Journal of Network Intelligence ©2022 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **7**, Number **1**, February **2022**

Overlapping Community Detection by Motif-aware Label Propagation

Jian-Chao Liao

College of Computer and Information Engineering
Xiamen University of Technology
Xiamen, 361024, P. R. China
2022031482@stu.xmut.edu.cn

Tong Li

College of Computer and Information Engineering
Xiamen University of Technology
Xiamen, 361024, P. R. China
2022031405@stu.xmut.edu.cn

Wei Weng*

College of Computer and Information Engineering
Xiamen University of Technology
Xiamen, 361024, P. R. China
*Corresponding Author: xmutwei@163.com

Jin-Bo Wang*

School of Economics
Xiamen University
Xiamen, 361005, P. R. China
*Corresponding Author: wangjb168@163.com

Juan Wen

School of Economics
Xiamen University
Xiamen, 361005, P. R. China
wenjuan@xmu.edu.cn

Received September 2021; revised November 2021

Abstract. *Community detection is a fundamental problem for analyzing the structural properties of complex networks. The existing approaches focus on detecting disjoint communities, even though communities in real graphs are well known to be overlapping. Another shortcoming is those approaches usually exploit the lower-order adjacent information and neglect the higher-order connective structure on networks. We deal with those problems by motif-aware Label Propagation. Specifically, the motif-based hypergraph is constructed to encode the higher-order structural characteristic of the network. Then, we unify the structure of hypergraph and original graph, calculate the belonging coefficient in the propagation process, and map each node to multiple community labels. Experimental results on multiple datasets have shown the superiority of the proposed method in improving the community detection performance.*
Keywords: Community detection, Motif, label propagation

---

1. **Introduction.** Graphs or networks provide a natural way of representing complex real-word systems, where the nodes stand for elementary units of the system and the edges represent their relations. Network data arises in a wide range of fields, such as social networks, collaboration networks, communication networks, biological networks and food webs [1]. Community detection aims to partition the network into subsets of vertices of relatively denser connections [2, 3]. It has been shown that many real-world networks have a significant property of community structure, and community detection is a key technique to understand the structure and function of the complex systems represented as networks. For example, in protein-protein interaction networks, community detection can identify function modules [4]; in World Wide Web, it can find groups of webpages associated with similar topics [5].

Lots of community detection methods have been proposed in the literature, including those based on modularity-maximization [6], generative and statistical models [7], local-metric optimization [8], hierarchical clustering [9, 10], spectral-clustering [11], link prediction [12] and matrix factorization [13]. Existing methods usually assume that each node belongs to exactly one community, i.e., they group the nodes into a set of disjoint clusters. In practice, such an assumption is rarely satisfied [14, 15]. For example, in social networks, it is expected that the individuals belong to multiple interest groups [16]; in collaboration networks, scientists may collaborate with several research groups [17]. In those scenarios, vertices can belong to more than one community, then the boundaries of communities overlap. Overlapping community detection is a much harder problem, as the number of communities a node belongs to is uncertain.

Another shortcoming is that those methods operate on simple graphs, in which edges are not denoted with weights to describe the degree of relations among nodes. In other words, those edges describe only the "existence" of relations between the connected pair of nodes. Simple graphs only possess the lower-order structure of the network and only the direct one-hop connections between nodes are taken into consideration, so existing community detection models learning from simple graphs might be suboptimal [18]. On the other hand, simple graphs constitute the majority of the real-world applications, so it is important to explore the higher-order structure among nodes from those simple graphs. In order to capture structural features of the graphs, one popular strategy is motif-based higher-order structure mining. Motifs are defined as recurring, significant patterns of interconnections or simply the building blocks of the networks [19, 20]. As it can provide higher-order connectivity patterns and are crucial for uncovering the organization of complex networks, motif-based structure has been gaining increasing attention in community detection [21].

Most real-world networks are large scale. For example, popular social networks, such as Facebook and Wechat, have reached hundreds of millions or even billions of users. Identifying communities in such big networks requires the algorithms to be computationally efficient. The majority of traditional community detection algorithms, like modularity or matrix factorization based algorithms, are not suitable for such circumstance, as they typically take $O(n^2)$ for optimizing and are too slow. On the other hand, label propagation-based community detection methods have been shown to be nearly linear time complexity and perform efficiently for its simplicity. For example, RAK [22] is one of the representative label propagation algorithm. It assigns each node in the network with a label denoting the community to which it belongs. In the process of label propagation, the node update its label with the labels belonging to most of its neighbors. RAK claimed that 95% of nodes can be identified correctly after five iterations, i.e., it has nearly linear time computational complexity. Because of its simplicity and effectiveness, there are a amount of methods which utilize label propagation for community detection [23–27].

Although the merits of motif and label propagation are obvious, many community detection algorithms utilize them independently. To the best of our knowledge, the first work combining the motif and label propagation for community detection was presented by Li et al. [18] recently, named MWLP. It exploits higher-order structure characteristics of the network by means of motif and designs a re-weighted network, then employs a novel voting strategy for label propagation. Experimental results revealed the superiority of MWLP for community detection. However, it identifies the disjoint community structure while can not deal with the problem of overlapping community. In terms of node confidence calculation, Wang et al. [28] proposed a confidence calculation method to improve the calculation accuracy. Kang et al. [29] use a non-inertial particle swarm optimization with elite mutation-Gaussian process regression is proposed to optimize the hyper-parameters of GRP. Yang et al. [30] proposed a multigroup multistrategy SCA algorithm. Dong et al. [31] proposed a Model Affiliation Graph with Interacting Communities. In this paper, we extended the MWLP and propose a novel method to deal with overlapping community detection for large scale networks. The main contribution of this paper is that the proposed method considers both high-order and low-order structures when detecting networks. Then in the process of label propagation, nodes can belong to multiple communities at the same time, and the label of the node itself is considered. This method can effectively improve the classification results of community detection.

2. **Related works.** In 2005, the first algorithm that deal with overlapping communities was proposed by Palla et al. [32]. They constructed a community with a series of k-cliques such that any one clique can be reached from another via a series of adjacent k-cliques. A multitude of overlapping community detection algorithms were proposed to date, which mainly falls into five broad categories: edge partition [15], matrix factorization [33], clique searching and merging [34, 35], label propagating and spectral method. In this section, we focus those representative algorithms which utilize label propagation for overlapping community detection. Another focus is motif, i.e., small patterns that appear frequently in a variety of graphs. However, though few methods dealing with motif clustering on disjoint community detection have been proposed, the motifs based overlapping community detection methods have not yet been comprehensively investigated.

