Computer Science and Information Systems 13(2):515–530 DOI: 10.2298/CSIS160217012W

Improved Community Mining Method Based on LFM and EAGLE

Min Wang[1], Shenggang Yang[1], Liyuan Wu[2]

[1] College of Finance & Statistics, Hunan University,
410082 Hunan, China
wangmin@hnu.edu.cn
sgyang@hnu.edu.cn
[2] College of Business Administration, Huaqiao University,
362021 Fujian, China
yuanyuan_wu1979@sina.com

Abstract. Community structures are crucial topological characteristic of complex networks. Consequently, network community structure mining has great significance to the real world. Complex networks have both hierarchy and overlaps, but it is still a problem to excavate the hierarchy and overlaps of networks efficiently and accurately at the same time by algorithm. This paper proposes an efficient and functional complex network community partition algorithm by combining fitness function optimization and community similarity, which can uncover both overlapping and hierarchical community structure of complex networks. Its basic idea is to use fitness function optimization at the bottom of hierarchy division to identify efficiently and accurately the underlying community structure which is with overlaps. Hierarchical structure is based on the community similarity to merge the underlying sub-communities with the principle of maximum similarity circulation. The experimental results utilizing Karate Club Network and US college football network show that the proposed algorithm is a manageable and accurate method for not only discovering the gradation community structure, but also overlap between excavated club.

Keywords: complex network, community detection, EAGLE, LFM, overlapping & hierarchical community.

1. Introduction

Complex network is a useful tool to understand the complex system, and by analyzing networks' properties, many complex system problems could be simplified. One of the properties common to many networks is community structure, the division of network nodes into groups within which the network connections are dense, with a lower density of edges between groups. The ability to find and analyze such groups can provide invaluable help in understanding and visualizing the structure of networks. The study of community structure in networks has been a focus in the effort of understanding complex system. The method proposed in this paper employs the local optimization of the fitness function, and then combines the two most similar communities based on the idea of the recycling community similarity. This method is suitable for the detection of hierarchical structure and overlapping communities simultaneously, therefore to improve the accuracy of community detecting and make the results closely representing the real world.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 outlines the algorithm. In section 4, the experimental results are reported. And the final section offers concluding remarks and sheds light on future research directions.

2. Community Detection in Complex Networks

2.1. Complex Network and Community

The complex networks' structure is irregular, complex and dynamically evolving in time, which is made up the collection of nodes and edges. In complex network, the community structure is the representation of subnet, which is groups of nodes that have a high density of edges within them, but between which there are only a lower density of external edges [1-2]. Uncovering communities in complex networks has attracted much attention because of the importance of the problem for many scientific areas [3-8]. The tightness of a community can be measured by subnet aggregation degree. According to different tightness, we could find different community structures. And even the nodes in the same community, the connection density between them are different. In the ideal case, the community is a fully connected network and contains at least three nodes.

Many of community detection algorithms are described in the survey paper [4-5, 9]. Girvan and Newman suggested an algorithm, called GN algorithm, which detects communities by progressively removing edges from the original network and the connected components of the remaining network are the communities [10]. After that, a lot of variations and extensions of the GN algorithm have been developed [11-23]. Some of them are employed with varying levels of success on speeding up the calculation speed over the original algorithm by focus on improving measurement of edges, such as the Monte Carlo resampled variation proposed by Tyler et al. [21], the algorithm based on counts of short cycles proposed by Radicchi et al. [22].

After years of works concerning module degree, the concept of hierarchical structure and overlapping communities of complex network are put forward. Many studies indicate that communities in real world network are simultaneously hierarchical and overlapped [18, 24]. Networks often show a hierarchical organization, with communities containing sub-communities, give a simple example: a few people form a family, thousands of families into a city, and then a number of cities will form a country; moreover, one node can participate in more than one community, resulting in overlapping communities naturally. For instance, people belong to different social communities, depending on their families, friends, hobbies, etc. Figure 1 shows the overlay structure in the network. However, most previous algorithms aim at detecting standard partitions, i.e. partitions within which each node is assigned to a single community; let alone few methods are capable of simultaneously detecting the overlapping and hierarchical community structure in networks. For this reason, the topic of deriving methods capable to detect both hierarchies and overlapping communities has recently attracted a large interest in the scientific community [25-31].

