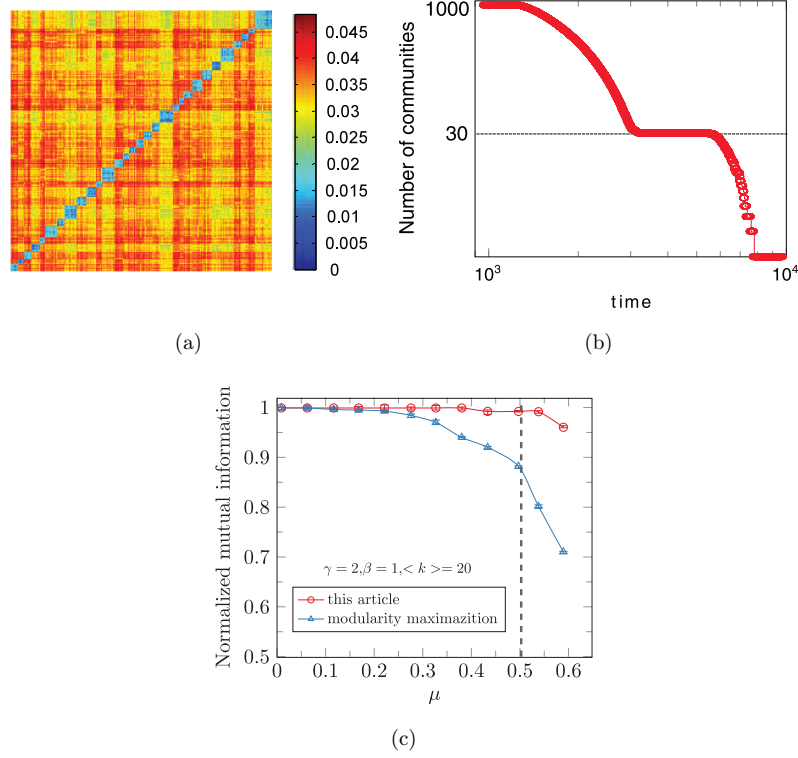


Fig. 4. (Color online) The verification of the temporary local balancing strategy by the dynamic diffusion process based on the LFR networks, (a) and (b) uncover the temporary local balancing state in the diffusion process on the LFR network with $\mu = 0.26$, and both of them show that the number of communities in the most stable state is 30 in consist with the ground truth; (c) using NMI as an evaluation criterion, the proposed algorithm is compared with modularity maximization, of which the experimental results suggest that proposed algorithm is much more robust. Note that in the testing experiments, the rate control parameter is set $\gamma = 0.01$, and the threshold for $c_{ij}(t)$ is 1×10^{-3} . The value of each $c_{ij}(t)$ and the score of NMI is respectively the average over 1000 realizations of the diffusion process at a random selected start node.

Furthermore, in order to better assess the quality of the detected communities, we introduce the normalized mutual information (NMI) between the detected community and the ground truth to evaluate the accuracy of community estimation.¹⁴ The higher the score of NMI is, the more accurate is the community estimation. In Fig. 4(c), we present the scores of NMI as a function of μ using the proposed algorithm and the modularity maximization.¹⁷ By increasing μ , the modularity becomes lower and the community structure of synthetic network is blurry, which is vividly shown in Fig. 4(c). Moreover, when the community structure strength is higher, both the proposed method and the modularity maximization can obtain an accurate result. However, when the community structure becomes more and more blurred, the modularity maximization will degenerate drastically, while the proposed algorithm can still maintain a high precision.

4.2. Real-world networks

The feasibility of the temporary local balancing strategy is further validated by the testing experiment based on these computer-generated networks. It cannot only improve the accuracy of community detection, but also explore the multiscale community structure with hierarchical features, such as the H15-2 and H13-4 networks. The real-world networks are much different from the random graphs in most cases, of which the distribution of node degrees is heterogeneous (i.e. the so-called scale-free feature). We will give its applications in real-world networks that are also widely used in community structure assessment tests.

