Abstract—Recently, the core-periphery (CP) structure as a type of meso-scale structure in complex networks has attracted some attention, which is composed by a dense core and a sparse connected periphery. In this paper, we propose an algorithm to detect the CP structure based on the 3-tuple motif, which is inspired by the conception of motif defined in a recent paper [Science, 353, 163-166, (2016)]. In this algorithm, we first define a 3-tuple motif by considering the property of nodes, and then a motif adjacency matrix is formed based on the defined motif, finally, the detection of the CP structure is converted to find a cluster that minimizes the smallest motif conductance. Our algorithm can detect different CP structures, including single or multiple CP structure, local or global CP structures. Results on the synthetic and the empirical networks indicate that the method is efficient and can apply to large-scale networks. More importantly, our algorithm is free parameter, where the core and periphery are detected without any predefined parameters.

Index Terms—Core-periphery structure, motif, the optimal conductance, complex networks.

I. INTRODUCTION

Network science has become a very topic field during the past two decades, since many complex systems can be expressed as networks in which entities and interactions can be represented by nodes and edges. The structures of networks can be described from micro-scale, meso-scale and macro-scale perspectives. Recently, myriad methods have been developed to detect one typical meso-scale structure: community structure [1]– [3]. However, as for another type of meso-scale structure: core-periphery (CP) structure, has received much less attention than they deserve. The CP structure has been examined in the study of society [4], transportation [5], scientific citation [6], international trade [7], [8] and other fields [9]–[11]. Although a core and a community are both a group of densely interconnected nodes, there is a remarkable difference between them, since a core

This work is supported by National Natural Science Foundation of China (61473001), and partially supported by the Young Talent Fundings of Anhui Provincial Universities (gxyqZD2017003).

C. Ma, B.-B. Xiang, and H.-F. Zhang are with the School of Mathematical Science, Anhui University, Hefei 230601, China (chuang@126.com; xiangbb2311@163.com; haifengzhang1978@gmail.com).

H. Chen is with the School of Physics and Material Science, Anhui University, Hefei 230601, China (chenhshf@mail.ustc.edu.cn).

M. Small is with the Department of Mathematics and Statistics, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009 (michael.small@uwa.edu.au.).

also addresses that it should densely connect with its periphery [12]. Though the intuitive notation of CP has been mentioned for a long history [13], the first quantitative formulation of CP structure was proposed by Borgatti and Everett in the late 1990s [14], in which a node belongs to a core if and only if it is well connected both with other core nodes and with peripheral nodes, and peripheral nodes have no connections with other peripheral nodes. Hereafter, some detection algorithms of CP structure have been proposed. There are several ways of design CP detection algorithms: one is to maximize the objective functions characterizing how well a partition match with the idealised CP structure [12], [14], [15]. Another is that the CP structure is detected according to a certain ordering sequence encoding the core value of each node [16]–[20]. In addition, Zhang et al. have proposed that the CP structure can be detected by fitting a stochastic block model (SBM) to empirical network data using a maximum likelihood method [21]. Many proposed algorithms can only deal with single CP structure, or the size of core nodes should be given in advance, or high computational complexity. Therefore, the investigation of the CP structure is still in an initial stage.

Recently, Benson et al. developed a generalized framework for clustering networks on the basis of structural motifs instead of edges [22]. The motif is described by a 2-tuple \((B,A)\), where \(B\) is a \(k \times k\) binary matrix and \(A \subset \{1,2,\ldots,k\}\) is a set of anchor nodes. Then they proved that the near optimal clustering of the network can be realized by finding the set with the smallest motif conductance. Such a method can be applied to directed, undirected, weighted networks, sign networks, and so forth. However, the motifs defined in the work mainly focused on the patterns of edges in the subgraph, which has no any restriction about the nodes of the subgraph. So we want to know: is it possible to uncover higher-organization structures of networks by define some motifs which incorporate the information of nodes.