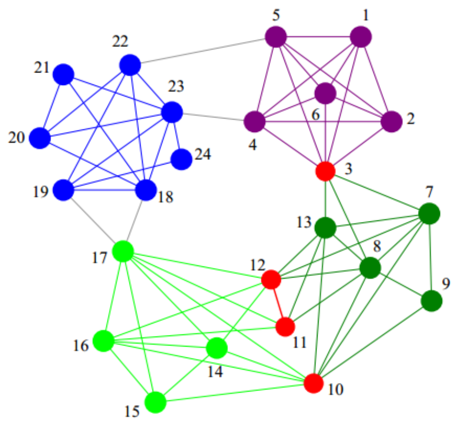

Fig 1. A network with overlapping community structure is shown. The nodes of different colors represent different communities. Red nodes represent the overlap between the two communities. Edges between communities are gray [32]

2.2. Local Optimization Method Based on the Fitness Function of LFM

LFM is a method based on local optimization of a fitness function with capabilities of calculating quickly and analyzing mass networks containing millions of nodes [25]. This algorithm is one of the typical community structure classification algorithms based on the ides of detecting communities by a local optimization. In LFM algorithm, the division of community is identified by the maximization of the fitness function f_G, and f_G is defined as

$$f_G = \frac{k_{in}^G}{(k_{in}^G + k_{out}^G)^\alpha} \quad \textbf{(1)}$$

where k_{in}^G (k_{out}^G) is the weighted sum of internal (external) degrees of the nodes of community G , and α is a parameter controlling the size of the communities, also known as resolution parameter. Given a fitness function, the fitness of a node i with respect to a community G is defined as the variation of the fitness of the community G with and without node i, i.e.

$$f_G^i = f_{G+i} - f_{G-i} \quad \textbf{(2)}$$

In equation (2), f_{G+i} (f_{G-i}) is the value of fitness function of community $\{G + i\}(\{G - i\})$, indicating the fitness of the community with node i included in or removed from community G. Where $f_G^i > 0$, means the value of fitness function increased with the node i joining community G, and thus node i should be included in community G; conversely, where $f_G^i < 0$, means the point i should be removed from G.

LFM consists of the following steps:

(1) Initialization, pick an node at random as the original member of community G, $k_{in}^G = 0$;

(2) Computing the fitness of all neighboring nodes to community G according to the formula (2);
(3) the neighbor with the largest fitness is added to G, yielding a new community G';
(4) the fitness of each node of G' is recalculated;
(5) if a node turns out to have negative fitness, it is removed from G', yield a new new community G'';
(6) If step (5) occurs, repeat from (4), otherwise repeat from step (2).

The process stops when the nodes examined in step (2) all have negative fitness, and the maxima of the fitness function is found. Then the algorithm continues to select other isolated nodes, repeat the process until all nodes in the network have been assigned to at least one community. The nodes of every community may either overlap with other communities or not.

The parameter α of the formula (1) reflects the hierarchy. Small α yield large communities, large α instead deliver small communities. By varying the α -value, we can explore the whole hierarchy of community structure. And the computational complexity of the algorithm mainly depends on the size of the community and the extent of their overlaps, which in turn strongly lie on the specific network being studied along with the value of the parameter α. Although the LFM need a lot of experiments and training to get the appropriate parameter α to breaks the dendrogram into communities, it is hopeful to find out α eventually.

2.3. EAGLE Algorithm Based on Condensation

The algorithm EAGLE is proposed as an agglomerative hierarchical clustering algorithm to investigate simultaneously the hierarchical and overlapping community structure of networks [26]. Different from traditional agglomerative algorithms, the algorithm deals with the set of maximal cliques rather than the set of sole nodes. A maximal clique is a clique that cannot be extended by including one more adjacent node, meaning it is not a subset of any other cliques. Algorithm EAGLE chooses Bron-Kerbosch algorithm [33] to find out all the maximal cliques in the network.

