



Research article

Motif adjacency matrix and spectral clustering of directed weighted networks

Yike Wang, Gaoxia Wang*, Ximei Hou and Fan Yang

College of Science and Three Gorges Mathematics Research Center, China Three Gorges University, Yichang, Hubei, 443002, China

* **Correspondence:** Email: gaoxiawang@163.com.

Abstract: In the spectral clustering methods, different from the network division based on edges, some research has begun to divide the network based on network motifs; the corresponding objective function of partition also becomes related to the motif information. But, the related research on the directed weighted network needs to be further deepened. The weight of the network has a great influence on the structural attributes of the network, so it is necessary to extend the motif-based clustering to the weighted network. In this paper, a motif-based spectral clustering method for directed weighted networks is proposed. At the same time, this paper supplements the method of obtaining matrix expressions of the motif adjacency matrix in directed unweighted networks and provides a method to deal with the weight of networks, which will be helpful for the application research of motifs. This clustering method takes into account the higher-order connectivity patterns in networks and broadens the applicable range of spectral clustering to directed weighted networks. In this method, the motif-based clustering of directed weighted networks can be transformed into the clustering of the undirected weighted network corresponding to the motif-based adjacency matrix. The results show that the clustering method can correctly identify the partition structure of the benchmark network, and experiments on some real networks show that this method performs better than the method that does not consider the weight of networks.

Keywords: directed weighted networks; motif adjacency matrix; spectral clustering; mapping; motif conductance

Mathematics Subject Classification: 62H30, 91D30

1. Introduction

Network science can help us to understand complex systems, and a common task of the analysis of networks is clustering [1]. The process of dividing a set into groups of similar elements is called clustering. Spectral clustering [2] is based on the spectral graph theory. By selecting some partition criterion [3,4], the objective function can be generated. The minimization of the objective function is closely related to the eigenvalues and eigenvectors of the Laplacian matrix [5]. Eventually, the problem becomes using the eigenvectors of the Laplacian matrix to divide the network. Spectral clustering has become a very popular clustering algorithm [6,7]. Unfortunately, most spectral clustering algorithms only operate on undirected networks because the adjacency matrix of a directed network is asymmetric. A common approach to clustering in directed networks is to build a symmetric adjacency matrix from the original asymmetric matrix [8].

The network motif [9] is a pattern of some kind of interconnection found in a complex network that is significantly higher than the number of interconnections in a random network. For the 13 three-node motifs of directed networks, according to the types of motifs, they can be divided into triangular motifs (closed subgraphs) and two-hop-path motifs (open subgraphs). In recent years, the traditional clustering method has begun to consider the motif of networks [10]. On the basis of ensuring the internal similarity of the subgraph [11], the traditional graph division should make sure the number of edges to be cut is as small as possible. While motif-based clustering takes into account the higher-order structure of the networks. The motif-based graph division should ensure that the nodes in the cluster should be involved in more instances of the motif and avoid cutting the instances of the motif as much as possible. Thus, the objective function of the cluster has also changed. Benson et al. [12] converted conductance to motif-based conductance in the partition of directed unweighted networks and generalized the motif-based spectral clustering to the directed networks by the motif adjacency matrix.

The motif adjacency matrix, which is a symmetric matrix, contains the information of a certain higher-order link pattern in the networks. For the three-node motif, Benson et al. [12] obtained the matrix expressions of motif adjacency matrices of triangular motifs (closed subgraphs) based on matrix operations. On this basis, the matrix operational method to obtain the matrix expressions of motif adjacency matrices for the remaining two-hop-path motifs (open subgraphs) is supplemented in this paper, and we use the lower triangular matrix to simplify the calculation difficulty. Therefore, the motif adjacency matrix expressions of all motifs in the directed network can be obtained by matrix operations.

Spectral clustering methods based on motifs are becoming increasingly important [13–15]. It is natural to search for motif-based applications for the weighted cases since weights are one of the important factors to describe the association between nodes and have a significant impact on the clustering effect of networks. For the directed weighted networks, most of the methods are to weight the motif [16–18]. In addition, a directed weighted network can be mapped to a directed unweighted network with multi-links which has the same adjacency matrix with it for analysis [19]. In this paper, a weighted network is considered as an unweighted network with multi-links, and by using the adjacency matrix as a bridge, the network can be mapped to an unweighted multilayer network.

This paper supplements the method of constructing matrix expressions of the motif adjacency matrix as we use the lower triangular matrix to simplify the expressions of motif adjacency matrices. And, this paper proposes a motif-based clustering algorithm for directed weighted networks (WMCA). The WMCA is tested on a benchmark network. Selecting two cluster evaluation indexes, the clustering effect of the WMCA method is compared with that of the method in [12] without considering the network weight on three real networks.

2. Introduction to motif and motif-based clustering

2.1. The three-node motif of directed networks

The motif used in this paper is a connected subgraph with three vertices. A certain connection relation of three nodes can be encoded by a binary coding matrix $H_{3 \times 3}$. Without considering the position order of nodes, H corresponds to a form of a connected subgraph, and the form of a connected subgraph is defined as a class of motif M . Figure 1 shows the types of three-node motifs in directed networks and their corresponding coding matrices. In the coding matrices, select the left node as node 1, the upper node as node 2 and the lower node as node 3.

The composition of the motifs is independent of the position order of the nodes, so the representation of the coding matrix is not unique. A connected subgraph with three vertices may have multiple coding matrices; the encoding matrices belonging to the same type of motif are permutation similarity. Take $M7$ as an example; its encoding matrix has three kinds, as shown in Figure 2(a). In the network shown in Figure 2(b), there are two motif instances of $M7$. At the same time, motif instances of other types in the network are shown in Figure 2(c).

2.2. Motif-based spectral clustering of directed unweighted networks

Consider a directed unweighted network with n nodes and a motif M ; $W_M = (w_{ij}^M)_{n \times n}$ is called the motif adjacency matrix [12], which w_{ij}^M represents the number of instances of M in which i and j participate jointly. The adjacency matrix of the directed unweighted network is not a symmetric matrix, while the motif adjacency matrix is a symmetric matrix, and its element values reflect the similarity between node pairs.

Benson et al. [12] proposed a general motif-based clustering algorithm (MCA). And, they prove that, when the specific conductance is selected as the objective function, the conductance of the directed unweighted network is equal to the conductance of the undirected weighted network corresponding to the motif adjacency matrix, which is called the motif conductance. The node group that makes the motif conductance minimum is the optimal division. Then the partition of the directed unweighted network is converted to the partition on the undirected weighted network that the motif adjacency matrix corresponds to. Thus, based on the motif adjacency matrix to form the normalized motif Laplacian matrix, classical spectral clustering method can be used. And the motif conductance is defined as:

$$\varphi_M(G)S = \frac{cut_M(G)S, \bar{S}}{\min(vol_M(G)S, vol_M(G)\bar{S})} \quad (1)$$

where $cut_M(G)S, \bar{S}$ represents the number of the motif instances with at least one endpoint in S and one endpoint in \bar{S} , and $vol_M(G)S$ represents the number of instances of M formed by nodes in S . The lower the conductance, the better the clustering.