From the definition of CP structure, it is known that core nodes are the high-degree nodes, even though the high-degree nodes are not necessarily the core nodes [14], [21]. Inspired by this fact, we first defined motifs by a 3-tuple \((B,A,\Phi)\) in this paper. The property \(\Phi\) is to highlight the degrees of core nodes are larger than the peripheral nodes, and also larger than the average degree of the network. As in Ref. [22], the motif
adjacency matrix is first constructed on the basis of the defined motif, then the dichotomization of the network is implemented by finding a cluster that minimizes the smallest motif conductance. Finally, the subgraph with larger average degree is the core, and the other is the periphery. We further generalize our algorithm to detect multiple cores and the global core, respectively. The performance of the algorithm is validated in different networks, more importantly, the algorithm is fast and free parameter.

II. RELATED WORK

The detection of CP structure was formally considered by Borgatti and Everett [14]. In the work, a discrete-version algorithm aims to find a vector C of length N whose entries can be either 1 or 0. The ith entry $C_i$ equals 1 if the corresponding node is assigned to the core, otherwise, it equals 0 if the node is assigned to the periphery. And define a pattern matrix with the element $C_{ij} = 1$ if $C_i = 1$ or $C_j = 1$, and let $C_{ij} = 0$ otherwise. A core quality $\rho = \sum_{i,j} A_{ij} C_{ij}$ is defined to measure how well a network approximates the ideal case, where $A_{ij}$ is the element of adjacency matrix $A$. The goal of the algorithm is to seek a value of $\rho$ that is high compared to the expected value of $\rho$ if $C$ is shuffled such that the number of 1 and 0 entries is preserved but their order is randomized [14], [15], [19]. They also presented a continuous version in which each node is assigned a continue “coreness” value between 0 and 1, and the element of the pattern matrix $C_{ij} = C_i C_j$. The continuous version was further developed in Ref. [15], where the aggregate core score of each node is defined by the core quality as well as the transition function. Very recently, the problem of CP detection is further developed by Kajoku et al., they define a more general objective function to find multiple CP pairs in networks, the multiple CP pairs can be found by maximizing the objective function [12].

Some methods detect core nodes by ranking nodes according to a defined core centrality. For instance, based on the fact that core nodes should have a high closeness centrality, Holme have proposed a CP coefficient using the closeness centrality and k-cores deposition technique to determine core nodes [11]. Silva et al. proposed a parameter called core coefficient to quantitatively evaluate the core-periphery structure of a network, which was defined based on the concept of closeness centrality and a newly defined parameter: network capacity [16]. A knotty centralization was proposed by Shanahan et al. to detect the core in networks, which attempts to find nodes that have high betweenness centrality but without high degree centrality [17]. In addition, other methods to measure the core values of nodes based on the path-core or random walk were studied in Ref. [19] and Ref. [18], respectively.

Methods to maximize the objective function characterizing how well a given network match an idealized CP structure may yield inefficient results, since empirical networks often significantly deviate from the idealised CP structure. Given that, Zhang et al proposed an algorithm to identify the CP structure by fitting a stochastic block model to empirical network data using a maximum likelihood method, rather than the idealised CP structure.

In this work, our algorithm is totally different from the proposed algorithms. In this algorithm, we first generalize the definition of 2-tuple motif to 3-tuple motif by considering the property of nodes, and then a motif adjacency matrix is constructed based on the 3-tuple motif, finally, the detection of the CP structure is converted to find a cluster that minimizes the smallest motif conductance. What’s more, our algorithm can detect single or multiple CP structure, as well as local or global CP structures.

It should be noted that the problem of the CP detection is still in an initial stage. For example, on the one hand, the standard index, like modularity index or normalized mutual information (NMI) index [23], to measure the performance of different algorithm has not been well investigated (which is also our future research direction, and leave the issue for the future consideration). On the other hand, some algorithms can only deal with single CP structure, or they should determine the core size in advance. Therefore, unlike the community detection algorithms, the performance of algorithm to detect CP is hard to be compared. To the best of our knowledge, systematical comparison between different algorithms in this field is lack. In this work, we generate a synthetic network and some empirical networks to highlight the performance of our algorithm.