EAGLE algorithm has two steps: firstly, a dendrogram is generated; secondly, choose a place to cut the dendrogram and get corresponding communities. In order to determine the place of the cut, an extension of modularity EQ is proposed to judge the quality of a cover, which is defined as

$$EQ = \frac{1}{2m}\sum_{i}\sum_{v\in C,w\in C}\frac{1}{O_v O_w}[A_{vw} - \frac{k_v k_w}{2m}] \quad \textbf{(3)}$$

Where O_v is the number of communities to which node v belongs. The place of the cut chosen is with the maximum value of EQ. It is found that a high value of EQ indicates a significant overlapping community structure. By further applying EAGLE to each community found until none of them can be divided into smaller ones, we could obtain a hierarchy of overlapping communities. Let n be the number of nodes in the network, s be the number of maximal cliques in the initial state of the algorithm, and$h$ be the number of pair of maximal cliques which are neighbors. The first step of the algorithm takes at most $O(n^2 + (h + s)s)$ operations and the second stage of the algorithm takes at most $O(n^2 s)$ operations. In the first step, finding out all the maximal cliques is a NP problem; however, it is easy due to the sparseness of these networks.

3. Improved Community Partition Method Based on LFM and EAGLE

3.1. Theoretical Algorithm Model

Communities in real world social networks have overlapping and hierarchical organizations. According to a highly relevant community detection study, a common framework that can extract overlapping and hierarchical organizations is characterized as follows.

(1) The first need to extract all meaningful units, such as maximal sub-graphs [1], maximal cliques [3] or even a single node from the original social network, and then view each meaningful unit as the initial community.

(2) Select a specific similarity function, repeated calculation the similarity of two adjacent communities, and merged with the maximum similarity of the club, until only a community still exists, the whole process can form a dendrogram.

(3) Through the dendrogram of step (2) to get the different levels of overlapping community structure, and then through a quality index function to assess the merits of each level division. Finally, the level of local or global optimal solution of quality index function is selected as the final result.

Without loss of generality, the community partition algorithm we proposed in this paper is similar to the common framework mentioned above. Our new method is the combination of LFM and EAGLE, which applies the idea of LFM in classification of base layers and adopts the agglomerative hierarchical clustering of EAGLE. We firstly deal with the base layer of communities, basing on fitness function optimization. The base layer of communities actually reflects the overlapping between communities. Then on the basis of the division of base layer, these communities are amalgamated according to the order determined by the community similarity.

Our algorithm' steps can be summarized as follows:

(1) Pick a node at random as the initial communities G, and the local optimization of the fitness function f_G is performed to obtain the first level of communities

$$f_G = \frac{k_{in}^G}{k_{in}^G + k_{out}^G}[\frac{1}{2m}\left(k_{in}^G - \frac{k_i k_{out}^G}{2m}\right)] \quad (4)$$

where $\mathrm{k}_{\mathrm{in}}^{\mathrm{G}}\left(\mathrm{k}_{\mathrm{out}}^{\mathrm{G}}\right)$ is the total internal (external) degrees of the nodes of community G; m is the total number of degrees in network.

(2) Calculate the fitness function f_G^i of neighbor node i;

(3) Select the neighbors with the maximum and positive fitness, and incorporate them into G to form a new community G'.

(4) Recalculate the fitness f_G^i of each node in community G'. If the nodes are with negative fitness, then remove them from community G'.

(5) Repeat step (3) through (4) until the fitness values of all nodes in community are positive.
(6) Return to step (2), calculate the fitness of neighbor nodes of the latest community.
(7) The process continues until all neighbors' fitness values are negative. And at this time, we get the maximal fitness value of community G. This means the first community in the first layer of community structure is determined.
(8) Reselect isolated nodes randomly from the remained and continue above procedures until all nodes at least assigned to one community. Then the first layer structure is generated, and the overlapping nodes are recorded.
(9) The hierarchical structure of dividing community base on similarity function of community, the similarity function proposed in this paper is an improved form of the EAGLE algorithm modularity (equation 3). It is suitable for the community similarity with overlapping network computing. The similarity function of community we propose is as follows:

$$M = \frac{1}{2m}\sum_{\substack{V\in C1, W\in C2\\ V\neq W}}[A_{VW} - \frac{k_V k_W}{2m}] + \frac{1}{2m}\sum_{V\in C1, V\in C2}\left[1 - \frac{k_v^2}{2m}\right] \quad \textbf{(5)}$$

A_{VW} is the number of edges between node V and node W, and K_V and K_W are the degrees of node V and node W respectively. We calculate the similarity of two communities in turn; then a pair of communities with the largest value of similarity is merged into a new one.