Our first example is based on Zachary's karate club (short of karate network), which is a network of friendship among 34 members as nodes and 78 edges denote the friendship between them.⁴² Due to differences in management opinions, the club was divided into two different groups (or communities). In Fig. 5, we show the community division detected by the temporary local balancing strategy. More concretely, the diffusion process occurring at different stages can be clearly observed, where the two most stable temporary local balancing states are labeled as $T1$ and $T2$. $T1$ and $T2$ are judged by the duration of the process to be longer than other states, which indicates that random walkers are more likely to consider the special stages as a community candidate than other stages. We consider $T3$ as a candidate for the community divisions because the duration is second to the $T1$. According to the ground truth, the karate network should be accurate in two community divisions. Also, based on the result according to the modularity maximization, the karate network can be divided into four communities. Thus, in such cases, it is difficult to identify which criterion is better. Perhaps, as the network changes, the members of the club are rejoined into one community or continue to be divided into more communities. No matter what, the multiscale community structure is unveiled by $T1$, $T2$ and $T3$, as shown in Fig. 5(b). It can be seen that the community divisions in respect to $T1$, $T2$ and $T3$ are distinguished according to the node color.

Another real-world network we investigated is the U.S. college football network (short of football network) collected in Ref. 42. It contains 115 nodes representing the college teams and 613 edges between these teams describing the games between pairs of teams during the regular season in the year 2000. All teams are divided into 12 alliances, and each team is competing more frequently with the team within the alliance than the team outside. In Fig. 6(a), the adjacency matrix of the football network is exhibited according to the alliances. It can be seen that the arrangement along the diagonal directions presents a concentrated trend, which indicates that there exists a community structure. In Fig. 6(b), the matrix $c_{ij}(t)$ at the temporary local balancing state with the longest duration (e.g. $T3$) is presented. We can see that the football network is divided into 13 communities, and two of them may be merged (as indicated by the boxes) because the community structure hidden in different scales is likely to appear at the temporary local balancing state with relatively longer duration (e.g. $T2$). Thus, the temporary local balancing strategy has brought