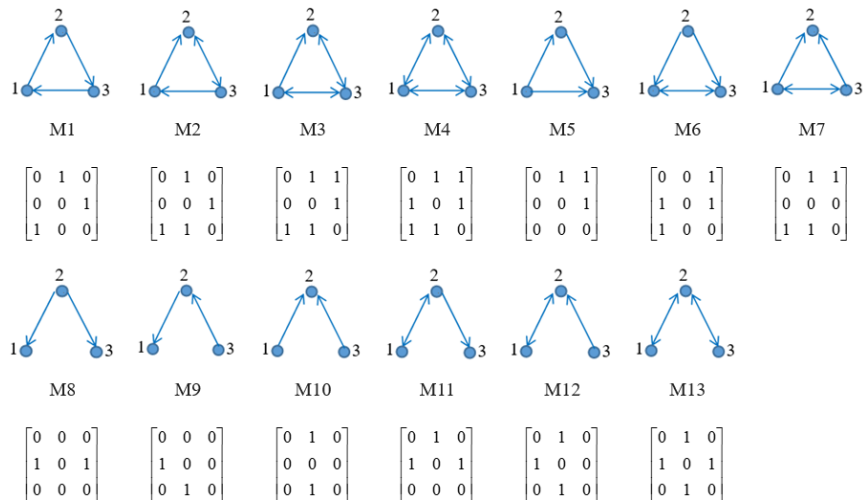


Figure 1. The types of three-node motifs and the coding matrices.

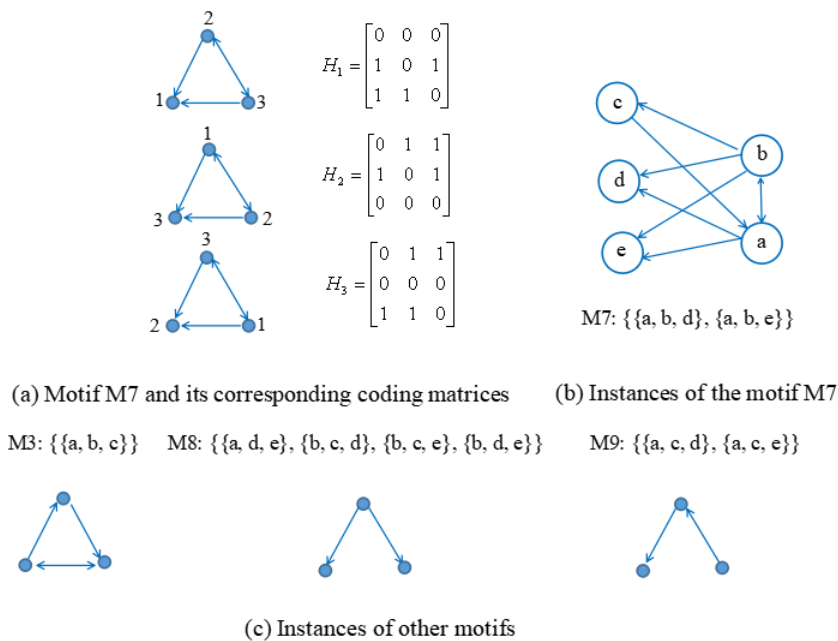


Figure 2. Instances of motif M7 and other motifs in a network.

3. Motif-based spectral clustering in directed weighted networks

This section introduces the motif-based spectral clustering method in directed weighted networks. The main process is shown in the Figure 3. In this section, the method of forming the expression motif adjacency matrix in directed unweighted networks is firstly introduced. Then, the method to deal with the weight of the network is introduced in order to form the motif adjacency matrix of the directed weighted network. Finally, based on the motif adjacency matrix, the directed weighted network can be clustered.

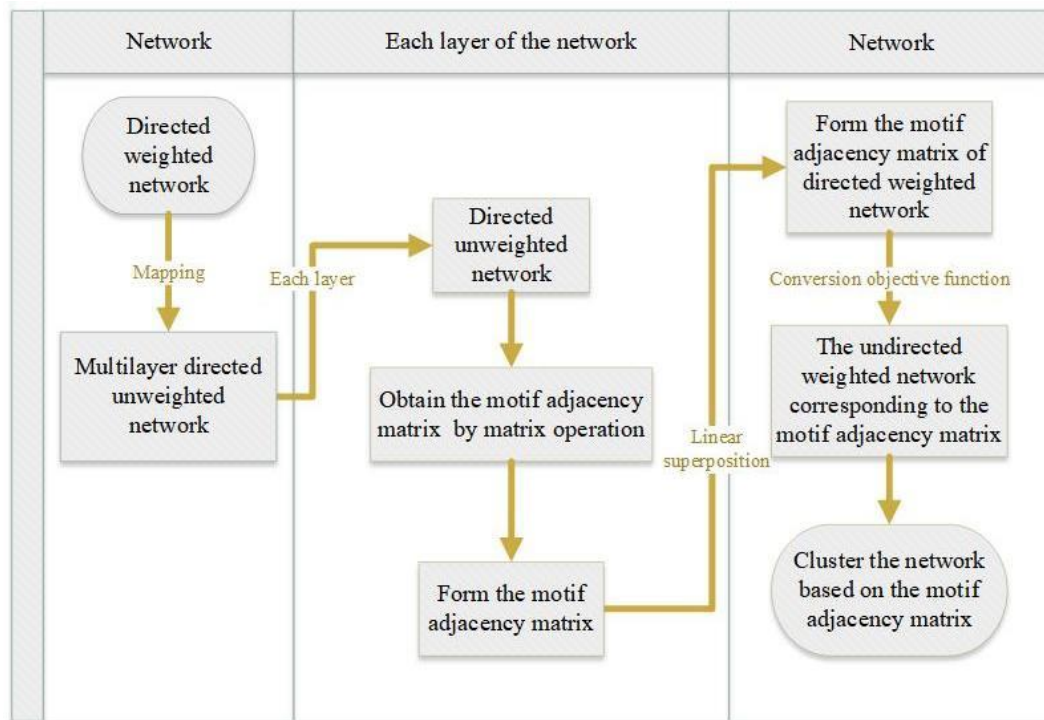


Figure 3. Flow chart of motif-based spectral clustering method in directed weighted network.

3.1. The expression of motif adjacency matrix in directed unweighted network

When the spectral clustering method is extended to the directed network, a symmetric matrix containing node similarity information is needed to form the Laplacian matrix. The adjacency matrix of directed network is asymmetric, while the motif adjacency matrix of the network is symmetric, and contains the information of higher-order connectivity between nodes. Therefore, spectral clustering can be realized on the directed weighted network through the motif adjacency matrix. So, forming motif adjacency matrix of the directed network becomes an important tool to the motif-based clustering. Benson et al. [12] give a method of obtaining the expressions of the motif adjacency matrix for triangular motifs (7 types) through matrix operation, and use the k-clique enumeration algorithm to obtain the motif adjacency matrix of two-hop paths motifs (6 types). This paper supplements the method of obtaining the motif adjacency matrix by matrix operation for the two-hop paths motifs, so that the motif adjacency matrix of any kind of motif can be obtained by matrix operation.

