

Selection of Significant Community Structure Based on Network Partition-based Cluster

Qingju Jiao*, Yuanyuan Jin

School of Computer and Information Engineering, Anyang Normal University, Anyang 455000, China

Corresponding Author Email: qjjiao@aynu.edu.cn

<https://doi.org/10.18280/jesa.520105>

ABSTRACT

Received: 10 November 2018

Accepted: 5 February 2019

Keywords:

complex network, module structure, multi-scale module detection, significant partition

Module (or community) structure detection, which has been successfully applied to many fields, is vital step for understand network dynamic and complex systems. Module structure not only is a crucial character of networks, but also is multi-scale. Therefore, many multi-scale module detection algorithms are proposed to resolve the problem. But a highly important issue for multi-scale methods is that of how to select crucial partitions among multi-scale network partitions so that these partitions can effectively help people to understand complex system. To solve the problem, we propose a novel partition-based hierarchical clustering to select significant network partitions. Experiments on selection of benchmark and real networks demonstrate that the new method for selecting significant partitions is very effectively.

1. INTRODUCTION

Many complex systems, such as biology, mathematics, sociology and physics can be abstracted as networks that only include nodes and edges. The merit is simple, but can provide significant convenience to perceive and explain mysteries of complex systems. The assembling property of network, i.e. community, has triggered a big activity to this field [1-3], which is considered to be capable of revealing the network structure. Many methods have been put forward to detect network communities, such as modularity [4] and its variations [5-7], and hierarchical clustering methods based on node [8] or edge similarity [9].

Although much progress has been obtained, many issues challenges existing community detection algorithms greatly in traditional single-scale approaches. Single-scale community detection algorithm may encounter the so-called resolution limit problem [10] (or termed underpartition) when applied on the real complex networks. For instance, the modularity orientated optimization is a branch of the single-scale community mining methods, which have been widely investigated. However, even the most recently reported infomap method [11], which is powerful for detecting non-overlapping communities, also suffers from this problem [12]. Other local-based single-scale community detection methods have overpartition problem since they have a tendency to group together those nodes with the strongest connections but leave out those with weaker connections, so that the divisions they generate may consist of smaller (even including several nodes) and dense cores [13]. One of the potential reasons for the underpartition and overpartition phenomenon in the single-scale community detection approaches is that the network itself has multi-scale community structures. Also, measuring the node similarity solely according to the links in the original observed network in existing methods will potentially result in the inaccuracy of the solutions.

Another difficulty for existing community mining algorithms is that networks often have several levels

organization [14-15], leading to different relevant communities at various scales. For instance, in the PPI networks, the protein complexes are often organized hierarchically [16-17]; the atomic-level networks of protein 3D structures are characterized by multiple scales in time and space [18-19]. Single-scale community detection methods are not suitable for the analysis of the multi-scale networks in which there is not a single ‘best’ mesoscopic level of description, but rather multiple levels association with multiple scales in the network [20]. Some recent methods [21-23] have been proposed to try to resolve the multiple scales of the complex system, but significant efforts are still needed. For instance, the stability method is a multi-scale community mining algorithm based on extended modularity idea, its performance will be affected by the inherent weaknesses of modularity definition as some networks. Another multi-scale map method is found to produce more groups without meaningful community structure at different scales. Apart from the two issues mentioned above, another most significant issue for people is that how to select crucial partitions among multi-scale network partitions, and these partitions can effectively help one to understand complex system.

In this work, we propose a network partition-based cluster method to select significant partitions based on ISIMB multi-scale method [24]. The proposed method first calculates similarity (or distance) between two partitions using normalization variation of information, and a similarity matrix is generated, second, hierarchical clustering with a defined objective function is employed to select significant partitions. Experiments on selection of benchmark and real networks demonstrate that the new method for selecting significant partitions is very effectively. Particularly, we apply the multi-scale concept to mine and select protein complexes in PPI networks and reveal their hierarchical organizations, which open a new way to understand and insight into protein complexes.