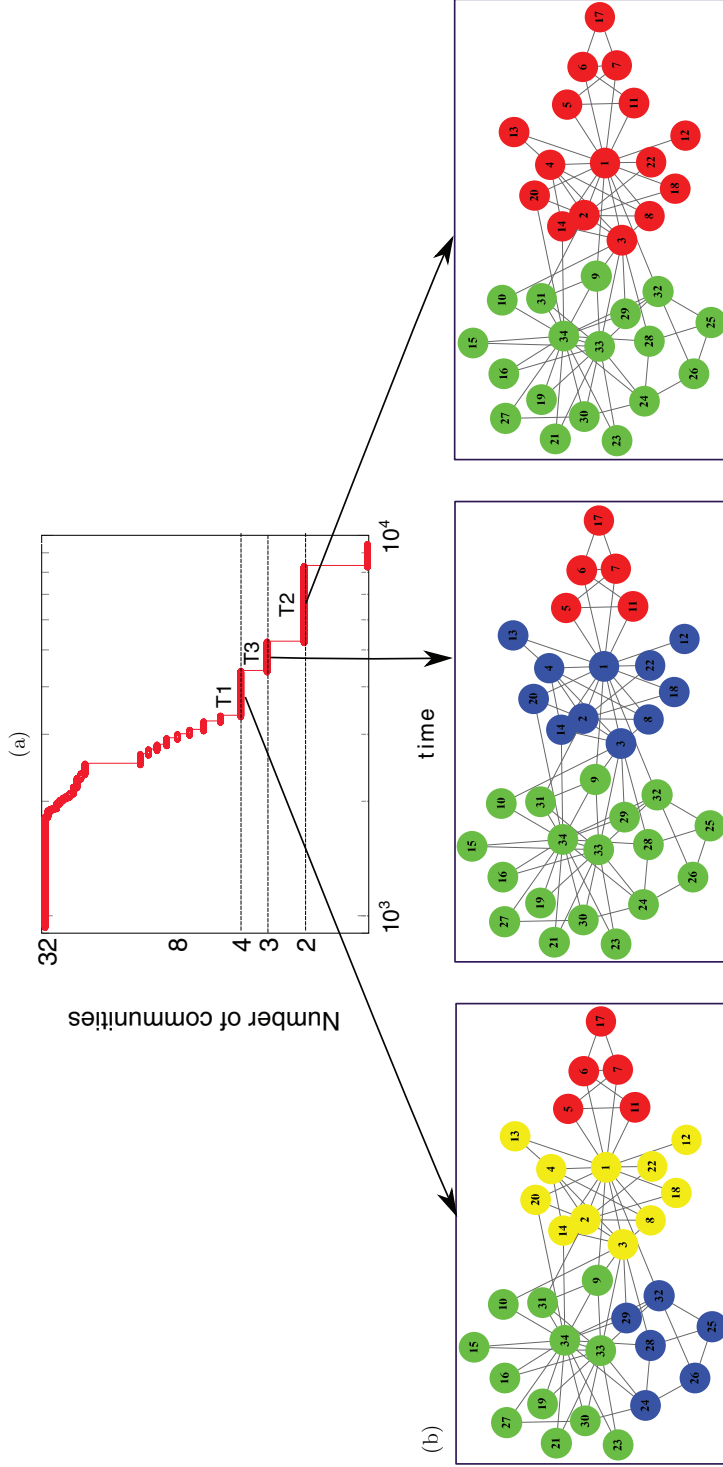


Fig. 5. (Color online) The verification of the temporary local balancing strategy by the dynamic diffusion process based on the karate network, (a) the number of communities at different temporary local balancing states changes with the time, which indicates three more stable states T_1 , T_2 , and T_3 with longer duration in the diffusion process; (b) the multiscale community structure corresponding to T_1 , T_2 and T_3 is unveiled by the nodes' color. Note that in the testing experiments, the rate control parameter is set $\gamma = 0.01$, and the threshold for $c_{ij}(t)$ is 1×10^{-3} .

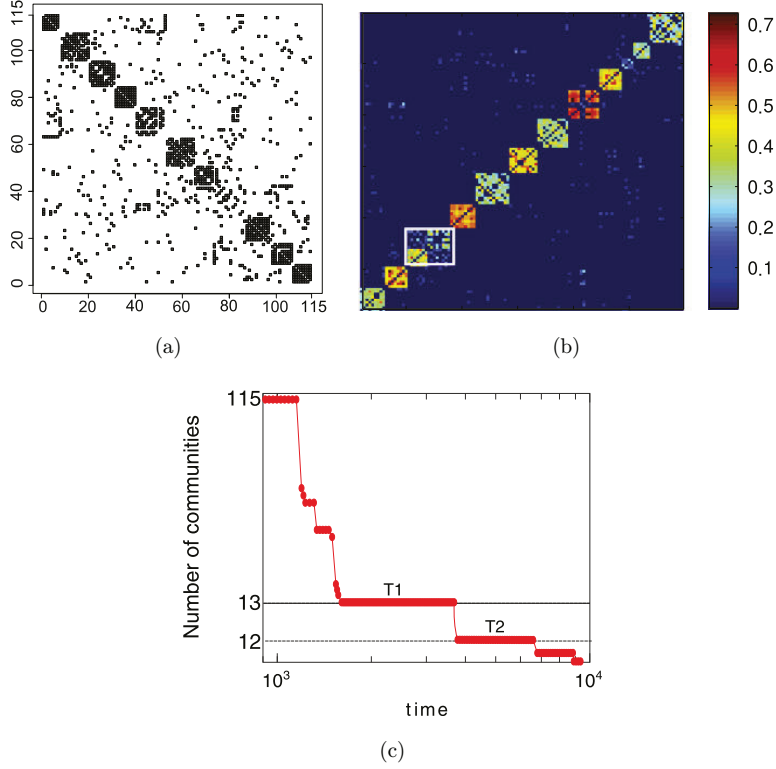


Fig. 6. (Color online) The verification of the temporary local balancing strategy by the dynamic diffusion process based on the football network, (a) the adjacency matrix of the football network ranged along the diagonal direction according to the alliances, which shows a concentrated trend implying the community structure; (b) matrix $c_{ij}(t)$ at the temporary local balancing state with the longest duration (e.g. T3), which indicates that the number of communities is 13; (c) The number of communities at different temporary local balancing states changes with time, which indicates several more stable states, among which T1 corresponds to the situation of 13 communities in (b) and T2 corresponds to 12 communities (merging two communities in the white box). We think that both of them can be considered as a candidate for the community division, however, according to the duration of the temporary local balancing state, T1 is closer to the ideal community division than T2. Note that in the testing experiments, the rate control parameter is set $\gamma = 0.01$, and the threshold for $c_{ij}(t)$ is 1×10^{-3} .

freedom to multiscale community estimation, and the proper intermediate state plays a key role on the accuracy of community detection. However, these temporary local balancing states with the long duration significantly provide candidates for the number of communities at multiple scale of the network. The results are confirmed by the number of communities at the temporary local balancing state T1 and T2, shown in Fig. 6(c). The number of communities may be 13, or it may be 12, because both T1 and T2 can maintain a longer duration. Because the duration of T1 is a little longer than T2, implying that the community division is more stable in the T1 than T2, we thus think that the resulting community structure obtained at T1 corresponds to the best community division.