2.1. **Label propagation.** Rahavan et al. [22] proposed the first label propagation based algorithm RAK for community detection. It initializes each node with a unique label. Then, RAK spread labels through the edges of the graph and updates the label of a node to the label that appears most frequently among its neighbours, iterating until a general consensus is reached. Because of its intuitive and quasi-linear time complexity, a number of algorithms are built based on label propagation for community detection, including a few work for overlapping community detection. COPRA [36] is extended from RAK and tries to find overlapping communities in the network. It assigns each node with one or more labels with different belonging factors. In each iteration the belonging factors would be updated and normalized according to the union of its neighbours. Because the degree of overlap is controlled by a hyper-parameter, this method detects inaccurate communities in high overlapping networks. DEMON [37] considers that the generic community detection algorithms are not suitable to deal with large scale and dense networks, as they typically try to cluster the whole structure and return some huge communities and a long list of small branches. To deal with this problem, it reveals local community structure for each node in its ego neighborhood through label propagation. For the current node, its local communities are merged with previous found communities into a global collection. As different node may share neighbors, their local communities or their merged communities may overlap. SLPA [38] is also an extension of the RAK. It employed a speaker-listener based information propagation process to update labels. Specifically, in each loop a node is selected as a listener, and each neighbor send out a label which is most frequent among its label set, then the listener accept the most popular label from the collection of labels received from neighbors. When the relatively stable outputs is produced, SLPA deleted the label whose occurrence is less than a given threshold, then connected nodes having a particular label are grouped together and form a community. NI-LPA [39] sorts the nodes in a fixed order to deal with the LPA instability problem, and allowed a node to contain many community labels for overlapping community detection. And, each label is associated with a belonging coefficient which is computed according to the node degree and its adjacent information among its neighbors. In propagation process, for each node NI-LPA uses the asynchronous mode to sum the coefficients of the same label among its neighbors and store label with the best coefficient.

2.2. **Motifs in Community Detection.** Real-world complex networks usually present and share patterns of interconnections or small subgraphs, called network motifs, occurring with a frequency higher than that in a random network [20, 40]. High frequencies of certain motifs indicate the important functions they play in a network, and can be recognized as fundamental units of networks [41]. It is surprising that there are few studies to explore the role those motifs have in community detection. Arenas et al. [42] generalized Newman-Girvan modularity [2] with motifs, and showed that it can reveal more detailed subdivision with respect to those obtained by optimizing the standard modularity. Benson et al. [43] encoded the higher-order structures of networks by means of tensors, then a tensor spectral clustering method was proposed to search a partitioning that does not cut the motifs. Another work [41] by the same authors extended the spectral graph clustering methodology to account for higher-order structures in networks, in which the entries of the adjacency matrix record the co-occurrence counts of corresponding nodes in an instance of motif. Li et al. [44] considered that the higher-order connections encoded in the motif-based hypergraph usually violate the original lower-order topological structure and may be fragmented, therefore, it will render the community structure with instability and degenerate the performance of the algorithm. To address the fragmentation issue, they proposed the algorithm termed EdMot to enhance the connectivity structure of the input network by integrating both the motif-based structure and the lower-order structure. MLM-MOGA [45] extracts a clustering on a multiplex network based on multi-objective optimization, in which it maximizes the number of instances of a motif inside the same community, while minimizing cutting instances of the same motif on all layers.

3. **The proposed algorithm.** In this section, we elaborate how we apply the motif and label propagation for overlapping community detection. Our method is adapted from COPRA [36], which is a well-established overlapping community method based on label propagation. COPRA utilizes the lower-order topology structure information, i.e., at the level of edges and nodes for label propagation, while we incorporate higher-order structure information into lower-order structures for label propagation. Motivated by MWLP [18], we adopt motif-based strategy to mine higher-order structure information, and then preserve both higher-order and lower-order structure by re-weighting the node adjacency matrix. Compared with the existing work, we add higher-order structure detection to the traditional overlapping community detection method, and then combine it with the original network. In the process of propagation, we also attach the label information of the

node itself, which can better find the real structure distribution of the network. Our algorithm have three key operations: motif-based higher-order structure mining, Multi-Label annotation and motif-aware weighted label propagation.

3.1. **Motif-based higher-order structure mining.** Given a network $G = (V, E)$, where V is the node set and E is the edge set. Let $n = |V|$ and $m = |E|$ represent the number of nodes and the number of edges, respectively. In this paper, we introduced the Motif-based adjacency matrix W_{M_m} [46] to obtain the higher-order structural features of the network. As shown in Eq.1, W_{M_m} can be expressed as follows:

$$W_{M_m}(i,j) = I_{ij}^{M_m}, \tag{1}$$

where $I_{ij}^{M_m}$ represents the number of the motif M_m containing both node i and node j. The most common higher-order structure is the small network subgraph, and its different types will reveal different structures. Many researchers have made great efforts and achieved good results for motif recognition, such FANMOD [47], Kavosh [48] and so on. Figure 1 shows the directional motif types for all three nodes. In this experiment, we focus on triangular motifs as the research object. The triangle is the main type of most network motifs, and its nature also reflects the communication relations in the social network. After determining the motif type, we can use the identified motif to reveal the higher-order structure of the network.

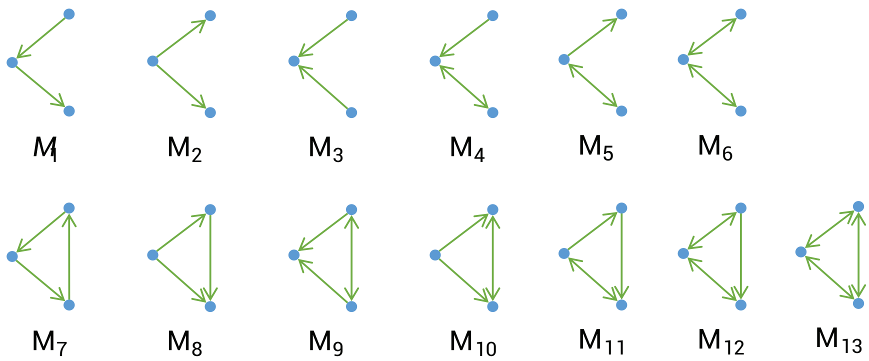

FIGURE 1. 13 kinds of triangle connected subgraph.

It has been shown in [18] that the bidirectional relationship of nodes can better reflect the connection structure of nodes. As shown in Figure 1, there are many bidirectional Motifs to reflect the higher-order structure information. According to the suggestion in [18], we choose the bidirectional Motif M_{13} in our experiments. The resulting adjacency matrix W_{M_m} can be viewed as a weighted matrix. Larger weight means that the connections between corresponding nodes are closer. Futhermore, there are not only higher-order connections but also lower-order connections in the network. If only focus on higher-order structure information, then the lower-order structure information will be lost. Therefore, we adopt a re-weighted network framework [18] to maintain higher-order and lower-order structural information. The weighted network can be expressed in matrix form as follows:

$$W = A + W_{M_m}, \tag{2}$$

where A is the adjacency matrix of the orginal network, in which the edges of the original network can be weighted or unweighted. W_{M_m} is an adjacency matrix with higher-order structure features based on Motif. After this transformation, the relationship between the elements of W can be seen as a measure of the intimacy between nodes. Therefore, we solve the problem that higher-order and lower-order connections can not be utilized at the same time. The structure mining process is described in Algorithm 1.

Algorithm 1 Motif-based higher-order structure mining

Input: A network $G = (V, E)$.

Output: An adjacency matrix W containing information of higher-order and lower-order structures of the network.

1: Identify the type of motif, i.e., M_m.
2: The motif adjacency matrix W_{M_m} is obtained by Eq.1.
3: The re-weighted network W is constructed by Eq.2.

3.2. **Multi-label annotation.** Firstly, we treat each node as a separate community. Like COPRA, each node is initialized to a unique label which indicates its community. Obviously, in the scenario of overlapping communities, each node may belong to multiple communities. Therefore, we need to allow each node to contain multiple labels. We associate each node n_i with a set of pairs $(l_1^i, b_1^i), (l_2^i, b_2^i), ..., (l_N^i, b_N^i)$, where l_x^i stands for the $x - th$ label of the node n_i, and b_x^i is the belonging coefficient of the node belonging to the label l_x^i. For each node n_i, the sum of all the belonging coefficient is 1, i.e., $\sum_{x=1}^{N} b_x^i = 1$.