2. METHODS

The ISIMB method has been proposed based on ISIM similarity. ISIM similarity first define transition probability from node i to node j at the $(t+1)th$ time is described below:

$$p_{ij}(t+1) = (1-\alpha) \frac{1}{SP(i,j)} + \alpha \sum_{k=1}^{|U_i|} \frac{1}{d_i} p_{kj}(t) \quad (1)$$

where, $\alpha \in (0,1)$ parameter is a regulative factor, and $SP(i,j)$ is the shortest path between node i and node j , d_i is the degree of node i , and U_i denotes the sets of neighbours of node i , and $k(k \neq i)$ is a member of U_i . The equation (1) can be calculated by iterative (equation 2) or convergent (equation 3) equations on whole network, and the ISIM node similarity between node i and node j is defined by equation (4).

$$\mathbf{P}(t+1) = (1-\alpha)\mathbf{SPV} + \alpha\mathbf{WP}(t) \quad (2)$$

$$\mathbf{P} = (1-\alpha)(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{SPV} \quad (3)$$

$$S_{ij} = \frac{\mathbf{P}_{ij} + \mathbf{P}_{ji}}{2} \quad (4)$$

For a network, ISIMB method first set some values (such as 60) for parameter α , and then, a series of similarity matrix of the network are obtained according ISIM, third, the number of modules are selected via the number of eigenvalues of similarity matrixes that larger than 0.2, at last, we employ hierarchical clustering with similarity matrix and preset number of communities to mine multi-scale community structure.

Using ISIMB method, we get a series of partitions of a network at different scales. However, one expects that the related or significant partitions can be identified from multi-scale community structures easily. Before identifying the significant scales, we should give a definition of it. A related or significant scale in a series of states is stable and it is more persistent than others when changing the resolution scale of the topology because they are usually more important to understand the system. In this work, a significant partition with the same number of communities and modular structure should span more parameters (α) than other scales. Here, we present a new method to detect related scales. The method uses hierarchical clustering to group different partitions revealed by the ISIMB method. The similarity between two partitions are measured by normalization variation of information (NVI) [25] which has all the properties of a proper distance measure.

$$\begin{aligned} V(X,Y) &= H(X) + H(Y) - 2I(X;Y) \\ &= H(X|Y) + H(Y|X) \\ &= -\sum_{xy} P(x,y) \log \frac{P(x,y)}{P(y)} - \sum_{xy} P(x,y) \log \frac{P(x,y)}{P(x)} \end{aligned} \quad (5)$$

where, X and Y are two network partitions, $H(X)$ is the entropy of X and $H(X|Y)$ is the conditional entropy, i.e., the additional information needed to describe X once we know Y . $I(X;Y)$ is defined by equation 6.

$$I(X;Y) = \sum_{x=1}^K \sum_{y=1}^{K'} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \quad (6)$$

where, K and K' are the number of network partitions X and Y respectively.

First, the distances among 60 partitions (a partition is similar with a node in traditional way) which are labeled by a series of digits from 1 to 60 in order are computed and a distance matrix is generated. Second, the hierarchical clustering is performed and the number of clusters (nc) are set from 1 to 60. Third, an objective function F is defined based on the average of intra-cluster distance and average inter-cluster distance.

$$F = \frac{\frac{1}{nc} \sum_{nc=1}^{60} (\sum_{X \in C_{nc}} \sum_{Y \in C_{nc}, Y \neq X} NVI(X,Y))}{2 \times nc \times (nc-1) \sum_{C_p, C_q} (\frac{1}{n_p \times n_q} \sum_{X \in C_p} \sum_{Y \in C_q} NVI(X,Y))} \quad (7)$$

where, X and Y are two network partitions, nc is the number of clusters, n_p and n_q are the number of partitions in clusters C_p and C_q respectively. In contrast to traditional cluster methods, the cluster methods used here is constrained clustering because the partitions in one cluster need to be ordinal. Therefore, for the final clustering result, it not only satisfies the minimum of the objective function F but also makes the following equation 0.

$$\sum_{i=1}^{C_{number}-1} (L_{i+1} - L_i) - (C_{number}-1) \quad (8)$$

where, C_{number} is the number of partitions in a cluster, L_i ($1, 2, \dots, 60$) is the label of partition i . At last, we select a partition to represent the cluster. For a given cluster, we compute the average NVI of each partition among other partitions, and pick the partition that has the minimum average of the NVI to represent the given cluster.