In all, it can be found through the above-mentioned applications based on two real-world networks that several stable temporary local balancing states emerge during the dynamic diffusion process on networks with community structure. These stable states usually reveal the inherent community structure of the underlying network. Therefore, in the process of practice, we must pay attention to associating the phenomenon with the causes, and then guide our research in community detection.

5. Conclusions

In this paper, we have described a temporary local balancing strategy in the dynamic diffusion process on a network, which takes the temporary local balancing state appearing in various stages of network diffusion as a candidate for community division. These stable states are based on different durations of time, which can be understood as the structural and functional characteristics of the network at different topological scales. The mutual transfer between different states is actually a process of self-regulation within the network and eventually reaching stability. However, the transfer of state is not completed in an instant. After practice, it is found that before it decays to the next state, there will be intermittent stagnation of varying degrees, which is closely related to the number of communities in the network.

The dynamic diffusion process is based on random walk on a network, which is similar to the current flow in the electric circuit, and the mutual attraction between them is determined by the voltage between adjacent nodes. By studying the diffusion process of random walk model, we illustrate that the temporary local balancing state can reveal the potential network community structure. In the testing experiments, we explore the transient of the whole dynamic processes of random walk compared to the final stable state. The purpose is that if there is a strong community structure in the network, random walkers will be trapped in such a cluster (or community), and through a continuous cycle, they can get a local temporary balancing state, which corresponds to the internal structural units with very similar properties.

Applying the temporal local balancing strategy to the computer-generated network with predefined community structure, we found that it not only recovers the inherent community structure of the network, but could also further identify multiscale community structure with hierarchical levels. We also give applications on the real-world network and find our strategy to give higher community detection accuracy. In addition, it also gives more temporary local balancing states as a candidate for community detection and can serve as a reference for the final overall assessment of the network.

The temporal local balancing strategy of this paper only discusses the situation of the stable state within the community during the diffusion process, without considering the mutual influence between them. For example, what is the connection between the presence of two neighboring stable states, and how to set the detection

thresholds for determining whether a community exists. Nevertheless, our work provides a new vision for the detection of the number of communities, and can be extended to more complex networks that contain multiple topological scales. In addition, in the actual application, we must also consider the comprehensive influence of more factors, so that we can get a good community division closer to the ground-truth.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61673086) and the Science Promotion Programme of UESTC, China (No. Y03111023901014006).