In each propagation step, we use synchronous update propagation, where the node labels in the $t - th$ iteration are always based on the neighborhood set labels in the $(t-1)th$ iteration. However, this method will make the label include all neighborhoods. We only need to keep the community that each node is most likely to belong to, not all communities. In our method, we first construct the node labels as described above. Then we set a parameter v to indicate the maximum number of communities that each node can belong to. We set the threshold to the reciprocal of v, pairs with an belonging coefficient less than $\frac{1}{v}$ is then deleted.

As mentioned in the [36], a situation that may arise during the selection process is that all pairs in a node may have belonging coefficients less than the threshold. If so, we select only the pair with the largest belonging coefficient and delete the others. If more than one pair has the same maximum belonging coefficient below the threshold, we keep a randomly selected pair among them.

3.3. **Motif-Aware Weighted Label Propagation.** Different from the traditional label propagation algorithm, our label propagation method is performed on a re-weighted network. The pseudo-code description is shown later. For the convenience of illustration, we designed two simple network and then run algorithms on it. The network description and running results of example 1 are shown in Table 1. In (a), $Node$ represents the node, and $Neighbor$ represents its neighbors. The weight relationship between nodes is shown in (b). For nodes that are not connected to each other, we set their weight relationship to 0. Since we will use the label of the node itself in the propagation process, we set the weight of the node to itself to 1. Then, the algorithm propagation result is shown in (c). N represents the node, and t_n represents the $n - th$ run result. In each result, the left side of the bracket represents the label, and the right side represents the belonging coefficient. Our algorithm ends after propagating three times on the example. The result of the third time is the same as the second time, so it will not be shown separately.

Taking the first node as an example, we first initialize the label pair of the node to $(1, 1)$. When propagating, we set the parameter $v = 1$, which means that each node can

Table 1. The results of the proposed MCOPRA method running on the example 1

Node	*Neighbor*
1	2 5 6
2	1 5 6
3	4 5 6 7
4	3 5 6
5	1 2 3 4
6	1 2 3 4
7	3 8 9
8	7 9
9	7 8

(a)

	1	2	3	4	5	6	7	8	9
1	1	3	0	0	2	2	0	0	0
2	3	1	0	0	2	2	0	0	0
3	0	0	1	3	2	2	1	0	0
4	0	0	3	1	2	2	0	0	0
5	2	2	2	2	1	0	0	0	0
6	2	2	2	2	0	1	0	0	0
7	0	0	1	0	0	0	1	2	2
8	0	0	0	0	0	0	2	1	2
9	0	0	0	0	0	0	2	2	1

(b)

N	t_0	t_1	t_2
1	[1,1]	[2,1]	[3,1]
2	[2,1]	[1,1]	[3,1]
3	[3,1]	[4,1]	[3,1]
4	[4,1]	[3,1]	[3,1]
5	[5,1]	[3,1]	[3,1]
6	[6,1]	[3,1]	[3,1]
7	[7,1]	[9,1]	[9,1]
8	[8,1]	[9,1]	[9,1]
9	[9,1]	[8,1]	[9,1]

(c)

only select one label. In the first propagation, we know that node 1 has three neighbors $2, 5, 6$. Their corresponding labels are $2, 5, 6$, respectively. We know that the label of the first neighbor of node 1 is 2, and the belonging coefficient is 1. Then, after multiplying the weight between nodes by the belonging coefficient, the first neighborhood label pair $(2, 3)$ is obtained. By analogy, plus the label pair of node 1 itself, we get the neighborhood label set $[(2, 3), (5, 2), (6, 2), (1, 1)]$. After that, the label set is normalized to $[(2, 0.375), (5, 0.25), (6, 0.25), (1, 0.125)]$. Since the belonging coefficient of label 2 is the largest, the label of node 1 is set to 2, and the coefficient is set to 1. Therefore, after the first iteration, we get the label pair of node 1 as $(2, 1)$. The propagation process of the remaining nodes is the same. The algorithm stops after three rounds of propagation. Finally, we get two communities. The first community contains nodes $(1, 2, 3, 4, 5, 6)$, and the second community contains nodes $(7, 8, 9)$. The results of the algorithm are consistent with the expected results. Intuitively, the running process of the algorithm is shown in Figure 2.

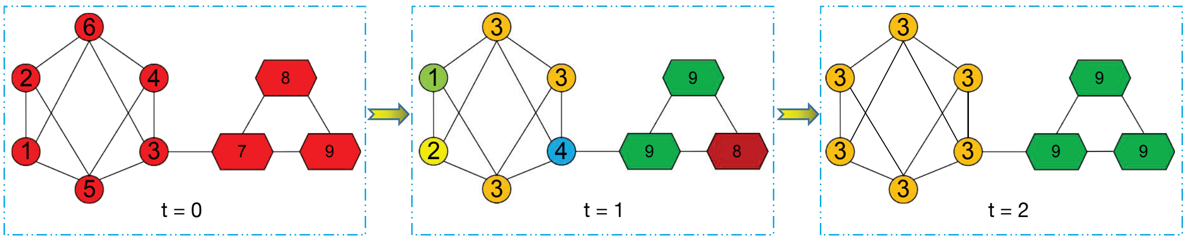

Figure 2. The results of MCOPRA running on the example 1.

In order to verify the effectiveness of the proposed algorithm for detecting overlapping communities, we conduct experiments on another example. We take the left part of example 1 as this example. The running process is shown in Figure 3 and Table 2. In this example, we set the parameter $v = 3$. We use the propagation process of node 5 to illustrate one of the experimental steps. First, we give node 5 an initial label pair $(5, 1)$. Then, in the first propagation, according to its neighbors, the label pairs we get are $[(1, 2), (2, 2), (3, 2), (4, 2), (5, 1)]$. Through normalization, we get the label pair [(1,0.22),(2,0.22),(3,0.22),(4,0.22),(5,0.11)]. The belonging coefficients of all label pairs are less than the threshold $1/3$. Among them, the belonging coefficients of labels $1, 2, 3, 4$ are equal and greater than the coefficient of label 5, so we randomly select a label from these four labels. Finally, the label pair of node 5 is $(1, 1)$. The propagation process is the same for the next few times. In the fourth propagation, we can get the label pair of node 5 as $[(1, 4.57), (4, 4.43)]$. The normalized label pair is $[(1, 0.51), (4, 0.49)]$. Because the belonging coefficients of these two pairs are greater than the threshold, both label pairs will be retained. Finally, the label pair of node 5 is $[(1, 0.51), (4, 0.49)]$. The algorithm stops after 4 propagations. In the end, the communities we get are $(1, 2, 5, 6)$ and $(3, 4, 5, 6)$. Among them, nodes 5 and 6 are overlapping nodes. Finally, the experimental results prove that our algorithm can effectively detect overlapping communities, and the results are the same as expected.