III. MOTIF SPECTRAL CLUSTERING

We first review the method developed in Ref. [22], which is to detect the higher-order organizations by finding a cluster with the smallest motif conductance. There are three main steps to detect higher-order organizations:

Define a motif $M$ and form motif adjacency matrix $W_M$: Consider an undirected graph $G = (V,E)$ with $|V| = n$, the definition of motif is described by a 2-tuple $(B,A)$ on $k$ nodes. The binary matrix $B$ includes the patterns of edges in the subgraph. $A \subset \{1,2,\ldots,k\}$ is a set of anchor nodes, which labels a relevant subset of nodes for defining motif conductance (see Fig. 1(a)). By letting $\chi_A$ be a selection function that takes the subset of a $k$-tuple indexed by $A$, and $\set$ be the operator taking a tuple to a set. Namely,

$$\set((v_1,v_2,\ldots,v_k)) = \{v_1,v_2,\ldots,v_k\}. \tag{1}$$

Assume a network $G$ is denoted by an adjacency matrix $A$, then the set of motifs is defined as:

$$M(B,A) = \{(\set(v),\set(\chi_A(v))) \mid v \in V^k, v_1,\ldots,v_k \text{ distinct}, A_v = B\}. \tag{2}$$
where $A_v$ is the $k \times k$ adjacency matrix on the subgraph induced by the $k$ nodes of the ordered vector $v$. For convenience, $(\text{set}(v), \text{set}(\chi_A(v)))$ is simply denoted as $(v, \chi_A(v))$ when considering the elements of $M(B,A)$. Moreover, any $(v, \chi_A(v)) \in M(B,A)$ is called a motif instance (see Fig. 1(b)).

Once the adjacency matrix $A$ and a motif set $M$ are determined, the motif adjacency matrix (Fig. 1(c)) can be formally defined as:

$$ (M_{W})_{ij} = \sum_{(v, \chi_A(v)) \in M} 1\{i,j\} \in \chi_A(v), \quad (3) $$

where $1(s)$ is the truth-value indicator function on $s$, i.e., $1(s)$ takes the value 1 if the statement $s$ is true and 0 otherwise.

**Apply spectral clustering on $W_M$:** Given a motif $M$, an optimal cluster of nodes $S$ have two goals: on the one hand, the nodes in $S$ should contain many instances of $M$; on the other hand, the set $S$ should avoid cutting instances of $M$. Thus, the optimal cluster is realized by minimizing the following motif conductance:

$$ \phi_M(S) = \frac{\text{cut}_M(M) \cdot S \cdot \bar{S}}{\min(\text{vol}_M(G) \cdot S, \text{vol}_M(G) \cdot \bar{S})}. \quad (4) $$

where $\bar{S}$ is the complement of $S$, namely, $\bar{S} = V \setminus S$. $\text{cut}_M(M) \cdot S \cdot \bar{S}$ is the number of instances of motif $M$ with at least one node in $S$ and one in $\bar{S}$, and $\text{vol}_M(G) \cdot S$ is the number of nodes in instances of $M$ that reside in $S$. The supplementary materials of Ref. [22] have proved that the near optimal clusters can be obtained by using motif spectral clustering on the motif adjacency matrix $W_M$.

First, define the normalized Laplacian matrix as:

$$ L_M = I - D_M^{-1/2}W_MD_M^{-1/2}, \quad (5) $$

where $(D_M)_{ii} = \sum_{j=1}^{n} (W_M)_{ij}$ is the diagonal degree matrix and $I$ is the identity matrix. Then compute the eigenvector of the second smallest eigenvalue of $L_M$, denoted by $z$. And let $\sigma_i$ be the index of $D_M^{-1/2}z$ with the $i$th smallest value.

**Output the clusters:** After the vector $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ is given, the last job is to find the prefix set of $\sigma$ with the smallest motif conductance, namely, $S = \arg \min \phi_M(S_i)$ with $S_i = \{\sigma_1, \sigma_2, \ldots, \sigma_i\}$. The detail steps are presented in Algorithm 1.