3. RESULTS

3.1 The performance of ISIMB method

The performance of ISIMB method is tested on both synthetic and real-world benchmark networks with other multi-scale methods, Stability, Map and RB [26] methods. The stability method detects module by optimizing a function ‘Stability’. The modular structure generated by stability method significantly depends on a time scale (Markov time). Using different time scales, one can obtain various community structures with different scales. Hence the Markov time acts effectively as a resolution regulator that reveals multi-scale community structures from fine to coarser as the Makov time grows. In particular, some standard partitioning measures, such as modularity, correspond to ‘Stability’ with different time scales. Likewise, the Map method reveals multi-scale community structure also using Markov time. In the Map method, the community detection problem is interpreted as compression of the description length of a random walk. In order to reveal multi-scale community structure, Map method

introduces dynamic Markov time into the process of map encoding. The RB method finds community structures in networks by minimizing the energy of the spin glass, and reveals modules at different scales using a resolution regulator in the energy. Different from other three methods, the resolution regulator in RB method varied from big to small: as one decreases the resolution regulator, communities in the network are merged, and more big communities are generated. When the resolution regulator is 1, the function energy in the RB method is the standard modularity quality function. In addition, the community results detected by four multi-scale methods are compared with real partitions (or real community structure) using Normalized Mutual Information (NMI, see equation 9). Given two partitions $\chi = (X_1, X_2, \dots, X_{n_X})$ and $\gamma = (Y_1, Y_2, \dots, Y_{n_Y})$ of a network, with n_X and n_Y communities, respectively:

$$NMI(\chi, \gamma) = \frac{-2 \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} n_{ij}^{XY} \log\left(\frac{n_{ij}^{XY} \cdot N}{n_i^X \cdot n_j^Y}\right)}{\sum_{i=1}^{n_X} n_i^X \log\left(\frac{n_i^X}{N}\right) + \sum_{j=1}^{n_Y} n_j^Y \log\left(\frac{n_j^Y}{N}\right)} \quad (9)$$

where, N is the number of nodes in a network, n_i^X and n_j^Y are the number of nodes in the communities X_i and Y_j respectively, n_{ij}^{XY} is the number of nodes shared by the communities X_i and Y_j : $n_{ij}^{XY} = |X_i \cap Y_j|$. The larger the NMI value, the better the community partition.

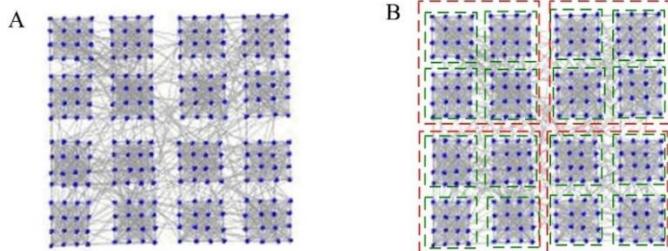


Figure 1. H 15-2 network structure (A) and two hierarchical community structures (B)

The first synthetic network is the H 15-2 network, which describes a homogeneous in degree network with two predefined hierarchical levels. It contains 256 nodes in total, and 15 edges are placed in the most internal community with 16 nodes. 2 edges are linked among the most internal communities and form the most external community including 64 nodes, and 1 edge is linked with any other node at random in the network. The H 15-2 network (see Figure 1A) has hierarchical community structures (see Figure 1B): the first level includes 16 communities and each community has 16 nodes, and the second level includes 4 communities and each community has 64 nodes. From the definition of community (or module) in which group of vertices within which connections are dense, but between which connection are sparse, we can infer that 16-way partition is more stable than 4-way partition because the nodes in 16-way partition are more densely connected than the nodes in 4-way partition. Therefore, a better multi-scale community detection method should find two phenomena. The one is that the method should find hierarchical community structures (both 16-way and 4-way partitions in H 15-2 network), and the other one is that the

multi-scale community detection should find more stable network partition, that is the 16-way partition should span more parameters (the parameters are resolution regulators in ISIMB and RB methods, and are Markov time in Stability and Map methods. see Figure 2) than 4-way partition.