After above merger, we further recalculate the similarity of a pair of community. The combining process for those pairs of communities with maximum similarity stops when all nodes are merged into a single community. This means the hierarchical structure is uncovered.

3.2. Performance indicators

The process of our algorithm corresponds to a dendrogram, basing on the agglomerative progression. The next task is to evaluate the goodness of overlapped community decomposition. The most widely applied measurement for community metrics is modularity, which first put forward by Newman and Girvan [10]. In this paper, we apply the extension of modularity suit for overlapping community structure to judge the strength of community's characteristic, which is defined as [26]:

$$Q = \frac{1}{2m}\sum_{ij}\frac{1}{O_i O_j}(A_{ij} - \frac{k_i k_j}{2m})\delta(C_i, C_j) \quad \textbf{(6)}$$

where O_i is the number of communities that node i belongs to, A_{ij} is the number of edges between node i and j, k_i is the degree of node i, m is the total number of edges in the network. $\delta(C_i, C_j) = 1$ if i and j belong to the same community, and $\delta(C_i, C_j) = 0$ otherwise.

4. Empirical Results and Conclusions

In this section, we validate our method by applying it to two real-world complex networks in social fields, the karate club network and network of American college football teams. The results show that this improved algorithm can provides a possible way to investigate a more complete picture of the community structure.

4.1. Karate Club Network

Karate Club Network is observed by Wayne Zachary over the course of two years in the early 1970s [34], which is a well-known benchmark regularly used to test community detection algorithms. The network consists of 34 nodes and 78 undirected edges. Each node represents a member of karate club, and each edge represents friendship between members. A consensus network structure extracted from Zachary's observations before the split is shown in Fig. 2.

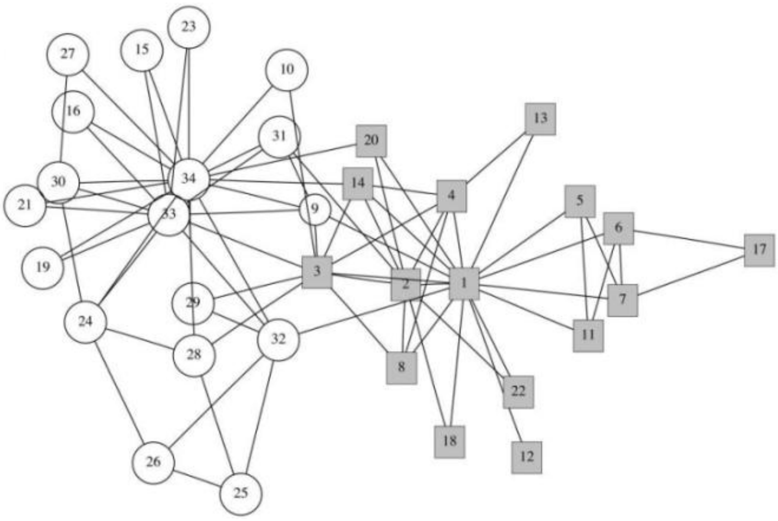

Fig. 2. Network of friendships between individuals in the karate club study of Zachary [34]

Table 1 demonstrates the results for karate club network found with our algorithm. By feeding this network into our algorithm, the overlapping nodes are uncovered even in the bottom lay; consequently, it seems more appropriate to illustrate the process of community detection of our algorithm by table rather than figure. In Table 1, we show the entire procession of agglomerative clustering. The community overlapping is found in each layer, and some nodes in the network are assigned to two or more communities. Also it is important to find the algorithm gives the highest value for the modularity when the network is split into two communities. Fig. 3 shows the modularity change of karate club network partition of our method. Accordingly, we select the layer with the maximal modularity as our output result, corresponding to two communities. The comparison between final partition results of our algorithm and the original partition of karate club network is shown in Table 2. The nodes contained in the same brace are the ones within a same community, and those bold Italic underline are overlapping vertices between communities.