**Algorithm 1:** (Input: graph $G$ and motif $M$. Output: optimal cluster $S$)

**Step 1:** Compute $W_M$ according to Eq. (3);

**Step 2:** Compute the normalized motif Laplacian matrix according to Eq. (5);

**Step 3:** Compute the eigenvector $z$, which is the second smallest eigenvalue of $L_M$;

**Step 4:** Let $\sigma_i$ be the index of $D_M^{-1/2}z$ with the $i$th smallest value;

**Step 5:** Find $S = \arg \min \phi_M(S_i)$ by increasing size $r$, where $S_r = \{\sigma_1, \sigma_2, \ldots, \sigma_r\}$;

**Step 6:** If $|S| < |S|$, then return $S$, else return $S$.

From the description of the Algorithm 1, one can find that there is an assumption in the Step 6: the optimal clustering $S$ is the smallest set of $S$ and $S$. However, there is no evidence to support this assumption. Moreover, according to this algorithm, one optimal cluster $S$ can be detected, however, the cluster is wrong if there is no any optimal cluster at all. This problem is very similar to the detection of community networks. The modularity optimization algorithm can produce a wrong partition even though one network has no obvious community structure. As a result, some improvements can be made in this method. In this work, we generalize the motif spectral clustering method to detect the CP structure in networks, which also overcomes the mentioned shortcomings.

**IV. Methods**

The definition of motif defined in the above section just considers the patterns of edges, but ignores the property of the nodes. In many cases, the property of the nodes is also important, and considering the property of nodes may help us find more important organizations. For example, given that the core nodes in the networks must be the high-degree nodes, therefore, we define a new type of motif by considering the degree property of nodes. Namely, motif is extended into a 3-tuple: $(B,A,\Phi)$, where $\Phi$ is the property of nodes.

By extending a 2-tuple motif definition to a 3-tuple motif definition, which is defined as:

$$ M(B,A,\Phi) = \{(\text{set}(v), \text{set}(\chi_A(v))) \mid v \in V^k, v_1, \ldots, v_k \text{ distinct}, A_v = B, f(\Phi, (\text{set}(v), \text{set}(\chi_A(v)))) = 1\}. \quad (6) $$

where $f(\Phi, (\text{set}(v), \text{set}(\chi_A(v)))) = 1$ indicates that the node of $\text{set}(v)$ and $\text{set}(\chi_A(v))$ should satisfy the property of $\Phi$.

According to the definition of CP in Ref. [14], one core node should connect with all other nodes, and a peripheral node should connect with all core nodes but
the degrees of all nodes in $S$ will be obtained by Step 1 - Step 5 in Algorithm 1 (the optimal cluster is the smallest set of $S$ and $\bar{S}$). We only need to judge whether the degrees of all nodes in $S$ or $\bar{S}$ are larger than the average degree $\bar{d}$. Obviously, the one who satisfies the condition is the core, and the other is the periphery.

It is noteworthy that the definition of $W_{M_1}$ in Eq. (7) is different from Eq. (3), which is defined to ensure the weight among core nodes and the weight among peripheral nodes are not changed. Therefore the difference between core and periphery is more obvious (see the diagrams in Fig. 3). Namely, such a definition is in favor of the detection of CP structure. For example, take the classical CP structure in Fig. 3(a) as an example, the network is divided into two well-connected subgraphs (see Fig. 3(b)) after the function mapping of $W_{M_1}$: core and periphery. From this example, one can also find that our method need not use the assumption in step 6 of Algorithm 1 (the optimal cluster is the smallest set of $S$ and $\bar{S}$). We only need to judge whether the degrees of all nodes in $S$ or $\bar{S}$ are larger than the average degree $\bar{d}$. Obviously, the one who satisfies the condition is the core, and the other is the periphery.