For a directed unweighted network G with adjacency matrix A , let B and U be the adjacency matrix of graphs formed by bidirectional edges and unidirectional edges, respectively [12]. Formally, $B = A \circ A^T$, $U = A - B$, where “ \circ ” means the Hadamard product of matrices. In mathematics, the Hadamard product is a special binary operation, where each element i, j is the product of elements i, j of the two original matrices [20]. That is, suppose $A = (a_{ij})_{m \times n}$ and $B = (b_{ij})_{m \times n}$, then the

Hadamard product of them is $C = A \circ B = (a_{ij} * b_{ij})_{m \times n}$. When G is regarded as an undirected graph

\bar{G} , let \bar{F} be the adjacency matrix corresponding to the complement graph of \bar{G} . An element value of 1 in \bar{F} indicates that there is no edge connected between the corresponding node pair in the graph G , so

\bar{F} can be called edge-missing matrix. In terms of the matrices U , B , and \bar{F} , matrix expressions of 13 motif adjacency matrices can be established. Let X, Y, Z represent one of U, B and \bar{F} . The central computational kernel in these computations is $XY \circ Z$, which means these operations are similar in form, that is, multiply two matrices and then take the Hadamard product with a third matrix. Furthermore, let B_L and \bar{F}_L denote the lower triangular matrices of B and \bar{F} , respectively, the matrix in some terms of the expression can be replaced by B_L and \bar{F}_L to simplify the calculation. For all three-node motifs, the simplified formulations that form the motif adjacency matrix W_M based on matrix operations in directed unweighted networks are shown in Table 1.

Table 1. Simplified matrix formulations of the motif adjacency matrix.

Motif	Matrix computations	W_M
M1	$C = U^2 \circ U^T$	$C + C^T$
M2	$C = B_L U \circ U^T + B_L^T U \circ U^T + U B_L \circ U^T + U B_L^T \circ U^T + U^2 \circ B$	$C + C^T$
M3	$C = B_L B \circ U + B_L^T B \circ U + B_L U \circ B + B_L^T U \circ B + U B_L \circ B + U B_L^T \circ B$	$C + C^T$
M4	$C = B_L B \circ B + B_L^T B \circ B$	C
M5	$C = U^2 \circ U + U U^T \circ U + U^T U \circ U$	$C + C^T$
M6	$C = U B_L \circ U + U B_L^T \circ U + B_L U^T \circ U^T + B_L^T U^T \circ U^T + U^T U \circ B$	C
M7	$C = U^T B_L \circ U^T + U^T B_L^T \circ U^T + B_L U \circ U + B_L^T U \circ U + U U^T \circ B$	C
M8	$C = \bar{F}_L U^T \circ U^T + \bar{F}_L^T U^T \circ U^T + U \bar{F}_L \circ U + U \bar{F}_L^T \circ U + U^T U \circ \bar{F}$	C
M9	$C = U U \circ \bar{F} + U \bar{F}_L \circ U^T + U \bar{F}_L^T \circ U^T + \bar{F}_L U \circ U^T + \bar{F}_L^T U \circ U^T$	$C + C^T$
M10	$C = U U^T \circ \bar{F} + U^T \bar{F}_L \circ U^T + U^T \bar{F}_L^T \circ U^T + \bar{F}_L U \circ U + \bar{F}_L^T U \circ U$	C
M11	$C = B_L U \circ \bar{F} + B_L^T U \circ \bar{F} + U \bar{F}_L \circ B + U \bar{F}_L^T \circ B + \bar{F}_L B \circ U^T + \bar{F}_L^T B \circ U^T$	$C + C^T$
M12	$C = B_L U^T \circ \bar{F} + B_L^T U^T \circ \bar{F} + U^T \bar{F}_L \circ B + U^T \bar{F}_L^T \circ B + \bar{F}_L B \circ U + \bar{F}_L^T B \circ U$	$C + C^T$
M13	$C = B_L B \circ \bar{F} + B_L^T B \circ \bar{F} + B \bar{F}_L \circ B + B \bar{F}_L^T \circ B + \bar{F}_L B \circ B + \bar{F}_L^T B \circ B$	C

In the first part of the Supplementary, the Matlab algorithm for obtaining the expression of the motif adjacency matrix for M8-M13 is given.

3.2. The conversion of directed weighted networks

Consider a directed weighted network, the data of all nodes in the network can be represented by its adjacency matrix $A = (a_{ij})_{n \times n}$, where n represents the number of nodes, a_{ij} represents the edge weight of node i to j . When the value of a_{ij} is a nonnegative integer, the directed weighted network

can be converted to a directed unweighted multilayer network.

Definition 1: let A be a nonnegative matrix, define a mapping $\phi(\cdot)$ as follows:

$$\phi(A) = A_l \begin{cases} = \psi(A), l = 1 \\ = \psi(A - \sum_{j=1}^{l-1} A_j), 1 < l \leq \max(A) \end{cases} \tag{2}$$

where $\psi(\cdot)$ means mapping non-zero elements in a matrix to 1, and $\max(\cdot)$ means taking the maximum value of the elements of a matrix.

The Mapping $\phi(\cdot)$ maps matrix A to a Boolean matrix sequence $A_l, l = 1, 2, \dots, \max(A)$. If the element values in A are all nonnegative integers, then has $A = \sum_{l=1}^{\max(A)} A_l$. If such matrix A is the adjacency matrix of the directed weighted network G , then each matrix A_l can correspond to a directed unweighted network G_l , and A_l is the adjacency matrix of the G_l . So, take the adjacency matrix as bridge, the mapping $\phi(\cdot)$ can map the network G to a multilayer network, $G = \cup_l G_l$. Thus, the motif adjacency matrix in each layer of the multilayer network can be obtained by the method given in Table 1. Naturally, the motif adjacency matrix of the directed weighted network is the sum of the motif adjacency matrix of each layer.

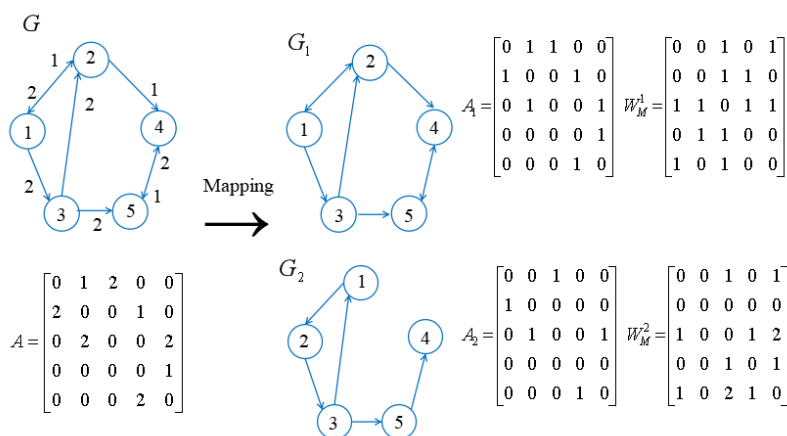


Figure 4. Mapping processing of directed weighted networks.