Table 1. The karate club network partitioning process of our algorithm. The nodes in the same brace are those contained within a same community. NM – number of merging, CM – NO. of community mergered, S – similarity of mergered community, Q – modularity, N – number of communities

NM	CM	S	Q	N	Community node details
0		\	0.2383	13	{3,14,28,29,32,33,34}{2,14,18,20,31,33,34}{1,3,5,6,7,9,11,14,17,20,28,29,31,32,33,34}{1,2,3,4,8,9,12,13,14,18,20,22,28,29,31,32,33,34}{3,10,28,29,32,33,34}{19,33,34}{15,33,34}{27,30,34}{23,33,34}{24,26,30,33,34}{25,26,32}{16,33,34}{21,33,34}
1	3,4	0.2014	0.2913	12	{3,14,28,29,32,33,34}{2,14,18,20,31,33,34}{1,2,3,4,5,6,7,8,9,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{3,10,28,29,32,33,34}{19,33,34}{15,33,34}{27,30,34}{23,33,34}{24,26,30,33,34}{25,26,32}{16,33,34}{21,33,34}
2	2,3	0.1252	0.3202	11	{3,14,28,29,32,33,34}{1,2,3,4,5,6,7,8,9,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{3,10,28,29,32,33,34}{19,33,34}{15,33,34}{27,30,34}{23,33,34}{24,26,30,33,34}{25,26,32}{16,33,34}{21,33,34}
3	3	0.0765	0.3317	10	{1,2,3,4,5,6,7,8,9,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{3,10,28,29,32,33,34}{19,33,34}{15,33,34}{27,30,34}{23,33,34}{24,26,30,33,34}{25,26,32}{16,33,34}{21,33,34}
4	1,2	0.0545	0.3428	9	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{19,33,34}{15,33,34}{27,30,34}{23,33,34}{24,26,30,33,34}{25,26,32}{16,33,34}{21,33,34}
5	4,6	0.0223	0.3683	8	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{19,33,34}{15,33,34}{24,26,27,30,33,34}{23,33,34}{25,26,32}{16,33,34}{21,33,34}
6	2,4	0.0215	0.3716	7	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{19,24,26,27,30,33,34}{15,33,34}{23,33,34}{25,26,32}{16,33,34}{21,33,34}
7	2,3	0.0273	0.3773	6	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{15,19,24,26,27,30,33,34}{23,33,34}{25,26,32}{16,33,34}{21,33,34}
8	2,3	0.0331	0.3818	5	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{15,19,23,24,26,27,30,33,34}{25,26,32}{16,33,34}{21,33,34}
9	2,4	0.0388	0.3871	4	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{15,16,19,23,24,26,27,30,33,34}{25,26,32}{21,33,34}
10	2,4	0.0446	0.3929	3	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{15,16,19,21,23,24,26,27,30,33,34}{25,26,32}
11	2,3	0.0335	0.4095	2	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,32,33,34}{15,16,19,21,23,24,25,26,27,30,32,33,34}
12	1,2	0.0196	\	1	{ All nodes in network }

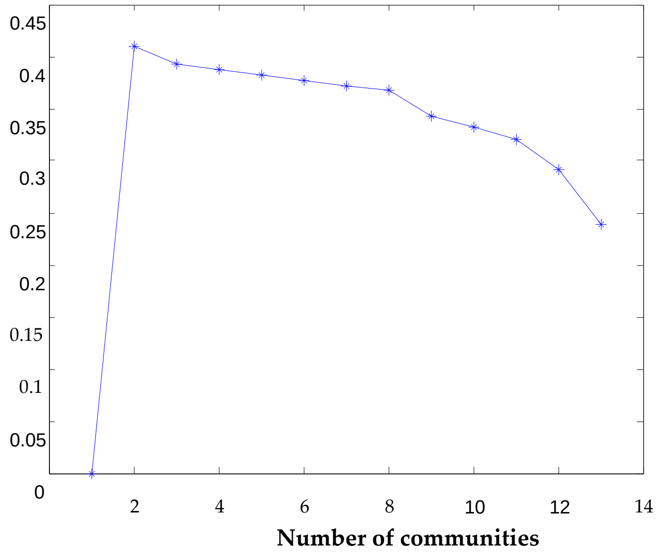

Fig. 3. Modularity change of karate club network partitioning by applying our algorithm