The CP structure shown in Fig. 3(a) is too strict to meet, in many cases, the diagram in Fig. 3(c) cannot be detected since $W_{M_1}$ is a zero-matrix. To do this, we define a looser motif $M_2$ as in Fig. 2(b), then the motif adjacency matrix $W_{M_2}$ is formed according to Eq. (7). As shown in Fig. 3(d), the illustration in Fig. 3(c) can be divided two separated sets after the function mapping of $W_{M_2}$: core and periphery.

Since the definition of motif $M_1$ is too strict and the definition of motif $M_2$ is very loose, a modest motif $M_3$ intermediated motif $M_1$ and $M_2$ is also defined (see Fig. 2(c)). It is hard to know whether the CP structure is obvious in a given network owing to the richness and diversity of real networks. In view of this, we define motif adjacency matrix $W_M$ as the combination of $W_{M_1}$, $W_{M_2}$ and $W_{M_3}$:

$$W_M = \alpha W_{M_1} + \gamma W_{M_2} + \beta W_{M_3}, \quad (8)$$

where $\alpha \geq \beta \geq \gamma \geq 0$. Since motif $M_2$ is encoded in motif $M_3$, and further encoded in motif $M_1$. In this paper, without loss of generality, we set $\alpha = 2\beta$, $\beta = 2\gamma$ and $\gamma = 1.0$.

In summary, the motif adjacency matrix is first obtained according to Eq. (8) for a given network, then $S$ and $\bar{S}$ will be obtained by Step 1 - Step 5 in Algorithm 1. Finally, the core is the set $S$ or $\bar{S}$ in which degrees of all nodes are larger than the average degree. In the
next section, we will validate the good performance of our algorithm by implementing it on a synthetic network and real networks.

V. MAIN RESULTS

A. Detection of single CP structure

First, we generate a synthetic network model with single CP structure, which is determined by five parameters: \( C, P, P_{CC}, P_{CP} \) and \( P_{PP} \). \( C \) is a set of core nodes, \( P \) is a set of peripheral nodes, \( P_{CC} \) is the connection probability between core nodes, \( P_{CP} \) is the connection probability between core nodes and peripheral nodes, and \( P_{PP} \) is the connection probability between the peripheral nodes. \( P_{CC} \geq P_{CP} > P_{PP} \) is to make sure there is a CP structure in the network.

By generating a network with \( |C| = 30, |P| = 70, P_{CC} = 0.5, P_{CP} = 0.5 \) and \( P_{PP} = 0.25 \), the performance of our algorithm is checked in Fig. 4. The blue curve is the motif conductance \( \phi \) which is computed from the Step 1-Step 4 of Algorithm 1. The smallest motif conductance \( \phi \) divides the nodes into two sets, one set in the right (red nodes and the red nodes covered by green circles) is core since their degrees are larger than the average degree (red dash line), another set in the left (black nodes) is the peripheral nodes. One should note that our algorithm gives rise to 33 core nodes, that is to say, three additional core nodes (red nodes covered by green circles) are detected. However, such an inconsistency cannot deny the efficiency of our algorithm. As we know, owing to the randomness in the connection probabilities of \( P_{CC}, P_{CP} \) and \( P_{PP} \), the real number of core nodes may be larger or less than 30 even though we have set \( |C| = 30 \).

We further validate our algorithm on two real networks. The first network is the Karate club network, which consists of 34 nodes that represent club members and 78 links that represent friendships among members [24], whose motif conductance \( \phi_M(S_r) \) is shown as Fig. 5(a). The minimum value of \( \phi_M(S_r) \) dichotomizes the network: core and periphery. The eight red nodes in the right are the core nodes since their degrees are all larger than the average degree of the network (dash line), and the remaining black nodes in the left are the peripheral nodes. Previous studies have proven that the Karate club network is a community network including two communities. So one possible scenario is that the network has two separated cores, where each community includes one core. However, our algorithm detects one core since there are some links connecting the core nodes (red nodes and red edges, as shown in Fig. 5(b)), which can merge two small cores into a larger core, leading to a single CP structure in the network.