The directed weighted network is mapped as a directed unweighted multilayer network as shown in Figure 4. Figure 4 shows a directed weighted network G and its adjacency matrix A . Select M9 as the interested motif, and show each layer network of the multilayer network and its corresponding motif adjacency matrix corresponding to M9 in Figure 4. First-layer network is represented by G_1 and its adjacency matrix is A_1 , second-layer network is represented by G_2 and its adjacency matrix is A_2 , and their motif adjacency matrices corresponding to M9 are denoted by W_M^1 and W_M^2 respectively. In the first layer network, nodes $\{1, 3, 5\}$, $\{2, 3, 4\}$ form an instance of M9 respectively; in the second layer network, nodes $\{1, 3, 5\}$, $\{3, 4, 5\}$ form an instance of M9 respectively. According to the definition of the motif adjacency matrix, the motif adjacency matrix of each layer network is

W_M^1 and W_M^2 respectively. Then the motif adjacency matrix corresponding to M9 of G is $W_M = W_M^1 + W_M^2$.

In the first part of the Supplementary, a program is given to obtain the expression of motif adjacency matrix for weighted networks using Matlab algorithm.

3.3. Motif-based clustering algorithm in directed weighted networks

For directed weighted networks, computing the conductance of motif-based conductance can be translated into computing the conductance on the graph corresponding to the motif adjacency matrix. See second part of the Supplementary for proof. In this paper, choose the motifs which have more instances in the network as the motifs of interest. In the network, the motifs with more instances are considered as important motifs, which can represent the basic unit of the network structure and have a certain influence on the structure of the network, while the motifs with fewer instances in the network have a limited influence on the network structure.

Consider the clustering based on the set of motifs $\{M_1, M_2, \dots, M_q\}$ for q different motifs. Let $W_{M_j}, j = 1, 2, \dots, q$ be the motif adjacency matrix of M_j . And let α_j be the weight of motif M_j . The value of α_j is the ratio of the number of M_j to the total number of q interested motifs. Obviously, $\sum_{j=1}^q \alpha_j = 1$ and $\alpha_j > 0$. Then the motif adjacency matrix of the network is $W_M = \sum_{j=1}^q \alpha_j W_{M_j}$. And the measures of cut and volume are simply linearly weighted sums. After forming the adjacency matrix, spectral clustering can be carried out according to the algorithm proposed by Benson et al. [12].

The steps of the motif-based clustering algorithms in directed weighted networks (WMCA) are as follows:

Motif-based clustering algorithm in directed weighted networks (WMCA)	
Step 1	Form the motif adjacency matrix. Give a directed weighted network and q motifs of interest to form the motif adjacency matrix W_M .
Step 2	Form the normalized motif Laplacian matrix. Construct the motif Laplacian matrix and normalize the motif Laplacian matrix $L_M = D^{-\frac{1}{2}}(D - W_M)D^{-\frac{1}{2}}$, where D is the diagonal matrix and is defined as $D_{ii} = \sum_j^n w_{ij}^M$.
Step 3	Compute the spectral order σ of nodes. σ is sorted by $D^{-\frac{1}{2}}z$, and z is the eigenvector corresponding to the second smallest eigenvalue of L_M , that is, σ_i is the node corresponding to the i th smallest value of $D^{-\frac{1}{2}}z$.
Step 4	Find the prefix set σ that minimizes the motif conductance, $S = \operatorname{argmin}_r \varphi_M(G)S_r$, where $S_r = \{\sigma_1, \dots, \sigma_r\}$.
Step 5	Find the smaller portion. If $ S < \bar{S} $, S is the module obtained by the clustering, otherwise \bar{S} is.

The algorithm divides the nodes into two clusters: S and \bar{S} . For the motif conductance, $\varphi_M(G)S = \varphi_M(G)\bar{S}$, so any set of nodes can be interpreted as a cluster.

After obtaining the motif adjacency matrix of the directed weighted network, if multiple

clustering is required, one can obtain multiple clusters by combining the method of Ng et al. [6]: Take eigenvectors of k smallest eigenvalues z_1, z_2, \dots, z_k for L_M in Step 2, and make $y_{ij} = \frac{z_{ij}}{\sqrt{\sum_{j=1}^k z_{ij}^2}}$ to form

the matrix Y . Then take the i th row of the matrix Y to embed node i into R^k , and on the embedded nodes k-means clustering can be run.

In Sections 3.1 and 3.2, the method to deal with the weight of directed weighted networks is proposed, and the method to obtain the motif adjacency matrix expression by matrix operation is supplemented and simplified. In Section 3.3, combine the motif adjacency matrix with the spectral clustering method to get two or more clusters in directed weighted networks. The motif adjacency matrix which contains the higher-order link information between nodes can also be used in other graph clustering of directed weighted networks.

4. Experiments

In this part, the clustering effect of WMCA method proposed in this paper is tested on a benchmark network with known partitions. Then, on the three real networks, the WMCA method and the MCA method (without considering the weight of the networks) proposed by Benson et al. [12] are respectively selected to cluster the networks and obtain multiple clusters. Two clustering evaluation metrics: Silhouette Coefficient (SC) [21] and Davies-Bouldin index (DBI) [22] are selected to compare the clustering effect of the two methods.

4.1. Experiments on a benchmark social network with two clusters

The performance of WMCA method is tested on the weighted Zachary karate club network [23], which is often used as a benchmark network to detect the effect of network partitioning. This network has two known divisions, and the weight of the edges represents the number of contexts in which interaction took place between the two individuals involved. When the network is used, the undirected edge is regarded as a bidirectional edge, so, there are only M13 and M4 motifs in the network, and the number of instances is 880 and 115, respectively. Table 2 presents some information on the Zachary karate club network.

Table 2. Information on the weighted Zachary karate club network.

Size	Number of edges	Maximum edge weight	Motifs of interested
34	156	7	M13, M4

In Figure 5(a) shows the network diagram of the Zachary karate club network, where the thickness of the line reflects the size of the weight. The known division of the network is shown in Figure 5(b), the color of a node indicates its category. Use the WMCA method based on M13, M4 and the combination of them respectively to cluster the network, the clustering results of weighted network is obtained. Compared with the known partition results, in the M13 based clustering, node 9 is wrongly divided, as shown in Figure 5(c), while in the M4 based clustering, node 10 is divided into another group as shown in Figure 5(d). The clustering based on the combination of motif M4 and M13 gives the correct classification as shown in Figure 5(b).

Next, a further analysis of node 9 in the M13-based cluster and node 10 in the M4-based cluster is performed. The local network composed of first-order neighbors of the node is called first-order neighborhood network, and the first-order neighborhood network of nodes 9 and 10 are shown in Figure 6 respectively. The colors of the nodes in Figure 6 are the same as those in Figure 5. The set of white nodes represents cluster S and the set of blue nodes represent cluster \bar{S} .