TABLE 2. The results of MCOPRA running on the example 2

N	t0	t1	t2	t3	t4
1	[1,1]	[2,1]	[1,1]	[1,1]	[1,1]
2	[2,1]	[1,1]	[2,0.5;1,0.5]	[1,1]	[1,1]
3	[3,1]	[4,1]	[3,0.5;4,0.5]	[4,1]	[4,1]
4	[4,1]	[3,1]	[4,1]	[4,1]	[4,1]
5	[5,1]	[1,1]	[1,1]	[1,0.57;4,0.43]	[1,0.51;4,0.49]
6	[6,1]	[4,1]	[4,1]	[1,0.43;4,0.57]	[1,0.49;4,0.51]

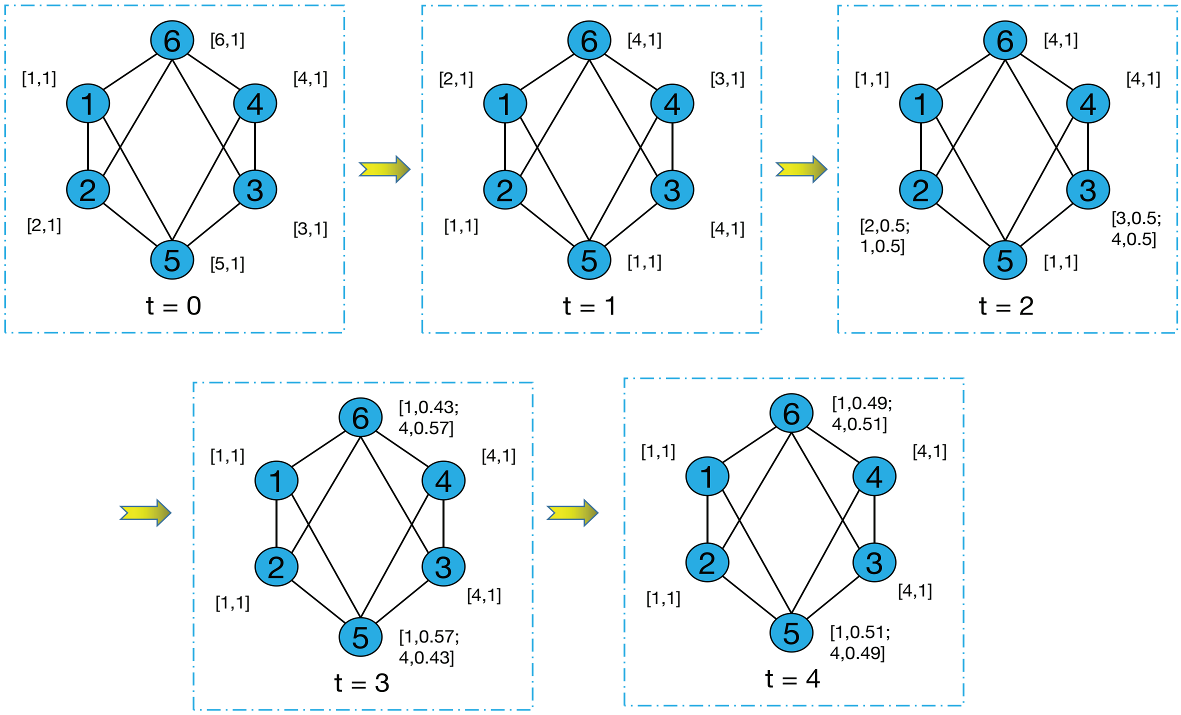

FIGURE 3. The results of MCOPRA running on the example 2.

Based on the results of the previous two examples, we summarize the pseudo-code of the complete MCOPRA algorithm as shown in Algorithm 2 and 3. Algorithm 2 illustrates the overall flow of our algorithm. In Algorithm 3, the detailed steps of *propagation* and *normalization* are explained. $N(x)$ is the neighborhood set of node x. The algorithm contains two node label vectors *new* and *old*. *new*.x and *old*.x represent the latest label and the previous label of node x, respectively. For the label of node x, it contains many pairs (c, b), where c and b represent the label and the belonging coefficient, respectively. $b \leftarrow b_y$ represents $b \leftarrow b_y w_{xy}$ in the propagation process, where w_{xy} represents the weight relationship between nodes x and y. t represents the number of iterations. When t reaches the maximum number of iterations, the program stops.

Algorithm 2 motif-based propagation process

Input: An adjacency matrix W.
Output: Node community classification result.

Initialize each node label : $old.x \leftarrow \{(x,1)\}$.
repeat
  repeat each node :
    $Propagate(x, old, new)$.
  end
  If $id(old) = id(new) : min \leftarrow mc(min, count(new))$.
  Else : $min \leftarrow count(new)$.
  If $min \neq oldmin$: $old \leftarrow new$ and $oldmin \leftarrow min$.
until $min = oldmin$ or $t > T$.
repeat each node x:
  $ids \leftarrow id(old.x)$.
  repeat each c in ids:
    If, for some g, (c,g) is in $coms$, (c,i) in sub:
      $coms \leftarrow coms – \{(c,g)\} \cup \{(c, g \cup \{x\})\}$.
      $sub \leftarrow sub – \{(c,i)\} \cup \{(c, i \cap ids)\}$.
    Else:
      $coms \leftarrow coms \cup \{(c, \{x\})\}$.
      $sub \leftarrow sub \cup \{(c, ids)\}$.
  end
end
For each (c,i) in sub: If $i \neq \{\}$: $coms \leftarrow coms – (c,g)$.

4. Experiments.

In this section, we applied the proposed algorithm MCOPRA on two types of networks to evalute effectiveness. The one is run on real-world networks, a few of them we know their community structure, and for many of them their community structure are not clear. The other is the synthetic network, we adjust their structure by modifying coresponding parameters.

4.1. Datasets.

(1)Synthetic networks. In order to verify the effective and accuracy of our algorithm, we generate the LFR [49] benchmark as our experimental network. The number of nodes has increased from 1000 to 20000 and mixing parameter(mu) has increased from 0 to 0.8. In Tabel 3, n is the number of nodes in the network, k and $maxk$ represent the average degree and the maximum average degree of the network, respectively. $Minc$ and $maxc$ are denote minimum and maximum numbers of communitise, respectively. The mixing parameter mu represents the ratio of the degree of external of a node relative to its community to the total degree of the node. On and Om are the number of the overlapping nodes and the community memberships of each overlapping node.

We use five different networks size, small networks(1000 nodes, 2000 nodes and 5000 nodes) and large networks(10000 nodes and 20000 nodes). And for a given size, we use two different types: overlapping communities and non-overlapping communities. For both communities share some common parameters: $k = 10$, $maxk = 30$, $minc = 10$, $maxc = 50$. The different is that we set On to 0 and Om to 0 on the non-overlapping community. In overlapping communities, we set On to be 1/100th of the number of nodes n and Om to be fixed at 2.

Algorithm 3 motif-based propagation process (contd.)

$Propagate(x, source, dest)$:

$dest.x \leftarrow \{\}$.
repeat each y in $N(x)$:
    repeat each (c, b_y) in $source.y$:
        $b \leftarrow b_y$.
        If, for some $b_x, (c, b_x)$ is in $dest.x$: $dest.x \leftarrow dest.x - \{(c, b_x)\} \cup \{(c, b_x + b)\}$.
        Else : $dest.x \leftarrow dest.x \cup \{(c, b)\}$.
    end
end
repeat each (c, b_y) in x:
    $b \leftarrow b_y$.
    If, for some $b_x, (c, b_x)$ is in $dest.x$: $dest.x \leftarrow dest.x - \{(c, b_x)\} \cup \{(c, b_x + b)\}$.
    Else : $dest.x \leftarrow dest.x \cup \{(c, b)\}$.
end
$Normalize(dest.x)$.
$b_{max} = 0$.
repeat each (c, b) in $dest.x$:
    If $b < \frac{1}{v}$:
        $dest.x \leftarrow dest.x - \{(c, b)\}$.
        If $b > b_{max}$:
            $b_{max} \leftarrow b$ and $c_{max} \leftarrow c$.
end
If $dest.x = \{\}$: $dest.x \leftarrow \{(c_{max}, 1)\}$.
Else : $Normalize(dest.x)$.