The second network is the USA airport network, which has 332 nodes representing airports and 2126 unweighted links describing the airlines between airports [25]. The blue curve in Fig. 5(c) plots the motif conductance \( \phi_M(S_r) \) as a function of \( S_r \). Also, the smallest motif conductance \( \phi_M(S_r) \) can automatically dichotomize the network as two sets: core and periphery. The core is composed of 27 nodes whose degrees are larger than the average degree (red nodes in Fig. 5(c)), another set is the periphery (black nodes in Fig. 5(c)). The illustration of the CP structure is presented in Fig. 5(d), where core nodes and peripheral nodes are labeled by red color and green color, respectively.
B. Detection of multiple cores

According to Algorithm 1, one possible scenario may emerge: even though the smallest motif conductance $\phi_M(S_r)$ can dichotomize the network into two sets, neither of them satisfies the condition that the degrees of all nodes are larger than the average degree (see Fig. 6(a) and (c)). This scenario may occur for two reasons: one is that the network itself does not exist CP structure at all, the other reason is that detection of the CP structure is influenced by the community structure. Because core nodes in different communities are sparsely connected [1], [24], the two sets obtained by our Algorithm 1 may be two communities rather than the core and periphery. In this case, the CP structure may be encoded in the communities. Therefore, our algorithm is further developed in the following.

If a network includes community structure, the smallest motif conductance $\phi_M(S_r)$ (the global minimum point) may give rise to two communities rather than the core and the periphery. But we cannot state that the network has no CP structure, it is possible that the CP structure encodes in communities. To this end, we can check whether there are some local minimum points in the curve of the motif conductance besides the global minimal point, and the local minimum points can be used to detect whether the CP structure exists in the community. Therefore, we need to define the local minimum point for the discrete sequence $x = \{x_1, x_2, \cdots, x_n\}$: we call a point $x_i (k < i \leq n - k)$ is a local minimum point of the function $h(x)$, if

$$h(x_i) \leq h(x_j) \quad (i - k \leq j \leq i + k),$$

(9)

here we choose $k = 3$.

For convenience, we define $S_G$ is a set whose elements are composed of the unhandled subgraph and $S_C$ is a set to save core nodes. Algorithm 2 is developed by extending Algorithm 1.

Algorithm 2: (Input: graph $G$ and motif $M$. Output: core set $S_C$)

Step 1: Initialize $S_G = \{G\}$, $S_C = \emptyset$;

Step 2: If $S_G$ is an empty set, then: the algorithm ends and returns $S_C$, else removes an element (subgraph) $\tilde{G}$ from the set $S_G$, i.e., $S_G = S_G/\{\tilde{G}\}$, $\tilde{G}$ is the first element of the set $S_G$;

Step 3: Two sets are obtained by Step 1 - Step 5 of Algorithm 1 for subgraph $\tilde{G}$, that is $S$ and $\tilde{S}$;

Step 4: If all nodes’ degrees of set $S$ are greater than the average degree, then $S_C = S_C \cup \{S\}$, and go to Step 2;

Step 5: If all nodes’ degrees of set $\tilde{S}$ are greater than the average degree, then $S_C = S_C \cup \{\tilde{S}\}$, and go to Step 2;

Step 6: If there exists a local minimum point in motif conductance function of sequence $S$, then $S_G = S_G \cup \{G_S\}$; If there exists a local minimum point in motif conductance function of sequence $\tilde{S}$, then $S_G = S_G \cup \{G_{\tilde{S}}\}$. Go to Step 2.

One point should be addressed: a network has no CP structure if all local minimum points in the curve have been checked, namely, each dichotomization based on the local minimum point cannot find CP structure.