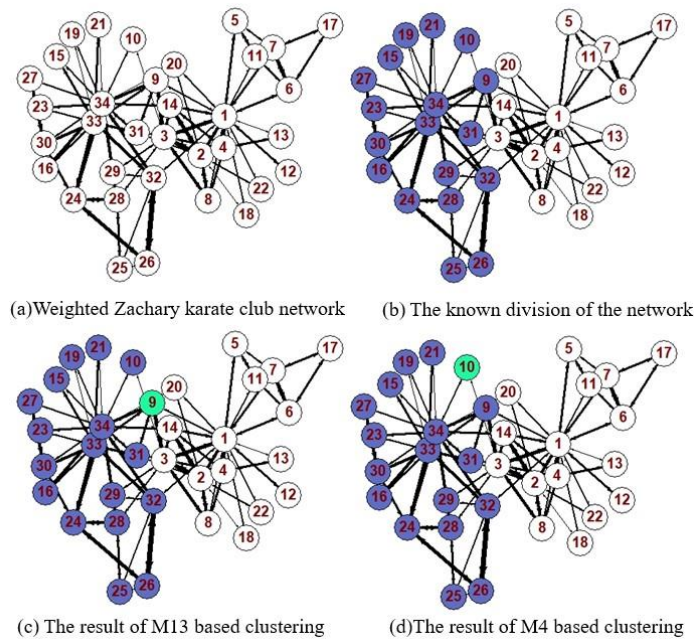


Figure 5. Weighted Zachary karate club network and the results of motif-based clustering.

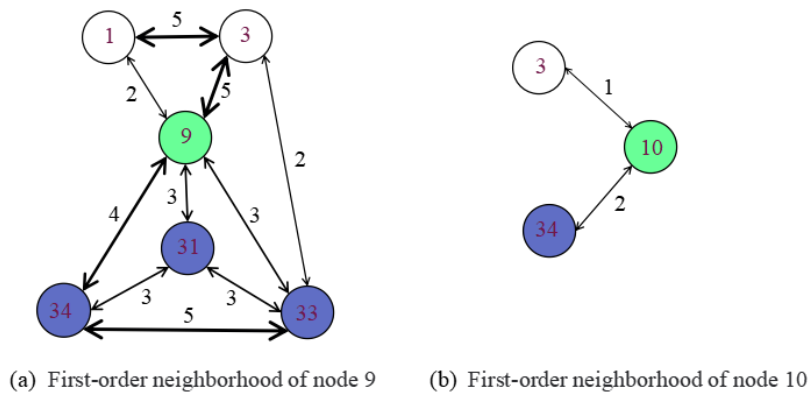


Figure 6. First-order neighborhood network of nodes.

In Figure 6(a), considering the combination of nodes belonging to a cluster and node 9, the instances with node 9 participating of the M13 and M4 and their number are shown in Table 3, respectively. From Table 3 can know that: the strength of this higher-order link form between nodes 9, 1 and 3 is higher than that between nodes 9, 31, 33, and 34 when considering it based on the M13. And this is the opposite when considered based on M4. In Figure 6(b), node 10 is only linked to node 3, 34, and they form an instance of motif M13, and this does not impact to M4-based clustering. When the

degree of node is considered, the similarity between node 10 and nodes in cluster \bar{S} is higher; while considering M4-based clustering, the similarity between node 10 and nodes in the two clusters is the same because node 10 does not form an instance of M4.

Table 3. The instances of M13 and M4 with node 9 participating.

M13		M4	
The nodes	Number	The nodes	Number
1, 3, 9	3	1, 3, 9	2
9, 31, 34	0	9, 31, 34	3
9, 31, 33	0	9, 31, 33	3
9, 33, 34	1	9, 33, 34	3

4.2. Experiments on some real networks

Select three networks in The Koblenz network collection [24]. The information of networks is as follows:

(1) Seventh graders network: This directed weighted network contains ratings of closeness between students at a school in Victoria. Students were asked to nominate their favorite classmates for three different activities. In the network, node represents students; If student A chooses student B to participate in an activity, there is an edge from A to B. The weights show how often this happens.

(2) Highschool network: This directed weighted network includes friendships between boys at a small high school in Illinois. A node represents a boy, boy A chooses B as his friend, then there is an edge from A to B. The weights show how often this happens.

(3) Caenorhabditis elegans (neural) network: This is a weighted network representing the neural network of Caenorhabditis elegans. In this version, the edges of the network are bidirectional, allows no multiple edges, and the given edge weights are the sum of the original edge weights.

Calculate the number of various instances of motifs in these networks. Table 3 gives some basic information about those networks and the motifs that whose instances appear relatively more often and the number of motif instances in the network. Meanwhile, the frequencies of all motif instances in the networks are shown in Figure 7.

To test the clustering effect of WMCA, the MCA method is selected for comparison. The motifs with more instances that appear in the network are selected as the motifs of interest to cluster the network. When using the MCA method, the networks are regarded as directed unweighted networks. In the non-weighted versions of these networks, the motifs M3, M6 and M11 are selected as the motifs of interest for seventh graders network, the selections of motifs for the other networks are shown in Table 4. For the WMCA method, the selections of motifs for the networks are shown in Table 4.

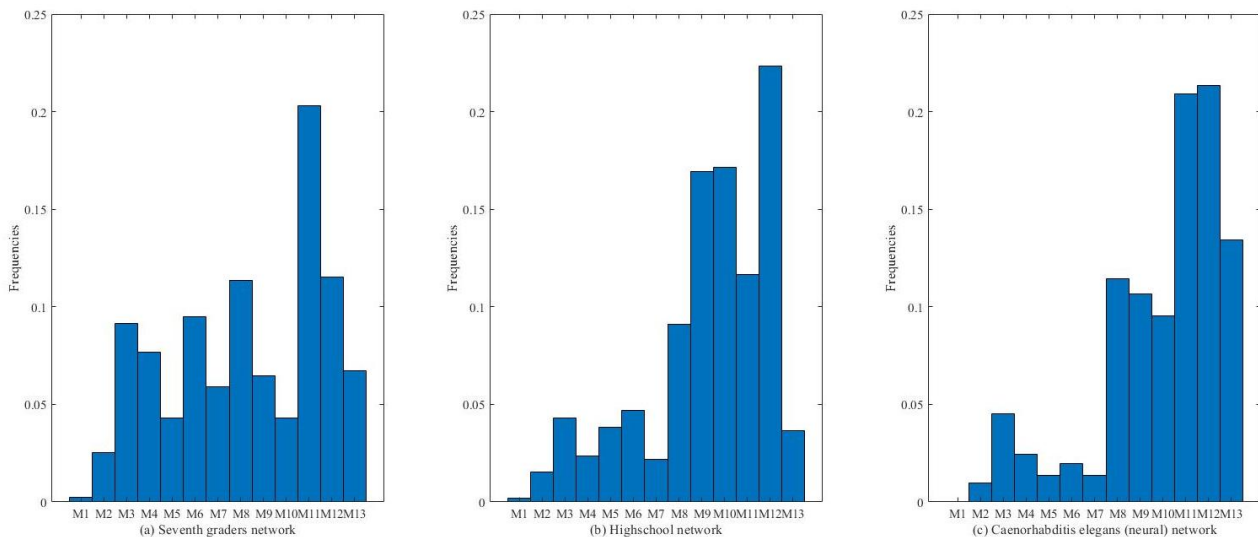


Figure 7. The frequencies of all motif instances in the networks.