Table 2. Comparison between final partition results of our algorithm and the original partition of karate club network [34]

Karate	Number of communities	Community node details
Original Partition	2	{1,2,3,4,5,6,7,8,11,12,13,14,17,18,22} {9,10,15,16,19,20,21,23,24,25,26,27,28,29,30,31,32,33,34}
Our Method Partition	2	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,20,22,28,29,31,***32***,***33***,***34***} {15,16,19,21,23,24,25,26,27,30,***32***,***33***,***34***}

4.2. The Bottlenose Dolphin Network

Bottlenose dolphin network [35] is also a commonly used benchmark network to test the communities mining algorithm. This benchmark network represents the associations between 62 dolphins living in Doubtful Sound, New Zealand. Links between dolphins represent the statistically significant frequent associations between them, that is, an unweighted and undirected network with 62 nodes and 159 links. Dolphin population contains two sub populations, of which the larger group contains 42 dolphins, while the smaller group of only 20 dolphins. The original network community is shown in Fig. 4. In order to facilitate the display of community mining results, give unique number to each dolphin in the benchmark network, the number of results as shown in Table 3.

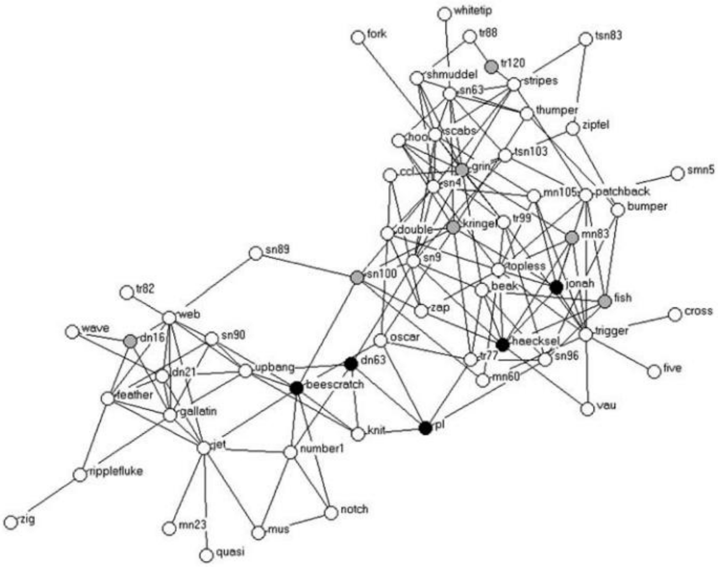

Fig. 4. Social network of bottlenose dolphins in Doubtful Sound, New Zealand; Side flops (SF) and upside-down lob tails (ULT) are rare events; Dolphins observed SF are in black and the ones observed ULT are in grey [36]

Table 3. Corresponding number of each dolphin in the bottlenose dolphin network

number	name	number	name	number	name	number	name	number	name
1	Beak	14	Gallatin	27	Notch	40	SN89	53	TSN103
2	Beescratch	15	Grin	28	Number1	41	SN9	54	TSN83
3	Bumper	16	Haecksel	29	Oscar	42	SN90	55	Upbang
4	CCL	17	Hook	30	Patchback	43	SN96	56	Vau
5	Cross	18	Jet	31	PL	44	Stripes	57	Wave
6	DN16	19	Jonah	32	Quasi	45	Thumper	58	Web
7	DN21	20	Knit	33	Ripplefluke	46	Topless	59	Whitetip
8	DN63	21	Kringel	34	Scabs	47	TR120	60	Zap
9	Double	22	MN105	35	Shmuddel	48	TR77	61	Zig
10	Feather	23	MN23	36	SMN5	49	TR82	62	Zipfel
11	Fish	24	MN60	37	SN100	50	TR88		
12	Five	25	MN83	38	SN4	51	TR99		
13	Fork	26	Mus	39	SN63	52	Trigger		

Table 5 shows the results of our algorithm on dolphin benchmark network and compared the results of the original network partitioning. We can see from the table that the results obtained by using our method are basically consistent with the original classification. What's more we found three overlapping nodes between the two communities. The overlapping nodes are marked in bold italic underline in the table. Because there are many nodes in the dolphin network, it is not convenient to list the whole hierarchy process. The whole partition method is consistent with the method of the first benchmark network (karate club network). Similarity of communities of the whole partition process and the corresponding changes in the modularity are shown in Table 4. The whole hierarchy process of the module changes as shown in Fig. 5. From Fig. 5 and Table 4 it is easy to draw the conclusion that when the whole network is divided into two groups will reach the maximum degree of modularity. Therefore, we choose the global optimum of the modules corresponding partition results of as the final optimal results.