Figure 2 illustrates the multi-scale community structures discovered by four multi-scale community detection methods of ISIMB, Stability, Map and RB methods. We can see that ISIMB and Map methods detect both 16-way and 4-way partitions successfully by varying α in ISIMB and the Markov time in Map. Likewise, For the Stability and RB algorithms, they also successfully find the two partitions. These results mean that all four methods show good performance on the first phenomenon mentioned above. For the second phenomenon, the 16-way partition found by ISIMB persists longest in four methods which indicates that the performance of ISIMB is best. Furthermore, for the Stability and RB algorithms, although they also successfully find the two partitions, the 4-way partition is a more stable partition than the 16-way partition in the two method. In fact, Stability and RB methods are extension of Modularity, so they may suffer from the same resolution limit problem.

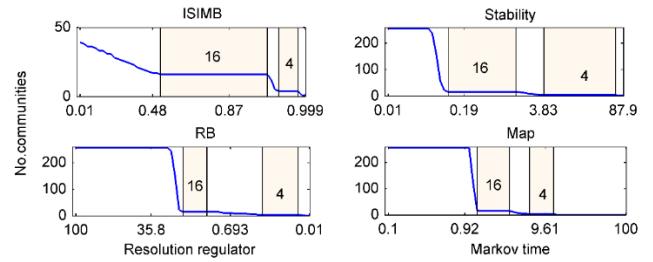


Figure 2. The multi-scale community structures of H 15-2 network discovered by four multi-scale community detection methods

The ISIMB method is then applied to another large synthetic network named LFR benchmark proposed by Lancichinetti et al. In the LFR benchmark [27], some parameters should be decided by a user: (1) the number of nodes N , the average degree $\langle k \rangle$ and maximum degree $\langle \max k \rangle$. (2) Minimum for the micro community sizes $\langle \min c \rangle$ and maximum for the micro community sizes $\langle \max c \rangle$. (4) The number of overlapping nodes (micro communities only) $\langle on \rangle$, the number of memberships of the overlapping nodes (micro communities only) $\langle om \rangle$. (5) Minimum for the macro community size $\langle \min C \rangle$, maximum for the macro community size $\langle \max C \rangle$. (6) Mixing parameter for the macro communities $\langle m \mu 1 \rangle$, mixing parameter for the micro communities $\langle m \mu 2 \rangle$. The parameters used to generate the synthetic network are: $N = 1000$, $k = \max k = 16$, $\min c = \max c = 10$, $on = om = 0$, $\min C = \max C = 50$, $m \mu 1 = 0.03$ and $m \mu 2 = 0.08$. Then, according to the LFR, the generated synthetic network will have two hierarchical community structures with 40 communities and 20 communities respectively. The first level will have stronger community structure than the second level because its communities have stronger local features set by the parameters.

Figure 3 compares the results from the four multi-scale algorithms. We also use the two phenomena mentioned above to measure the four multi-scale algorithms. From this figure, the four algorithms can discover the two hierarchical community structures which are perfectly consistent with the predefined community structures (NMI is equal to 1). But,

only ISIMB method correctly reveals the more importance of the 40-way partition by a longer time of persisted α compared to the 20-way partition. Although the Stability, Map and RB methods also successfully discover the predefined two-layer hierarchical community structures, they regard the 20-way partition as more stable. Overall, comparing with other three method, ISIMB method can find stable and significant community structure from networks more effectively.