$Normalize(l)$:

$sum = 0$.
repeat each (c, b) in l:
    $sum \leftarrow sum + b$.
end
repeat each (c, b) in l:
    $l \leftarrow l - \{(c, b)\} \cup \{(c, b/sum)\}$.
end

TABLE 3. The parameters used for the generation of synthetic networks

network	n	*k*	*maxk*	*minc*	*maxc*	*on*	*om*	*mu*
LFR1	1000	10	30	10	50	0	0	0-0.4
LFR2	1000	10	30	10	50	10	2	0-0.5
LFR3	2000	10	30	10	50	0	0	0-0.4
LFR4	2000	10	30	10	50	20	2	0-0.5
LFR5	5000	10	30	10	50	0	0	0-0.4
LFR6	5000	10	30	10	50	50	2	0-0.5
LFR7	10000	10	30	10	50	0	0	0-0.4
LFR8	10000	10	30	10	50	100	2	0-0.5
LFR9	20000	10	30	10	50	0	0	0-0.4
LFR10	20000	10	30	10	50	200	2	0-0.5

(2)Real-world networks. In real-world networks, we use karate, football and dolphin.

The node communities of these networks are known, and several of their properties are shown in Table 4. In which, N and E are the number of nodes and edges, C is the number of communities. A brief description of these three networks is given below.

TABLE 4. Properties of real-world networks

Datasets	N	E	C
Karate [50]	34	78	2
Dolphin [51]	62	159	2
Football [52]	115	613	12

Karate network [50]: The Zachary Network is a social network based on observations of a karate club at an American University. Zachary observed and described the network consists of 34 nodes and 78 edges, where nodes represent members of a club and edges represent friendships among members. Now the karate club network has become a well-known benchmark network in complex social network.

Dolphin network [51]: The second network is the dolphin network. The Dolphin dataset is a network of social relationships obtained by Lusseau et al., who observed the communication of 62 groups of dolphins in New Zealand's Doubrucsound Strait over a period of 7 years. The network has 62 nodes and 159 edges. Nodes in the network represent dolphins, and edges represent a frequent contact between dolphins.

Football network [52]: The third real network is the college football network. Newman created a complex social network based on the national college football association. The network contained 115 nodes and 616 edges, where the nodes represent football teams and the edges between two nodes indicates that a game has been played between two teams. The 115 teams are divided into 12 leagues. The format of the competition is that teams within the league play group games, followed by games between teams within the league. An alliance can then be represented as the real community structure of the network.

For the other networks, we don't know what the real community is. These networks are shown in Table 5. In which N and E are the number of nodes and edges, D represents the network node density. $Dmax$, $Dmin$ and $Davg$ indicate the maximum, minimum and average degrees of node, respectively. The property $Assor$ represents the correlation of degrees, and is used to examine whether vertices with similar degree values tend to be connected to each other.

TABLE 5. Properties of real-world networks

Datasets	N	E	D	$Dmax$	$Dmin$	$Davg$	$Assor$
Polbooks	105	441	0.081	25	2	8	-0.128
Jazz	198	2.7k	0.141	100	1	27	0.02
Lesmis	77	254	0.087	36	1	6	-0.165
Email-univ	1.1k	5.5k	0.009	71	1	9	0.078
Polblogs	1.5k	19k	0.017	467	0	25	-0.196
ca-GrQc	5.2k	14.5k	0.00181	81	0	5	0.659
TerroristRel	881	8.6k	0.022	36	1	19	0.851

4.2. **Comparison methods.** We select three community detection algorithms as a comparison method, all three algorithms can detect overlapping communities.

COPRA [36]: COPRA is a community detection algorithm based on label propagation proposed by Gregory in 2010. This algorithm can be regarded as an improved version of RAK. The biggest improvement of COPRA algorithm over RAK algorithm is that

COPRA can be used to find overlapping communities, while RAK algorithm can only be used to find non-overlapping communities.

NI-LPA [39]: NI-LPA is a method proposed by Ben et al., in 2019. It is an algorithm that focuses on the function of nodes. Because there are only two cases where the label has a high attribution factor: the label is sent by multiple source nodes, or the label is sent by an important node. Therefore, after the propagation phase, the pairs with ownership coefficient less than a certain threshold are deleted, and only the labels with the best ownership coefficient are retained.

LFM [53]: In 2009, Lancichinetti proposed a LFM algorithm based on the optimal fitness function. The algorithm needs to calculate the difference of fitness when the nodes join the community. The most important thing is that some nodes in the network may have high fitness for multiple communities.

4.3. **Evaluation metrics.**

Normalized mutual information(NMI): For overlapping community, we can use the extended NMI [53] to judge the difference between the communities by the algorithm and the real communities. Its values are distributed between 0 and 1. The higher the NMI value is, the more accurate the partition results are, and the lower the NMI value is, the less similar the communities are. For partition C', Eq.3,4 gives the main formula of overlapping NMI.

$$NMI = 1 - \frac{1}{2}[H(X|Y) + H(Y|X)], \tag{3}$$

$$H(X|Y) = \frac{1}{C'} \sum_k \frac{H(X_k|Y)}{H(X_k)}. \tag{4}$$

Modularity(Q): For non-overlapping communities, we need to evaluate the quality of the community through quantitative indicators, so as to further measure the advantages and disadvantages of several community detection algorithms. In order to evaluate the division of community, one idea is to ensure that there are as many edges within the community as possible. Newman and Girvan proposed a measure to evaluate the quality of community division, called the modulariy measure Q [54]. The modularity Q means that the value of Q tends to 1, the more obvious the modularity degree of the generated community division is. Therefore, the measurement standard proposed by the definition is :

$$Q = \sum_i (e_{ii} - {a_i}^2) = Tre - \left\| e^2 \right\|, \tag{5}$$

where $\|\gamma^2\|$ represents the sum of all the elements in the matrix γ. Eq.5 represents the ratio of the edges connecting two nodes of the same type in a network to the expected ratio of the edges connecting any two nodes in the same community structure.

Normalized mutual information(EQ): Modularity Q [54] was originally proposed by Newman and Girvan for non-overlapping community structure. However, with the development of network, modularity can not be applied to complex network structure. Therefore, Shen et al. [55] proposed a function to measure the performance of overlapping community detection algorithm based on the number of overlapping members. A high value of modularity indicates that the network partition is significant. In this paper, we use the extended modularity Eq.6 as follows:

$$EQ = \frac{1}{2m} \sum_{i} \sum_{u \epsilon c_i, v \epsilon c_i} \frac{1}{Q_u Q_v} [A_{uv} - \frac{k_u k_v}{2m}], \quad (6)$$

where Q_u represents the community to which node u belongs, A is the adjacency matrix of the network, k_u represents the degree of node u, and m is the total number of edges in the network.