Table 4. Basic information about the networks.

Network	Size	Number of edges	Maximum edge weight	Motifs
Seventh graders	29	376	3	M8(464), M11(829), M12(471)
Highschool	70	366	2	M9(324), M10(328), M12(427)
Caenorhabditis elegans (neural)	297	4296	72	M9(9111), M10(8138), M11(17884), M12(18224), M13(11496)

WMCA and MCA method are applied to these networks respectively, and the common internal clustering evaluation indexes: Silhouette Coefficient (SC) [21] and Davies-Bouldin index (DBI) [22] are selected to evaluate the effect. SC ranges between -1 and 1 , where a higher SC refers to a model with more coherent clusters. And the smaller the value of DBI, the better the clustering effect. Their formats are as follows:

$$SC = \frac{1}{n} \sum_{i=1}^n \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (3)$$

$$DBI = \frac{1}{K} \sum_{u,v=1}^K \max_{u \neq v} \left(\frac{s_u + s_v}{d_{uv}} \right), \quad (4)$$

where n is the number of nodes, $a(i)$ is the cluster cohesion which refers to the average distance between the node i and all other nodes within the same cluster, and $b(i)$ is the cluster separation which refers to the average distance between the node i and all other nodes in the nearest cluster. And K is the number of clusters, s_u and s_v respectively represent intra-cluster dispersions for cluster u and v , the average distance between each node within the cluster and its centroid. d_{uv} is the distance between is the distance between the centroid of cluster u and v .

Select the number of clusters within a certain range according to the number of network nodes. Under these number of clusters, the effect of two clusters is assessed by two evaluation metrics, respectively. The number of classifications K and the corresponding value of two clustering evaluation

indexes are shown in Figure 8. Figure 8(a),(b),(c) correspond to the cluster evaluation values of the above three networks, respectively. On the whole, the SC value of WMCA is larger than that of MCA method, especially when the number of clusters gets bigger and bigger, and the DBI values of WMCA are generally smaller than those of the MCA method. Therefore, the clustering effect of the WMCA method is better than that of the MCA method.

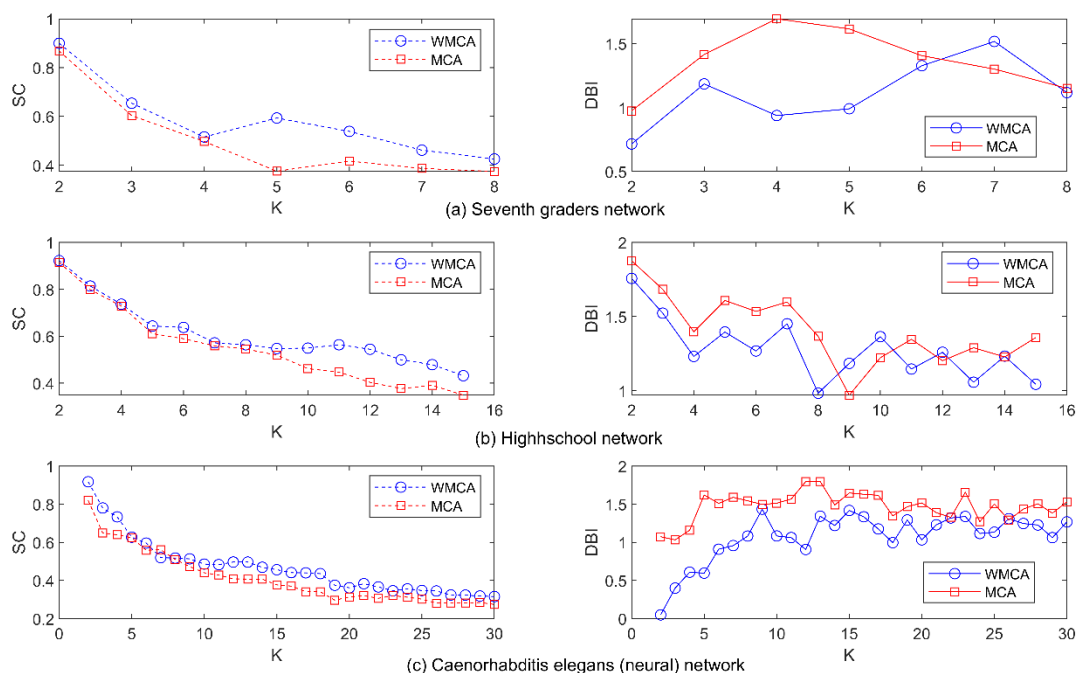


Figure 8. The number of different classifications and their corresponding SC and DBI value.

5. Conclusions

Based on matrix multiplication and Hadamard product of matrices, this paper supplements the matrix expressions of motif adjacency matrix in directed unweighted networks. The central computational kernel in these computations is $XY \circ Z$. Then, through the relationship between the corresponding adjacency matrix of the networks, the directed weighted networks can be transformed into directed unweighted multilayer networks. The transformation of weighted networks in this paper is suitable for networks with integer weights. If the weights are not integers, the weights may need to be processed. Thus, the motif adjacency matrix of the network can be obtained, and the method of spectral clustering can be applied.

The advantages of the methods proposed in this paper are as follows: firstly, it provides a convenient method to deal with the weight of the network, which will be helpful to the research of other weighted networks; secondly, the clustering method is applicable to both weighted and unweighted networks, and one can select one or more motifs of interest to divide the network; finally, it is very convenient to implement the method of obtaining the expression of motif adjacency matrix and other algorithms based on motif adjacency matrix for medium scale networks. When the networks are large and sparse, these computations become slower than standard fast triangle enumeration algorithms, the parallel algorithm of sparse matrix product can improve the efficiency of

calculation [25]. According to the structure of the motifs, the expression of the motif adjacency matrix based on matrix representation can also be extended to the cases of four-node motifs and higher order.

Overall, the complexity of the WMCA algorithm is governed by the mapping of weighted networks, the computations of the motif adjacency matrix W_M , an eigenvector, and the sweep cut procedure. When mapping a weighted network, it is necessary to traverse $n \times n$ elements of the adjacency matrix, whose time complexity is $O(n^2)$ in theory. For forming a motif adjacency matrix, suppose $X_{n \times n}, Y_{n \times n}, Z_{n \times n}$, the time complexity of matrix multiplication is $O(n^3)$. In theory, the time complexity of the method shown in Table 1 is $O(n^3)$. And the time complexity can be optimized to $O(n^{2.376})$ [26] through the optimization of matrix operations. The complexity of the remaining steps is consistent with the method proposed by Benson et al, which is $O\left(n^{\frac{3}{2}}\right)$.