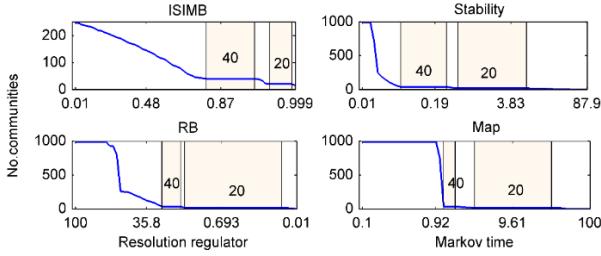


Figure 3. The multi-scale community structures in the LFR synthetic network reveled by four multi-scale methods

The third benchmark network is the Karate network with 34 nodes and 78 edges [28]. The network is widely used to test various module detection methods, and is naturally divided into 2 clusters. Figure 4 shows the multi-scale modular structure detected by four methods. From this figure, we can see that ISIMB captures two stable phases with 4 and 2 communities respectively. The 2-way partition (Figure 5A) agrees exactly with the benchmark (NMI is equal to 1). Figure 5B shows the stable structure of the 4-way partition which has large modularity found by the ISIMB on the Karate network. Apart from the large modularity, the 4-way partition also has higher partition density than 2-way partition. Compared to ISIMB, the Stability method discovers only one stable partition with 2 communities. But the Map does not detect any stable partition, even 2-way partition. The RB method discovers both 4-way and 2-way partition, but the 2-way partition is not consistent with the real partition (NMI is equal to 0.8372).

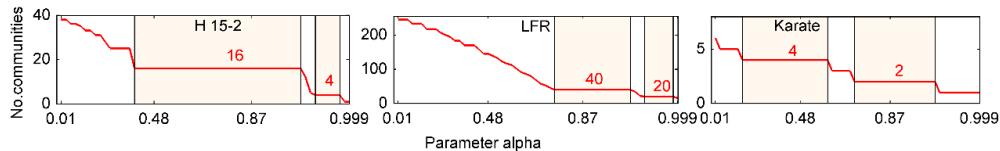


Figure 4. Multi-scale modular structures of Karate network detected by four methods

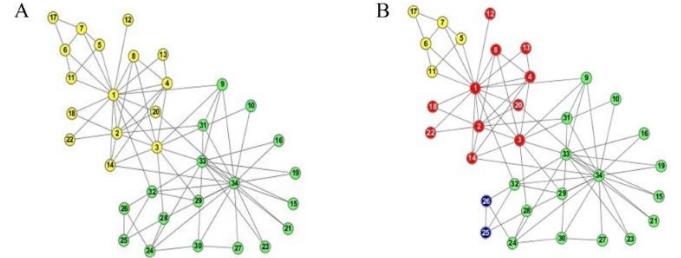


Figure 5. The community structures of 2-way (A) and 4-way (B) partition detected by ISIMB method

3.2 Selection of significant community structure on benchmark networks

The network partition-based method is first applied to detect significant community structure on three benchmark networks mentioned above: H 15-2 network, LFR synthetic network and Karate network, and the results are shown in Figure 6. From the figure we can see that the two significant modular structure with 16 and 4 communities in H 15-2 network are discovered by our method. Likewise, the significant community structure in LFR synthetic network are also revealed by our method. In the Karate network, the plot of the result is the same as the plot in ISIMB in the figure 4, which means that the significant community structure is also detected by our method.

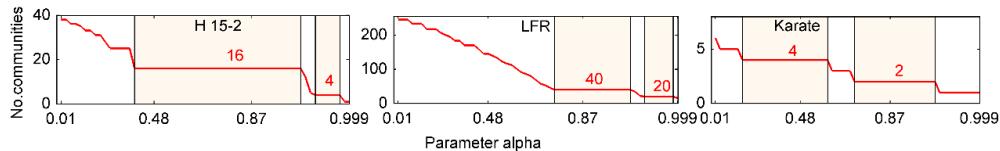


Figure 6. The significant community structure on three benchmark networks

3.3 Selection of significant modular structure on PPI network

Our method is then applied to detect hierarchical protein complexes in protein-protein interaction (PPI) network. The network includes 334 proteins and 1,879 interactions, and the 334 proteins are composed by 8 yeast protein root complexes and 70 protein subcomplexes. In PPI networks, protein complexes generally correspond to module (or community) structures. Protein complexes not only exhibit community structure, but also display multiple scales because they are organized hierarchically. First, we employ ISIMB method to reveal multi-scale of protein complexes. But, a most significant issue for biologists is that how to select crucial partitions among multi-scale network partitions, and these

partitions can effectively cover proteins complexes in biological networks. The multi-scale of the PPI network is analyzed by ISIMB method and the results are shown in Figure 7.