4.4. **Experimental results and discussion.**
(1) Experimental results on modularity Q. Based on the LFR synthetic network, we generated five kinds of non-overlapping networks. Then we evaluate the performance of the algorithms using the modularity Q. The results of these algorithms on non-overlapping networks are shown in Table 6. For both MCOPRA and COPRA, we set the parameter $v = 1$ of their algorithm. It can be seen from the table that with the increase of mixing parameters mu, the values of Q obtained by each algorithm begins to decrease. In the case of the same value of mu, we can find that as the number of network nodes increases, the performance of our algorithm is not greatly affected. Therefore, we can know that our algorithm can still obtain a high Q value even when the number of network nodes is large or small. As the results show, in most scenarios, our algorithm performance is better than the COPRA algorithm. Finally, we conclude that the method has good stability. By changing the number of nodes in the network, or changing the value of mu, our algorithm can still detect the community structure well. The best results on each dataset are highlighted in bold, the rest of the tables are the same.

Table 6. Experimental results Q on non-overlapping communities in synthetic network.

	LFR1		LFR3		LFR5		LFR7		LFR9	
mu	MCOPRA	COPRA	MCOPRA	COPRA	MCOPRA	COPRA	MCOPRA	COPRA	MCOPRA	COPRA
0.0	**0.9994**	0.9481	**0.9931**	0.9265	**0.9937**	0.9419	**0.9924**	0.9693	**0.9969**	0.9653
0.1	**0.9011**	0.8521	**0.9012**	0.8594	**0.8951**	0.8546	**0.8926**	0.8730	**0.8968**	0.8735
0.2	**0.7996**	0.7829	**0.7854**	0.7663	**0.7915**	0.7631	**0.7905**	0.7659	**0.7914**	0.7771
0.3	**0.6913**	0.6904	**0.6914**	0.6724	0.6836	**0.6877**	**0.6871**	0.6686	**0.6868**	0.6730
0.4	**0.5918**	0.5778	**0.5800**	0.5506	**0.5814**	0.5741	0.5763	**0.5793**	**0.5783**	0.5639

(2) Experimental results on NMI. In order to verify the effectiveness of the algorithm in the detection of overlapping communities, we also conducted experiments under five overlapping networks generated by the synthetic network. For both COPRA and MCOPRA, we set the parameter $v = 5$ of their algorithm. Figure 4 shows the experimental results of several algorithms in overlapping communities. In the table, the symbol "–" means that the algorithm has been unable to find the community structure. The rest of the tables below are the same. We measure the quality of the detected network community structure by calculating the NMI value of overlapping communities. As can be seen from Figure 4, compared with the other three algorithms, MCOPRA has advantages in most cases. And we can find that MCOPRA is in a relatively stable state between the value of mu from 0.0 to 0.4, which reflects the stability of our algorithm. When the value of mu is 0.5, some algorithms can no longer detect the community structure, but the performance of our algorithm is still better than other algorithms. Finally, experimental results also show that our algorithm can maintain great performance in the case of simple or complex community structure.
(3) Experimental results on EQ. EQ values of the four algorithms-LFM [53], COPRA [36], NI-LPA [39] are listed in Table 7. It can be seen from the table that our algorithm has the best classification effect in most cases. In several networks such as email, Polbooks and GrQc, EQ is close to 0.5. This shows that the proposed algorithm

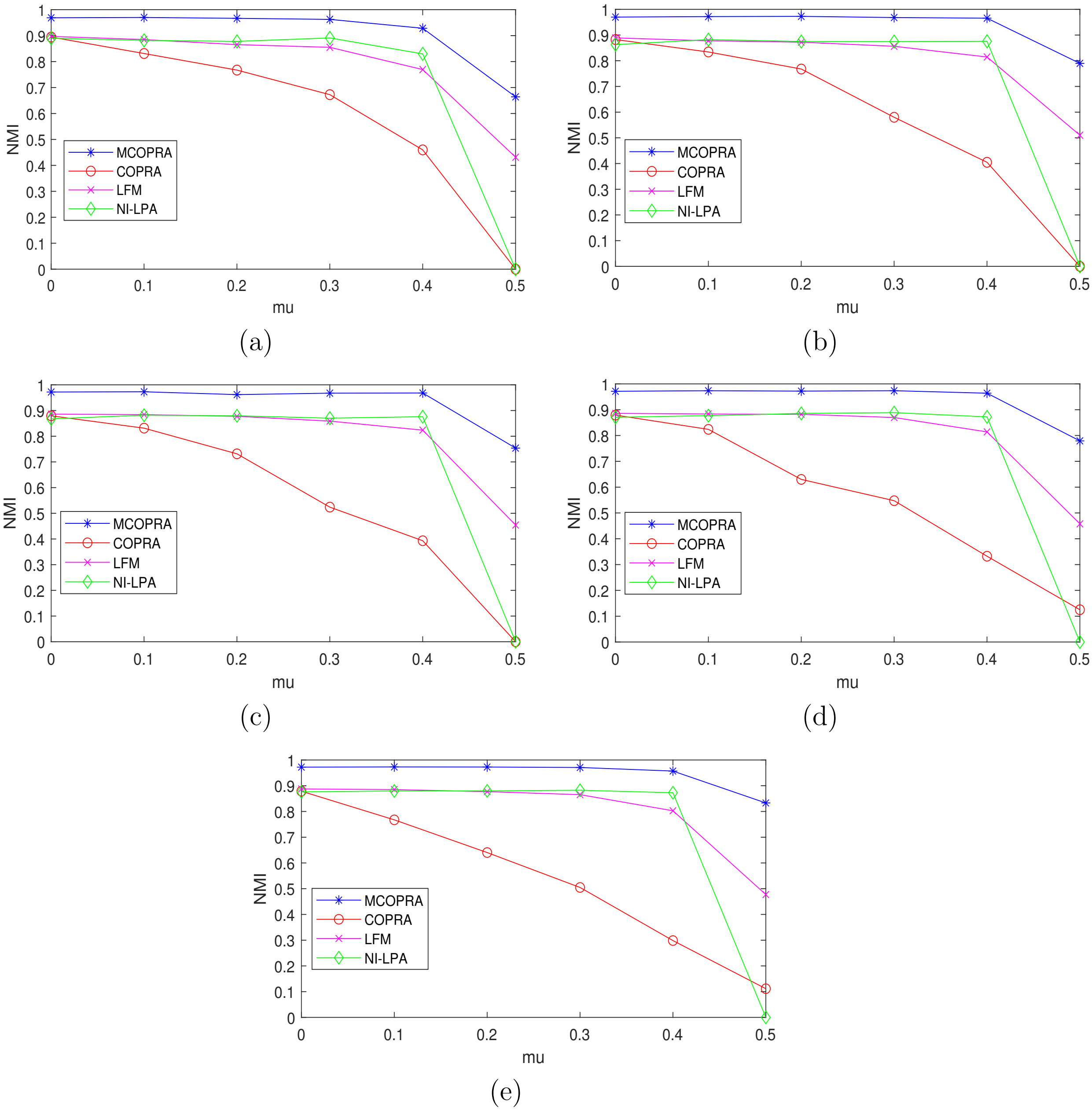

FIGURE 4. Experimental results NMI on non-overlapping communities in synthetic network. (a), (b), (c), (d), (e), representing network LFR2, LFR4, LFR6, LFR8, LFR10.

can detect overlapping communities in real networks. From Table 7, it can be observed that MCOPRA has advantages over the other three algorithms. The reason is that our proposed algorithm introduces higher-order structure factors in the process of detecting overlapping communities.