We implement Algorithm 2 on two real networks. The first network is the Dolphin social network, which consists of 62 nodes representing the dolphins and 159 links denoting the frequent associations between dolphins [1]. The network is composed of two communities. Fig. 6(a) indicate that the two sets obtained from the smallest motif conductance $\phi_M(S_r)$ (first dichotomization) cannot give rise to CP structure, since neither of them can ensure that their degrees are larger than the average degree of the network. However there are local minimum points on the both sides of the smallest motif conductance. We then respectively dichotomize the two local points to check whether there are cores in the both sets (as shown in the inset of Fig. 6(a)). Luckily, two cores (red nodes in the inset of Fig. 6(a)) are detected based on the second dichotomization regarding to the two local minimum points. As illustrated in Fig. 6(b): red and purple nodes are two cores.

The second network is the Political blogs network (PB), which has 1222 nodes and 16714 connections in the network [26]. The nodes of this network are blogs about US politics and the edges are hyperlinks between these blogs. Since this network displays a marked division into groups of conservative and liberal blogs, which has been viewed as a typical example of community structure [27]. Also, Fig. 6(c) indicates that the first dichotomization based on the smallest motif conductance cannot find the core and the periphery. However, the second di-
C. Detection of global CP structure by joining a leader node

Take a schematic illustration in Fig. 7(a) as an example, there are two cores $C_1$ and $C_2$ in two separated communities, dichotomization based on Algorithm 1 gives rise to two communities, but the cores in communities cannot be detected (see Fig. 7(a)). Therefore, we further developed Algorithm 2 to detect multiple cores. For example, by implementing Algorithm 2 on the Email network [28], several local cores are detected in the network (Fig. 8(a), green nodes are peripheral nodes and the nodes labeled by other colors are core nodes). However, the network’s integrity is destroyed due to the multiple dichotomizations (see Fig. 8(a)). Sometimes, we are more concerned with whether there is a global core (or more precisely, “hidden” global core) from the whole network perspective. To do this, we need to design a method to detect the global core, no matter whether the network is a single CP structure or multiple CP structure structure network.

To overcome the difficulty in Algorithm 1, we can join a leader node into the network, which connects with all the nodes in the network (see Fig. 7(c)). The motivation is that, once one leader is joined into the network and with the topological transformation function $W_M$, the weight among core nodes and the weight among peripheral nodes are respectively increased, but the weight between core nodes and peripheral nodes is increased slightly. Now the dichotomization based on Algorithm 1 can detect the global core and its periphery (see the schematic illustration in Fig. 7(d)). More importantly, such a global core still includes the local cores in different communities. As shown in Fig. 8(b), the global core (red nodes) is obtained in Email network after joining a leader node, which not only contains the local cores detected by Algorithm 2 but also contains some peripheral nodes who are misclassified due to multiple partition.

VI. CONCLUSIONS

In this paper, we have defined a 3-tuple motif to detect CP structure, and then the motif adjacency matrix based on the 3-tuple motif is constructed. At last, the detection of CP structure can be realized by the smallest motif conductance, which is obtained by applying spectral clustering on the motif adjacency matrix. Our method has the following advantages: 1) our method can detect not only single or multiple CP structure, but also local or global CP structure; 2) our method is fast and can apply large-scale networks. The complexity of our algorithm mainly depends on the computations of the motif adjacency matrix and an eigenvector. As stated in the supplementary material of Ref. [22], the motif adjacency matrix can be computed in $O((m+n)\log n)^{O(1)}$ time and the eigenvector can be computed in $O\left(\left((m+n)\log n\right)^{O(1)}\right)$ time by using fast Laplacian solvers [29]. So our algorithm is fast, for instance, the CP structure in the BlogCatalog network [30] can be quickly detected by our method (see Fig. 9). This network has 10312 nodes, 333983 links and the average degree is 64.78. It takes 1856.85 seconds when using MATLAB2015b to implement our algorithm on a PC with 3.60GHZ Inter Core i7-4790 CPU 3.60GHZ and the Windows 7 64bit operating system; 3) our algorithm has the free parameter, which can avoid to preset the size of core or the number of cores.

REFERENCES

Fig. 9. Detection of CP structure in BlogCatalog network based on Algorithm 1, where red nodes and green nodes are core and periphery, respectively. The sizes of nodes denote their degrees.