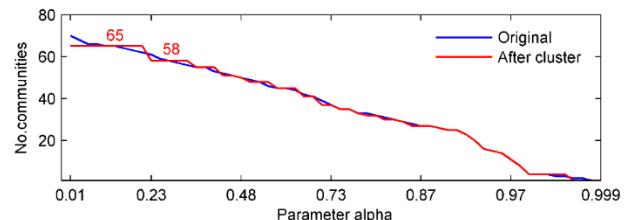


Figure 7. The multi-scale and significant community structure of the PPI network

From the Figure 7, we can see that the original multi-scale structure (blue line) does not reveal significant community structure because they do not have more persistent phenomena. After significant community selection, more stable community structure is discovered (red line in Figure 7), such as, 65-way partition, 58-way partition, and so on. Apart from, the significant partitions (or states) of the PPI network are measured using the real protein complexes, and 4 evaluation indexes [29], $Recall_c$ (see equation 11), $Precision_c$ (see equation 12), $Recall_p$ (see equation 13) and $Precision_p$ (see equation 14) are computed. The $Recall_c$ and $Precision_c$ are defined at complex level and the other two are defined at complex-protein pair level. Let S be a cluster, C be a reference complex, V_S be the set of proteins contained in S and V_C be the set of proteins contained in C . The match score is defined by equation 10.

$$match_score(S, C) = \frac{|V_S \cap V_C|}{|V_S \cup V_C|} \quad (10)$$

If $match_score(S, C) \geq 0.5$, we say S and C match each other. Given a set of reference complexes $C = \{C_1, C_2, \dots, C_n\}$ and a set of predicted complexes $\rho = \{S_1, S_2, \dots, S_m\}$, the 4 evaluation indexes are defined as follows:

$$Recall_c = \frac{|\{C_i \mid C_i \in C \wedge \exists S_j \in \rho, S_j \text{ matches } C_i\}|}{|C|} \quad (11)$$

Table 1. The performance on 4 evaluation indexes in various of methods

Methods	No. communities	Recall _c	Precision _c	Recall _p	Precision _p	Total
ISIM	65	0.5190	0.5538	0.2193	0.5539	1.8460
	58	0.4937	0.5690	0.2415	0.5629	1.8670
ClusterONE	52	0.4231	0.5577	0.2108	0.4048	1.6000
CFinder	504	0.5128	0.1667	0.2246	0.1858	1.0899
CMC	45	0.1923	0.3111	0.1472	0.2361	0.8868
MCODE	17	0.1795	0.6471	0.1070	0.3702	1.3037

4. CONCLUSION

Many complex systems in nature can be described by networks, and then be resolved easily. One of significant characterization in complex networks is community or modular structure. Comparing with single-scale community structure, detecting multi-scale community structure is important for understanding the structure and dynamics of complex network. Selection of crucial partitions among multi-scale network partitions can be applied in many fields. For example, multi-scale module structure in PPI network can reveal evolution of protein complex or proteins with similar function, and selection of crucial partitions among multi-scale community structure in PPI networks can find more stable protein complexes or groups of proteins with similar functions. Furthermore, take social network as an example, multi-scale community structure should be able to reveal evolution of people in friendship, collaboration or others, and selection of crucial partitions among multi-scale community structure in social network can find more stable friendship or collaboration. Although some multi-scale community structure detection methods have been proposed, few algorithms for selecting significant network partition are designed. In the work, we

$$Precision_c = \frac{|\{S_j \mid S_j \in \rho \wedge \exists C_i \in C, C_i \text{ matches } S_j\}|}{|\rho|} \quad (12)$$

$$Recall_p = \frac{\sum_{i=1}^n \max\{overlap(C_i, S_j) \mid \forall S_j \text{ that matches } C_i\}}{\sum_{i=1}^n |C_i|} \quad (13)$$