(4) The effect of the value of v on the modularity. In order to understand how v affects the quality of community classification, we measured the EQ value obtained by the algorithm on the real network TerroristRel. When $v = 1$, the value of EQ obtained by the algorithm is 0.65, which is better than $v = 2$. Between $v = 3$ and 8, the EQ value reaches a better situation, reaching the maximum value of 0.87 when $v = 7$. Starting from $v = 9$, the EQ value drops to about 0.3, until $v = 20$, the value fluctuates slightly between 0.3 and 0.4. Then the EQ value began to decline, and stabilized at 0.2 with the

Table 7. Results of modularity EQ running in real-world networks

Datasets	MCOPRA	LFM	COPRA	NI-LPA
Karate	**0.3718**(v=2)	0.2863	0.1358(v=2)	0.3703
Dolphin	0.3784(v=3)	**0.3984**	0.1119(v=3)	0.0403
Polbooks	0.4556(v=4)	0.4360	**0.5527**(v=4)	0.2227
Jazz	**0.3484**(v=3)	0.1816	0.2938(v=3)	-
Lesmis	**0.2888**(v=3)	0.2594	0.2415(v=3)	0.0756
Football	0.5715(v=6)	**0.6386**	0.2472(v=6)	-
Email-univ	**0.5109**(v=2)	-	0.4578(v=2)	0.0302
Polblogs	0.3770(v=2)	0.3684	**0.3788**(v=2)	-
ca-GrQc	**0.4334**(v=6)	0.0911	0.4282(v=6)	0.4276
TerroristRel	**0.8527**(v=4)	0.3127	0.5637(v=4)	0.8423

increase of v. So we know that v has a big effect on EQ. The result is also show in Figure 5.

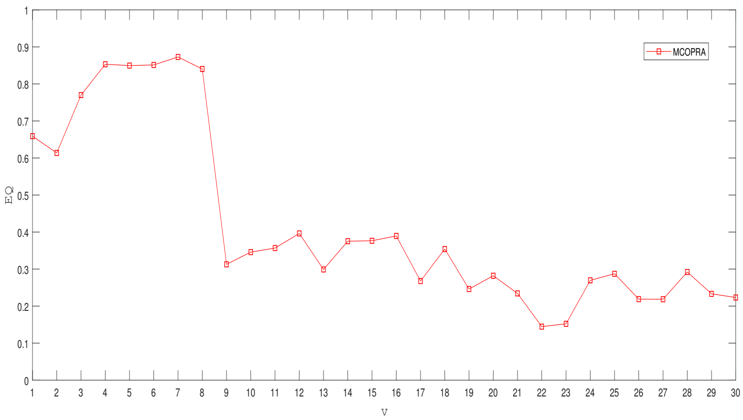

Figure 5. EQ of MCOPRA on TerroristRel network.

5. **Conclusions.** In this paper, we propose a novel method, the motif-aware Label Propagation Algorithm (MCOPRA), which is based on label propagation techniques to detect overlapping communities in complex networks. MCOPRA retains the good propagation method of COPRA, and uses the higher-order structural features of the network to improve the label selection process. Therefore, our algorithm uses a new filtering method to remove unnecessary labels. The proposed method improves the accuracy and robustness of community detection. We have done a lot of experiments on real and artificial networks, and the quality of MCOPRA for detecting overlapping communities has been proved in experiments. We found that when it comes to networks with complex node relationships, our algorithm can find the community closest to the correct network and maintain high stability. On the real network, our method can reveal the real community distribution. In the future, we can improve the MCOPRA model to discover communities in the context of dynamic networks.

Acknowledgment. This work was supported in part by the National Social Science Fund of China (No. 21BTJ011), the Natural Science Foundation of Fujian Province, China (No. 2021J011187), the Joint Funds of 5th Round of Health and Education Research Program of Fujian Province (No. 2019-WJ-41), the Fundamental Research Funds for the Central Universities (Project No.2072021052), the Science and Technology Planning Project of

Fujian Province (No. 2020H0023). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] A. Todeschini, X. Miscouridou, F. Caron, Exchangeable random measures for sparse and modular graphs with overlapping communities, *Journal of the Royal Statal Society Series B*, 2020, https://doi.org/10.1111/rssb.12363.

[2] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821-7826, 2002.

[3] E. Abbe, Community detection and stochastic block models: Recent developments, *Journal of Machine Learning Research*, vol. 18, no. 177, pp. 1-86, 2018.

[4] T. Nepusz, H. Yu, A. Paccanaro, Detecting overlapping protein complexes in protein-protein interaction networks, *Nature Methods*, vol. 9, no. 5, pp. 471-472, 2012.

[5] Y. Dourisboure, F. Geraci, M. Pellegrini, Extraction and classification of dense implicit communities in the web graph, *ACM Transactions on the Web*, vol. 3, no.2, pp. 1-36, 2009.

[6] V. D. Blondel, J. L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, 10008, 2008.

[7] E. M. Airoldi, D. M. Blei, S. E. Fienberg, E. P. Xing, Mixed membership stochastic blockmodels, *2008 Neural Information Processing Systems*, pp. 33-40, 2009.

[8] D. Imre; P. Gergely, V. Tamás, Clique percolation in random networks, *Physical Review Letters*, vol. 94, 160220, 2005.

[9] B. Yang, J. Di, J. Liu, D. Liu, Hierarchical community detection with applications to real-world network analysis, *Data & Knowledge Engineering*, vol. 83, pp. 20-38, 2013.

[10] D. Gómez, E. Zarrazola, J. Yáñez, J. Montero, A divide-and-linkalgorithm for hierarchical clustering in networks, *Information Science*, vol. 316, no C, pp. 308-328, 2015.

[11] U. V. Luxburg, A tutorial on spectral clustering, *Statistics and Computing*, vol. 17, no. 4, pp. 395-416, 2007.

[12] D. P. T. Kim, Improve the Accuracy of Link Predictions on Sparse Networks Based on Similarity Measures and Multiple Ensemble Learning, *Journal of Information Hiding and Multimedia Signal Processing*, vol 11, no. 4, pp. 151-161, 2020.

[13] F. Wang, T. Li, X. Wang, S. Zhu, C. Ding, Community discovery using nonnegative matrix factorization, *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 493-521, 2011.

[14] C. Pizzuti, Overlapped community detection in complex networks, *Genetic and evolutionary computation*, pp. 859-866, 2009.

[15] S. Gregory, An Algorithm to Find Overlapping Community Structure in Networks, *European Conference on Principle & Practice of Knowledge Discovery in Databases*, pp. 91-102, 2007.

[16] J. S. Wu, L. C. Jiao, C. Jin, F. Liu, M. G. Gong, R. H. Shang, W. S. Chen, Overlapping community detection via network dynamics, *Physical Review E: Statistical, Nonlinear & Soft Matter Physics*, vol. 85, 16115, 2012.

[17] S. Boudebza, R. Cazabet, F. Azouaou, O. Nouali, OLCPM: An online framework for detecting overlapping communities in dynamic social networks, *Computer Communications*, vol. 123, pp. 36-51, 2018.

[18] P. Z. Li, L. Huang, C. D. Wang, J. H. Lai, D. Huang, Community Detection by Motif-Aware Label Propagation, *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 2, pp. 1-19, 2020.