The method (WMCA) in this paper is to cluster nodes based on the participation of nodes in forming the motif, rather than the direct links between nodes. In this paper, the directed weighted network is decomposed into a multilayer network composed of directed unweighted networks. And it has been proved in the supporting materials of Benson et al. [12] that the motif-based spectral clustering of directed unweighted networks can be converted into the spectral clustering problem of undirected weighted networks corresponding to the motif-based adjacency matrix, and the motif adjacency matrix of multilayer networks is the sum of the motif adjacency matrix of each layer network. Therefore, the motif-based clustering of directed weighted network can also be solved by the spectral clustering method of undirected weighted network corresponding to the motif-based adjacency matrix. In essence, the methods that take the conductivity of the matrix (graph) as the optimization objective can adopt the hierarchical processing method proposed in this paper for the directed weighted network to carry out motif-based clustering. Sieranoja et al. [27] introduce two new algorithms K-algorithm and M-algorithm, they are adaptations of the k-means algorithm for graph clustering. The method is applicable to undirected networks. When the direction of the network needs to be considered, the undirected weighted network based on the motif adjacency matrix of network motifs may be helpful. While a certain kind of higher-order structure is being considered in networks, some lower-order links are being ignored. In the future research, the two may be combined to consider. And when there is more than one important motif of the network, it may be worth studying which one or several motifs have the best clustering effect.

Acknowledgments

This work was supported by National Natural Science Foundation of China (grant numbers 11601267).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. S. E. Schaeffe, Graph clustering, *Comput. Sci. Rev.*, **1** (2007), 27–64. <https://doi.org/10.1016/j.cosrev.2007.05.001>

2. W. E. Donath, A. J. Hoffman, Algorithms for partitioning of graphs and computer logic based on eigenvectors of connection matrices, *IBM Techn. Discl. Bull.*, **15** (1972), 938–944.
3. Z. Wu, R. Leahy, An optimal graph theoretic approach to data clustering: theory and its application to image segmentation, *IEEE T. Pattern Anal.*, **15** (1993), 1101–1113. <https://doi.org/10.1109/34.244673>
4. L. Hagen, A. B. Kahng, New spectral methods for ratio cut partitioning and clustering, *IEEE T. Comput. Aid. D.*, **11** (1992), 1074–1085. <https://doi.org/10.1109/43.159993>
5. Estrada, Ernesto, D. J. Higham, Network properties revealed through matrix functions, *SIAM Rev.*, **4** (2010), 696–714. <https://doi.org/10.1137/090761070>
6. A. Y. Ng, M. I. Jordan, Y. Weiss, On spectral clustering: Analysis and an algorithm, *Adv. Neural Inf. Process. Syst.*, 2001, 849–856.
7. M. Boedihardjo, S. Deng, T. Strohmer, A performance guarantee for spectral clustering, *SIAM J. Math. Data Sci.*, **3** (2021), 369–387. <https://doi.org/10.1137/20M1352193>
8. V. Satuluri, S. Parthasarathy, Symmetrizations for clustering directed graphs, *Proceedings of the 14th International Conference on Extending Database Technology*, 2011, 343–354. <https://doi.org/10.1145/1951365.1951407>
9. R. Milo, Network motifs: Simple building blocks of complex networks, *Science*, **298** (2002), 824–827. <https://doi.org/10.1126/science.298.5594.824>
10. C. Li, Y. Tang, Z. Tang, J. Cao, Y. Zhang, Motif-based embedding label propagation algorithm for community detection, *Int. J. Intell. Syst.*, **37** (2021), 1880–1902. <https://doi.org/10.1002/int.22759>
11. J. Shi, J. Malik, Normalized cuts and image segmentation, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, 731–737. <https://doi.org/10.1109/CVPR.1997.609407>
12. A. R. Benson, D. F. Gleich, J. Leskovec, Higher-order organization of complex networks, *Science*, **353** (2016), 163–166. <https://doi.org/10.1126/science.aad9029>
13. Y. L. Zhang, B. Wu, Y. Liu, J. N. Lv, Local community detection based on network motifs, *Tsinghua Sci. Technol.*, **24** (2019), 716–727. <https://doi.org/10.26599/TST.2018.9010106>
14. Y. Ge, H. Lu, P. Peng, Mixed-order spectral clustering for complex networks, *Pattern Recogn.*, **117** (2021), 107964. <https://doi.org/10.1016/j.patcog.2021.107964>
15. F. Xia, S. Yu, C. F. Liu, J. X. Li, I. Lee, CHIEF: Clustering with higher-order motifs in big networks, *IEEE T. Netw. Sci. Eng.*, **9** (2022), 990–1005. <https://doi.org/10.1109/TNSE.2021.3108974>
16. Y. F. Wang, H. Y. Wang, S. H. Zhang, A weighted higher-order network analysis of fine particulate matter ($PM_{2.5}$) transport in Yangtze River Delta, *Phys. A (Amsterdam, Neth.)*, **496** (2018), 654–662. <https://doi.org/10.1016/j.physa.2017.12.096>
17. X. Shen, X. Gong, X. Jiang, J. Yang, X. Hu, High-order organization of weighted microbial interaction network, *2018 IEEE International Conference on Bioinformatics and Biomedicine*, 2018, 206–209. <https://doi.org/10.1109/BIBM.2018.8621218>
18. W. G. Underwood, A. Elliott, M. Cucuringu, Motif-based spectral clustering of weighted directed networks, *Appl. Netw. Sci.*, **5** (2020), 62. <https://doi.org/10.1007/s41109-020-00293-z>
19. M. E. J. Newman, Analysis of weighted networks, *Phys. Rev. E*, **70** (2004), 05613. <https://doi.org/10.1103/PhysRevE.70.056131>
20. A. H. Roger, R. J. Charles, *Matrix analysis*, 2 Eds., Cambridge: Cambridge University Press, 2012.

21. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, **20** (1987), 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
22. D. L. Davies, D. W. Bouldin, A cluster separation measure, *IEEE T. Pattern Anal.*, **PAMI-1** (1979), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
23. W. Zachary, An information flow model for conflict and fission in small groups, *J. Anthropol. Res.*, **33** (1977), 452–473.
24. J. Kunegis, KONECT: The Koblenz network collection, *Proceedings of the 22nd International Conference on World Wide Web*, 2013, 1343–1350. <https://doi.org/10.1145/2487788.2488173>
25. A. Azad, A. Buluç, J. R. Gilbert, Parallel triangle counting and enumeration using matrix algebra, *2015 IEEE International Parallel and Distributed Processing Symposium Workshop*, 2015, 804–811. <https://doi.org/10.1109/IPDPSW.2015.75>
26. D. Coppersmith, S. Winograd, Matrix multiplication via arithmetic progressions, *J. Symb. Comput.*, **9** (1990), 251–280.
27. S. Sieranoja, P. Frnti, Adapting k-means for graph clustering, *Knowl. Inf. Syst.*, **64** (2022), 115–142. <https://doi.org/10.1007/s10115-021-01623-y>

Supplementary

(1) The Matlab algorithm “ $W = UWMA(A, \text{motif})$ ” to form the expression of the motif adjacency matrix corresponding to the motif M8-M13 is shown in Figure 1. Referring to the method “ $W = \text{MotifAdjacency}(A, \text{motif})$ ” of Benson et al. [12] (<http://snap.stanford.edu/higher-order/>), Matlab algorithm, “ $W = UWMA(A, \text{motif})$ ”, that can form the expressions of motif adjacency matrix for all motifs can be given. Figure 1 also shows the algorithm “ $W = WMA(A, \text{motif})$ ” that forming the motif adjacency matrix for the directed weighted network.