References

1. W. Fan and K. Yeung, *Commun. Nonlinear Sci. Numer. Simul.* **20**, 1015 (2015).
2. J. Yang and J. Leskovec, *Knowled. Inform. Syst.* **42**, 181 (2015).
3. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
4. A. L. Barabási and R. Albert, *Science* **286**, 509 (1999).
5. S. Fortunato and D. Hric, *Phys. Rep.* **659**, 1 (2016).
6. M. E. Newman, arXiv preprint arXiv:1606.02319 (2016).
7. N. Masuda, M. A. Porter and R. Lambiotte, *Phys. Rep.* **716**, 1 (2017).
8. M. X. Liu, W. Wang, Y. Liu, M. Tang, S. M. Cai and H. F. Zhang, *Phys. Rev. E* **95**, 052306 (2017).
9. X. X. Zhan, C. Liu, G. Zhou, Z. K. Zhang, G. Q. Sun, J. J. Zhu and Z. Jin, *Appl. Math. Comput.* **332**, 437 (2018).
10. C. Qu, X. X. Zhan, G. Wang, J. Wu and Z. K. Zhang, *Chaos* **29**, 033116 (2019).
11. X. Liu, H. M. Cheng and Z. Y. Zhang, *IEEE Trans. Knowl. Data Eng.* (2019). DOI: 10.1109/TKDE.2019.2911943.
12. S. Fortunato, *Phys. Rep.* **486**, 75 (2010).
13. X. Q. Chen, F. X. Ren, H. W. Shen, Z. K. Zhang and T. Zhou, *J. Statist. Mech.* **2010**, P10011 (2010).
14. M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West and R. Lambiotte, *Nat. Commun.* **5**, 4630 (2014).
15. C. Zhe, A. Sun and X. Xiao, Community detection on large complex attribute network, in *Proc. 25th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Anchorage, Alaska, USA (2019).
16. M. E. Newman and T. P. Peixoto, *Phys. Rev. Lett.* **115**, 088701 (2015).
17. M. E. Newman, *Phys. Rev. E* **94**, 052315 (2016).
18. A. Clauset, M. E. Newman and C. Moore, *Phys. Rev. E* **70**, 066111 (2004).
19. V. D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, *J. Statist. Mech.* **2008**, P10008 (2008).
20. Z. Zhang, P. Pu, D. Han and M. Tang, *Physica A* **506**, 975 (2018).
21. J. Duch and A. Arenas, *Phys. Rev. E* **72**, 027104 (2005).
22. J. Shao, Z. Zhang, Q. Yang, Z. Yu, J. Wang, Y. Zhao and Q. Yang, Community detection and link prediction via cluster-driven low-rank matrix completion, in *Proc. 28th Int. Joint Conf. Artificial Intelligence*, Macao, China (2019).
23. M. Gong, L. Ma, Q. Zhang and L. Jiao, *Physica A* **391**, 4050 (2012).

24. X. Zhang, C. Moore and M. E. Newman, *Eur. Phys. J. B* **90**, 200 (2017).
25. Y. Li, K. He, K. Kloster, D. Bindel and J. Hopcroft, *ACM Trans. Knowl. Discov. from Data* **12**, 17 (2018).
26. A. Arenas, A. Diaz-Guilera and C. J. Pérez-Vicente, *Phys. Rev. Lett.* **96**, 114102 (2006).
27. S. Li, Y. Chen, H. Du and M. W. Feldman, *Complexity* **15**, 53 (2010).
28. M. Y. Zhou, Z. Zhuo, S. M. Cai and Z. Fu, *Chaos* **24**, 033128 (2014).
29. J. Shao, Z. Han, Q. Yang, and T. Zhou, Community detection based on distance dynamics, in *Proc. 21th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, California, USA (2016).
30. Z. Zhuo, S. M. Cai, M. Tang and Y.-C. Lai, *Chaos* **28**, 043119 (2018).
31. H. J. Li, Y. Wang, L. Y. Wu, Z. P. Liu, L. Chen and X. S. Zhang, *EPL* **97**, 48005 (2012).
32. Z. K. Zhang, C. Liu, X. X. Zhan, X. Lu, C. X. Zhang and Y. C. Zhang, *Phys. Rep.* **651**, 1 (2016).
33. P. Pons and M. Latapy, Computing communities in large networks using random walks, in *Proc. 20th Int. Symp. Computer and Information Science*, Istanbul, Turkey (2005).
34. R. Lambiotte, J. C. Delvenne and M. Barahona, *IEEE Trans. Net. Sci. Eng.* **1**, 76 (2014).
35. S. Fortunato and M. Barthélemy, *Proc. Natl. Acad. Sci.* **104**, 36 (2007).
36. C. K. Volos, I. Kyprianidis, I. Stouboulos, E. Tlelo-Cuautle and S. Vaidyanathan, *J. Eng. Sci. Technol. Rev.* **8** (2015).
37. M. T. Schaub, J. C. Delvenne, S. N. Yaliraki and M. Barahona, *PloS One* **7**, e32210 (2012).
38. M. E. Newman, *Proc. Natl. Acad. Sci.* **103**, 8577 (2006).
39. A. Barrat, M. Barthélemy and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).
40. A. Arenas and A. Diaz-Guilera, *Eur. Phys. J. Spec. Top.* **143**, 19 (2007).
41. A. Lancichinetti, S. Fortunato and F. Radicchi, *Phys. Rev. E* **78**, 046110 (2008).
42. M. Girvan and M. E. Newman, *Proc. Natl. Acad. Sci.* **99**, 7821 (2002).