Table 4. Bottlenose dolphin network partitioning process of algorithm we proposed. NM – number of merging, S – similarity of mergered community, Q – modularity, N – number of communities

NM	S	Q	N	NM	S	Q	N
0	\	0.22204	10	5	0.012346	0.30465	5
1	0.041234	0.27457	9	6	0.010235	0.35894	4
2	0.033454	0.28568	8	7	0.008346	0.42567	3
3	0.024223	0.29455	7	8	0.004235	0.46986	2
4	0.018543	0.28458	6	9	0.003247	\	1

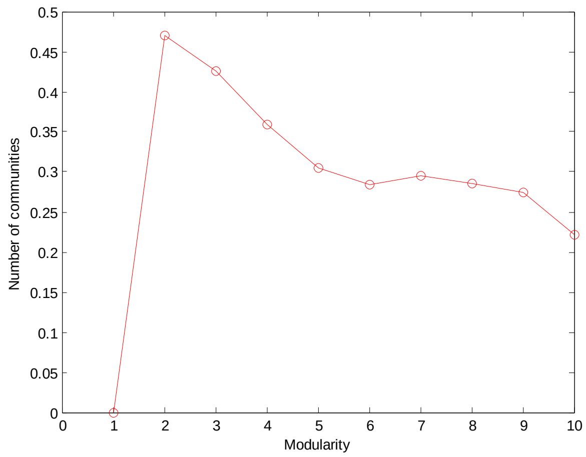

Fig. 5. Modularity change of bottlenose dolphin network partitioning by applying our algorithm

Table 5. Comparison between final partition results of our algorithm and the original partition of bottlenose dolphin network

Dolphins	Number of communities	Community node details
Original Partition	2	{2,6,7,8,10,14,18,20,23,26,27,28,32,33,42,49,55,57,58,61} {1,3,4,5,9,11,12,13,15,16,17,19,21,22,24,25,29,30,31,34,35,36, 37,38,39,40,41,43,44,45,46,47,48,50,51,52,53,54,56,59,60,62}
Our Method Partition	2	{2,6,7,***8***,10,14,18,***20***,23,26,27,28,***30***,32,33,40,42,49,55,57,58,61} {1,3,4,5,***8***,9,11,12,13,15,16,17,19,***20***,21,22,24,25,29,***30***,31,34,35,36, 37,38,39, 41,43,44,45,46,47,48,50,51,52,53,54,56,59,60,62}

4.3. Football Network of American University

The next network that we investigated is a American college football network which represents the game schedule of the 2000 season of Division 1 of the US college football league [1]. There are 115 nodes, representing the teams, and two nodes are connected if their teams play against each other. Totally, 613 edges are in the network, representing the games played in the course of the year. The community structure of the network is shown in Fig. 6.

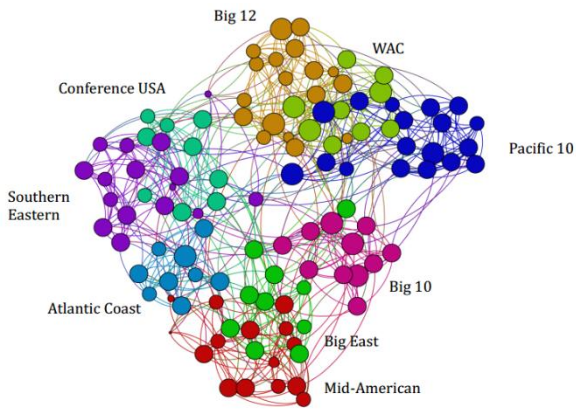

Fig. 6. The original network of American college football teams [37]

Table 6 shows the results for this network obtained with our algorithm. Due to there are too many network nodes, here we do not listed the node details of each layer during the partitioning process. The modularity change of football network is shown in Fig. 7, from which we find that when the number of communities is 12, corresponding to the maximal modularity; hence we take the layer as the final partitioning results. We could observe the different communities structure of American football network detected by Girvan and Newman [1] in Table 7 and by using our algorithm in Table 8.