$$Precision_p = \frac{\sum_{j=1}^m \max\{overlap(S_j, C_i) \mid \forall C_i \text{ that matches } S_j\}}{\sum_{j=1}^m |S_j|} \quad (14)$$

$$\text{where, } overlap(C_i, S_j) = |V_{C_i} \cap V_{S_j}|.$$

In order to verify the performance of our method, four single-scale methods (such as ClusterONE, CFinder, CMC and MCODE) are employed to discover protein complexes. Comparing these methods, the 2 stable states are also the winner in the 4 indexes. For example, the values of indexes of ClusterONE, CFinder [30], CMC and MCODE [31] are 1.6, 1.0899, 0.8868 and 1.3037 respectively (See Table 1). These results indicate that the new method for selecting significant partitions is very effectively because the 2 stable partitions are consistent with real protein complexes, and also that the methods for revealing multi-scale community is successful.

introduce a network partition-based cluster method for selecting significant community structure which can help one to understand complex system. Comparing with other single-scale and multi-scale methods, our method can select significant network partitions effectively.

Although the proposed algorithm is efficient, it has problems for further discussion. For example, the performance of ISIMB is measured on both synthetic networks and real-world networks, but the measured method is qualitative (by two phenomenons), not is quantitative. So, a quantitative method should be proposed to measure different multi-scale community structure detection algorithms. The second problems are that the proposed method in this work relies on ISIMB method totally. Therefore, the proposed method may inherit some drawbacks from ISIMB method. Such as, the proposed method cannot select crucial partitions from very sparse networks.

ACKNOWLEDGMENTS

The authors acknowledge the supports from the National Language Committee scientific research projects of China (No.

YB135-50), the National Natural Science Foundation of China (No. 61806007) and Science and Technology Development Plan Project of Henan (No. 182102310920).