[19] A. R. Benson, D. F. Gleich, J. Leskovec, Higher-order organization of complex networks, *Science*, vol. 353, no. 6295, pp. 163-166, 2016.

[20] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network Motifs: Simple Building Blocks of Complex Networks, *Science*, vol. 298, no. 5594, pp. 824-827, 2002.

[21] Y. Hao, A. R. Benson, J. Leskovec, D. F. Gleich, Local Higher-Order Graph Clustering, *Knowledge Discovery and Data Mining*, pp. 555-564, 2017.

[22] U. N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, *Physical Review E: Statistical, Nonlinear & Soft Matter Physics*, vol. 76, 036106, 2007.

[23] Y. Zhang, Y. G. Liu, J. K. Zhu, C. H. Yang, W. Yang, S. Q. Zhai, NALPA: a node Ability Based Label Propagation Algorithm for Community, *IEEE Access*, vol. 8, pp. 46642-46664, 2020.

[24] X. Liu, J. S. Firoz, M. Zalewski, M. Halappanavar, A.H. Gebremedhin, Distributed Direction-Optimizing Label Propagation for Community Detection, *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1-6, 2019.

[25] Z. H. Wu, Y. F. Lin, S. Gregory, H. Y. Wan, S. F. Tian, Balanced Multi-Label Propagation for Overlapping Community Detection in Social Networks, *Journal of Computer Science & Technology*, vol. 37, no. 3, pp. 468-479, 2012.

[26] S. H Li, H. Lou, W. Jiang, J. H Tang, Detecting community structure via synchronous label propagation, *Neurocomputing*, vol. 151, pp. 1063-1075, 2015.

[27] B. Y. Xu, Z. Liang, Y. Jia, B. Zhou, Community Detection Based on Robust Label Propagation Algorithm, *Intternational Conference on Trustworthy Computing and Services*, pp. 303-310, 2013.

[28] H. Wang, K. H. Kang, Y. F. Wang, Node Confidence Calculation Method Based on D-separation of theorem Bayesian Network, *Computer Applications and Software*, vol. 33, no. 11, pp. 288-296, 2020.

[29] L. L. Kang, R. S. Chen, N. X. Xiong, Y. C. Chen, Y.X. Hu, C. M. Chen, Selecting Hyper-Parameters of Gaussian Process Regression Based on Non-Inertial Particle Swarm Optimization in Internet of Things, *IEEE Access*, vol. 7, pp. 59504-59513, 2019.

[30] Q. Y. Yang, S. C. Chu, J. S. Pan, C. M. Chen, C. H. Chen, Sine Cosine Algorithm with Multigroup and Multistrategy for Solving CVRP, *Mathematical Problems in Engineering*, vol. 2020, 8184254, 2020.

[31] R. Dong, J. J. Yang, Y. G. Chen, Overlapping Community Detection in Weighted Temporal Text Networks, *IEEE Access*, vol. 8, pp. 58118-58129, 2020.

[32] P. Gergely, D. Imre, F. Illés, V. Tamás, Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, vol. 435, no. 7043, pp. 814-818, 2005.

[33] F. Sheikholeslami, G. B. Giannakis, Overlapping Community Detection via Constrained PARAFAC: A Divide and Conquer Approach, *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 127-136, 2017.

[34] X. Y. Zhang, C. T. Wang, Y. S. Su, L. Q. Pan, H. F. Zhang, A Fast Overlapping Community Detection Algorithm Based on Weak Cliques for Large-Scale Networks, *IEEE Transactions on Computational Social Systems*, vol. 4, no. 4, pp. 218-230, 2017.

[35] H. J. Li, J. H. Zhang, Z. P. Liu, L. N. Chen, X. S. Zhang, Identifying overlapping communities in social networks using multi-scale local information expansion, *European Physical Journal B*, vol. 85, no. 6, p. 1-9, 2012.

[36] S. Gregory, Finding overlapping communities in networks by label propagation, *New Journal of Physics*, vol. 12, 103018, 2009.

[37] M. Coscia, G. Rossetti, F. Giannotti, D. Pedreschi, DEMON: a Local-First Discovery Method for Overlapping Communities, *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 603-611, 2012.

[38] J. R. Xie, B. K. Szymanski, X. M. Liu, SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process, *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 344-349, 2011.

[39] I. B. E. Kouni, W. Karoui, L. B. Romdhane, Node Importance based Label Propagation Algorithm for overlapping community detection in networks, *Expert Systems With Applications*, vol. 162, 113020, 2020.

[40] L. Huang, C. D. Wang, H. Y. Chao, Higher-order multi-layer community detection, *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 9945-9946, 2019.

[41] A. R. Benson, D. F. Gleich, J. Leskovec, Higher-order organization of complex networks, *Science*, vol. 353, no. 6295, pp. 163-166, 2016.

[42] A. Arenas, A. Fernández, S. Fortunato, S. Gómez, Motif-based communities in complex network, *Journal of Physics A: Mathematical and Theoretical*, vol. 41, 224001, 2008.

[43] A. R. Benson, D. F. Gleich, J. Leskovec, Tensor spectral clustering for partitioning higher-order network structures, *Proceedings of the Siam International Conference on Data Mining. Siam International Conference on Data Mining*, vol. 2015, pp. 118-126, 2015.

[44] P. Z. Li, L. Huang, C. D. Wang, J. H. Lai, EdMot: An Edge Enhancement Approach for Motif-aware Community Detection, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, pp. 479-487, 2019.

[45] P. Clara, S. Annalisa, Motif-Based Community Detection in Multiplex Networks, *International Workshop on Complex Networks and their Applications. Springer*, pp. 190-201, 2017.

[46] A. R. Benson, D. F. Gleich, J. Leskovec, Higher-order organization of complex networks, *Science*, vol. 353, no. 6295, pp. 163-166, 2016.

[47] W. Sebastian, R. Florian, FANMOD: a tool for fast network motif detection, *Bioinformatics*, vol. 22, no. 9, pp. 1152-1153, 2006.

[48] K. Zahra, A. Hayedeh, E. Elahe, N.D. Abbas, A. Elnaz, A. Sahar, M. Shahin, S. Falk, M.N. Ali, Kavosh: A new algorithm for finding network motifs, *BMC Bioinformatics*, vol. 10, no. 1, 318, 2009.

[49] A. Lancichinetti, S. Fortunato, F. Radicchi, Benchmark graphs for test community detection algorithms, *Physical review. E*, vol. 78, no. 4, 046110, 2008.

[50] W. W. Zachary, An Information Flow Model for Conflict and Fission in Small Groups, *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452-473, 1977.

[51] L. David, The emergent properties of a dolphin social network, *Proceedings of the Royal Society B: Biological Sciences*, vol. 270, no. 2, pp. S186-S188, 2003.

[52] M. Girvan, M. E. J Newman, Community structure in social and biological networks, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821-7826, 2002.

[53] A. Lancichinetti, S. Fortunato, J. Kertesz, Detecting the overlapping and hierarchical community structure of complex networks, *New Journal of Physics*, vol. 11, 033015, 2009.

[54] M. Girvan, M. E. J Newman, Finding and evaluating community structure in Networks, *Physical Review*, vol. 69, 026113, 2004.

[55] H. W. Shen, X. Q. Cheng, K. Cai, M. B. Hu, Detect overlapping and hierarchical community structure in networks, *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706-1712, 2009.