(2) In a directed weighted network, mapping $\phi(\cdot)$ maps the network to a multilayer network. Let G_M be the graph corresponding to the motif adjacency matrix, and $G^l, W_M^l = (w_{ij}^l)_{n \times n}, D^l = (d_{ii}^l)_{n \times n}$ and G_M^l respectively represent the network graph, the motif adjacency matrix, the degree matrix of the motif adjacency matrix and the graph corresponding to the motif adjacency matrix of the l -layer network. Let $|\beta|$ be the number nodes that make up the motif, and when the number of nodes that make up the motif is 3, $|\beta|=3$.

Proposition 1: Consider the motif M , after the weighted network is mapped to the multi-layer network, the motif-based cuts and volumes of the multi-layer network are respectively equal to the sum of the motif-based cuts and volumes of each layer network, i.e (1) $cut(G_M)S, \bar{S} = \sum_l cut(G_M^l)S, \bar{S}$, (2) $vol(G_M)S = \sum_l vol(G_M^l)S$.

Proof. (1) Known from the paper [12] that: $cut_M(G_M^l)S, \bar{S} = \sum_{i \in S, j \in \bar{S}} w_{ij}^l$, and $cut(G_M)S, \bar{S} = \sum_{i \in S, j \in \bar{S}} w_{ij}^M$.

Then $\sum_{l=1} cut_M(G_M^l)S, \bar{S} = \sum_{l=1} \sum_{i \in S, j \in \bar{S}} w_{ij}^l$, and $cut(\sum G_M^l)S, \bar{S} = \sum_{i \in S, j \in \bar{S}} w_{ij}^M$.

And $\sum_l W_M^l = W_M$, thus $cut(G_M)S, \bar{S} = \sum_l cut(G_M^l)S, \bar{S}$.

(2) It is known that [12]: $vol(G_M^l)S = \sum_{i \in S} d_{ii}^l$.

Then $\sum_l \text{vol}_M(G_M^l)S = \sum_l \sum_{i \in S} d_{ii}^l = \sum_l \sum_{i \in S} \sum_{j=1}^n w_{ij}^l$, and $\text{vol}(G_M)S = \sum_{i \in S} \sum_{j=1}^n w_{ij}^M$.

And $\sum_l W_M^l = W_M$, thus $\sum_l \text{vol}(G_M^l)S = \text{vol}(G_M)S$.

```

1.
function W=UWMA(A, motif)
% A: Adjacency matrix of the directed unweighted network;
% motif: Two-hop paths motifs. Use a string, such as 'Mn';
B = A&A';      % Bidirectional edges matrix;
U = A - B;     % Unidirectional edges matrix;
S1 = A | A';   % The adjacency matrix of the corresponding undirected graph;
F = ~S1-eye(size(A)); % edge-missing matrix
BL=tril(B);    %Take the trigonometric matrix;
UL=tril(U);
Fl=tril(F);
if strcmp(motif,'M8')==1
    W=Fl*U'.*U'+Fl'*U'.*U'+U*Fl.*U +U*Fl'.*U +U'*U.*F;
elseif strcmp(motif,'M9')==1
    C=U*U.*F + U*Fl.*U' + U*Fl'.*U'+ Fl*U.*U'+ Fl'*U.*U';
    W=C+C';
elseif strcmp(motif,'M10')==1
    W=U*U'.*F + U*Fl.*U' + U*Fl'.*U' + Fl*U.*U + Fl'*U.*U;
elseif strcmp(motif,'M11')==1
    C=BL*U.*F +BL'*U'.*F + U*Fl.*B+ U*Fl'.*B +Fl*B.*U'+Fl'*B.*U';
    W=C+C';
elseif strcmp(motif,'M12')==1
    C=BL*U'.*F +BL'*U'.*F + U*Fl.*B+ U*Fl'.*B +Fl*B.*U'+Fl'*B.*U;
    W=C+C';
elseif strcmp(motif,'M13')==1
    W=BL*B.*F +BL'*B.*F+ B*Fl.*B + B*Fl'.*B +Fl*B.*B+Fl'*B.*B;
else
    error("Motif must be one of M8-M13.")
End

2.
function w = WMA(A, motif)
% A: Adjacency matrix of the directed weighted network;
% motif: Three-node motifs. Use a string, such as 'Mn';

B=min(A,1);
w= UWMA(B, motif); %The first layer network.

a=max(max(A));
n=1;
while n<=a
    A=A-min(A,1);
    W=UWMA(A, motif); %The each layer (except Layer 1).
    w=w+W;
    n=n+1;
end

```

Figure 1. The method of obtaining motif adjacency matrix by Matlab.

Proposition 2: In the multi-layer network, consider the three-node motifs, and the conductance of the directed weighted network is equivalent to the motif conductance of the network corresponding to the motif adjacency matrix, i.e $\varphi(G)S = \varphi_M(G)S$.

Proof. Known from the paper [12] that $vol_M(G_l)S = \frac{1}{|\beta|-1} vol(G_M^l)S$, and $cut_M(G)S, \bar{S} = \frac{1}{2} cut(G_M)S, \bar{S}$.

$$\text{Then } vol_M(G^l)S = \frac{1}{|\beta|-1} vol(G_M^l)S, cut_M(G^l)S, \bar{S} = \frac{1}{2} cut(G_M^l)S, \bar{S}.$$

From Proposition 1 know: $cut_M(G)S, \bar{S} = \sum_l cut_M(G^l)S, \bar{S}$, and $vol_M(G)S = \sum_l vol_M(G^l)S$.

$$\text{Then } cut_M(G)S, \bar{S} = \sum_l cut_M(G^l)S, \bar{S} = \frac{1}{2} \sum_l cut(G_M^l)S, \bar{S} = \frac{1}{2} cut(G_M)S, \bar{S},$$

$$\text{and } vol_M(G)S = \sum_{l=1} vol_M(G^l)S = \frac{1}{|\beta|-1} \sum_{l=1} vol(G_M^l)S = \frac{1}{|\beta|-1} vol(G_M)S.$$

When $|\beta| = 3$,

$$\begin{aligned} \varphi_M(G)S &= \frac{\frac{1}{2} cut(G_M)S, \bar{S}}{\min\left(\frac{1}{2} vol(G_M)S, \frac{1}{2} vol_M(G_M)S\right)} \\ &= \frac{cut_M(G)S, \bar{S}}{\min(vol_M(G)S, vol_M(G)\bar{S})} \\ &= \varphi(G)S \end{aligned}$$

The proof is completed.



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)