REFERENCES

- [1] Radicchi F, Castellano C, Cecconi F, Loreto V, Parisi D. (2003). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America* 101(9): 2658-2663. <https://doi.org/10.1073/pnas.0400054101>.
- [2] Girvan M, Newman MEJ. (2002). Community structure in social and biological networks *Proceedings of the National Academy of Sciences* 99(12): 7821-7826. <https://doi.org/10.1073/pnas.122653799>
- [3] Newman MEJ. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103(23): 8577-8582. <https://doi.org/10.1073/pnas.0601602103>
- [4] Newman MEJ, Girvan M. (2004). Finding and evaluating community structure in networks. *Physical Review E* 69(2): 026113. <https://doi.org/10.1103/physreve.69.026113>
- [5] Guimera R, Sales-Pardo M, Amaral LAN. (2004). Modularity from fluctuations in random graphs and complex networks. *Physical Review E* 70(2): 025101. <https://doi.org/10.1103/PhysRevE.70.025101>
- [6] Duch J, Arenas A. (2005). Community detection in complex networks using extremal optimization. *Physical Review E* 72(2): 027104. <https://doi.org/10.1103/PhysRevE.72.027104>
- [7] Richardson T, Mucha PJ, Porter MA. (2009). Spectral tripartitioning of networks. *Physical Review E* 80(3): 036111. <https://doi.org/10.1103/PhysRevE.80.036111>
- [8] Jiao QJ, Huang Y, Shen HB. (2015). Community mining with new node similarity by incorporating both global and local topological knowledge in a constrained random walk. *Physica A: Statistical Mechanics and its Applications* 424(2015): 363-371. <https://doi.org/10.1016/j.physa.2015.01.022>
- [9] Ahn YY, Bagrow JP, Lehmann S. (2010). Link communities reveal multiscale complexity in networks. *Nature* 466(7307): 761-764. <https://doi.org/10.1038/nature09182>
- [10] Fortunato S, Barthelemy M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences* 104(1): 36-41. <https://doi.org/10.1073/pnas.0605965104>
- [11] Rosvall M, Bergstrom CT. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105(4): 1118-1123. <https://doi.org/10.1073/pnas.0706851105>
- [12] Schaub MT, Jean-Charles D, Yaliraki SN, Mauricio B, Olaf S. (2012). Markov dynamics as a zooming lens for multiscale community detection: Non clique-like communities and the field-of-view limit. *PLoS ONE* 7(2): e32210. <https://doi.org/10.1371/journal.pone.0032210>
- [13] Newman MEJ. (2011). Communities, modules and large-scale structure in networks. *Nature Physics* 8(1): 25-31. <https://doi.org/10.1038/nphys2162>
- [14] Simon HA. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society* 106(6): <https://doi.org/10.1109/MCS.2007.384127> 467-482.
- [15] Le Martelot E, Hankin C. (2013). Fast multi-scale detection of relevant communities in large-scale networks. *The Computer Journal* 2013: bxt002. <https://doi.org/10.1093/comjnl/bxt002>
- [16] Nepusz T, Yu H, Paccanaro A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nature methods* 9(5): 471-472. <https://doi.org/10.1038/nmeth.1938>
- [17] Mewes HW, Amid C, Arnold R, Frishman D, Güldener U, Mannhaupt G. (2004). MIPS: Analysis and annotation of proteins from whole genomes. *Nucleic Acids Research* 34: 169-72. <https://doi.org/10.1093/nar/gkj148>
- [18] Delvenne JC, Yaliraki SN, Barahona M. (2010). Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences* 107(29): 12755-12760. <https://doi.org/10.2307/25708610>
- [19] Delmotte A, Tate EW, Yaliraki SN, Barahona M. (2011). Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin-myosin light chain interaction. *Physical Biology* 8(5): 055010. <https://doi.org/10.1088/1478-3975/8/5/055010>
- [20] Schaub MT, Lambiotte R, Barahona M. (2012). Encoding dynamics for multiscale community detection: Markov time sweeping for the map equation. *Physical Review E* 86(2): 026112. <https://doi.org/10.1103/PhysRevE.86.026112>
- [21] Arenas A, Fernandez A, Gomez S. (2008). Analysis of the structure of complex networks at different resolution levels. *New Journal of Physics* 10(5): 053039. <https://doi.org/10.1088/1367-2630/10/5/053039>
- [22] Chen C, Fushing H. (2012). Multiscale community geometry in a network and its application. *Physical Review E* 86(4): 041120. <https://doi.org/10.1103/PhysRevE.86.041120>
- [23] Reichardt J, Bornholdt S. (2006). Statistical mechanics of community detection. *Physical Review E* 74(1): 016110. <https://doi.org/10.1103/PhysRevE.74.016110>
- [24] Jiao QJ, Huang Y, Shen HB. (2016). A new multi-scale method to reveal hierarchical modular structures in biological networks. *Molecular Biosystems* 12: 3724-3733. <https://doi.org/10.1039/C6MB00617E>
- [25] Karrer B, Levina E, Newman MEJ. (2008). Robustness of community structure in networks. *Physical Review E* 77(4): 046119. <https://doi.org/10.1103/PhysRevE.77.046119>
- [26] Arenas A, Díaz-Guilera A, Pérez-Vicente CJ. (2006). Synchronization reveals topological scales in complex networks. *Physical Review Letters* 96(11): 114102. <https://doi.org/10.1103/physrevlett.96.114102>
- [27] Lancichinetti A, Fortunato S. (2009). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E* 80(1): 016118. <https://doi.org/10.1103/PhysRevE.80.016118>
- [28] Zachary W.W. (1977). An information flow model for conflict and fission in small groups. *Journal of anthropological Research* 1977: 452-473. <https://doi.org/10.1086/jar.33.4.3629752>
- [29] Liu G, Wong L, Chua HN. (2009). Complex discovery from weighted PPI networks. *Bioinformatics* 25(15):

- 1891-1897.
<https://doi.org/10.1093/bioinformatics/btp311>
- [30] Adamcsek B, Palla G, Farkas IJ, Derenyi I, Vicsek T. (2006). CFinder: locating cliques and overlapping modules in biological networks. Bioinformatics 22(8): 1021-1023.
- <https://doi.org/10.1093/bioinformatics/btl039>
- [31] Bader GD, Hogue CW. (2003). An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4(1): 2.
<https://doi.org/10.1186/1471-2105-4-2>