Table 6. Football network partitioning process of algorithm we proposed. NM – number of merging, S – similarity of mergered community, Q – modularity, N – number of communities

NM	S	Q	N	NM	S	Q	N
0	\	0.42234	18	9	0.006757	0.53245	9
1	0.036607	0.44151	17	10	0.003605	0.51345	8
2	0.035072	0.46347	16	11	0.003566	0.49234	7
3	0.031883	0.51615	15	12	0.003350	0.46739	6
4	0.028277	0.54813	14	13	0.001797	0.45398	5
5	0.024765	0.56454	13	14	0.000237	0.46434	4
6	0.018128	0.59844	12	15	0.000232	0.44827	3
7	0.017130	0.57234	11	16	-0.0042	0.39856	2
8	0.006947	0.55245	10	17	-0.0073	\	1

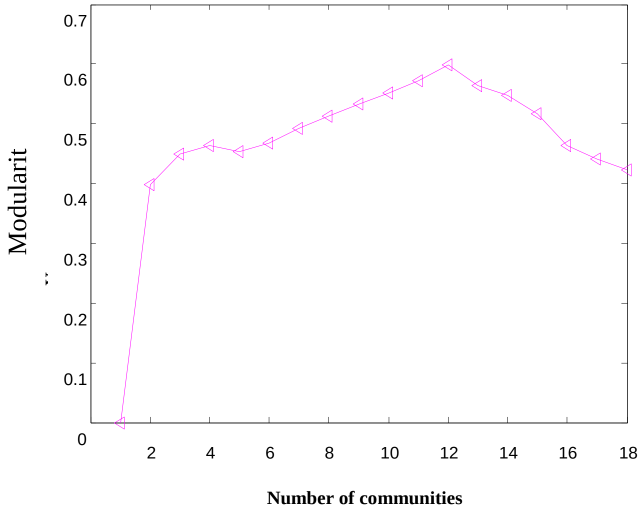

Fig. 7. Modularity change of football network partitioning by using our algorithm

Table 7. Original partitioning of football network

NO. of communities	Community node details
1	1 5 10 17 24 42 94 105
2	2 26 34 38 46 90 104 106 110
3	3 7 14 16 33 40 48 61 65 101 107
4	4 6 11 41 53 73 75 82 85 99 103 108
5	8 9 22 23 52 69 78 79 109 112
6	12 25 51 60 64 70 98
7	13 15 19 27 32 35 39 44 55 62 72 86 100
8	18 21 28 57 63 66 71 77 88 96 97 114
9	20 30 31 36 56 80 95 102
10	29 47 50 54 59 68 74 84 89 115
11	37 43 81 83 91
12	45 49 58 67 76 87 92 93 111 113

Table 8. Partitioning results of football network by applying our algorithm

NO. of communities	Community node details
1	***1*** 5 10 17 24 42 ***68*** 94 105
2	***1*** 2 26 34 38 46 ***54*** 90 104 106 110
3	3 7 14 ***16*** 33 40 48 61 65 101 107
4	4 6 11 41 53 ***59*** 73 75 82 85 99 103 108
5	8 9 22 23 52 ***69*** 78 ***79109112***
6	12 25 29 51 70 ***91***
7	13 15 ***19*** 27 32 35 ***37*** 39 ***43*** 44 55 62 72 86 100
8	18 21 28 57 63 ***64*** 66 71 ***76*** 77 88 96 97 114
9	***119*** 20 30 31 36 ***37*** 56 ***586776*** 80 87 92 93 95 102
10	47 50 ***545968*** 74 84 89 111 ***115***
11	***374359*** 60 ***64*** 81 83 ***91*** 98
12	***16*** 45 49 ***6979109112*** 113 ***115***

5. Discussion

In this paper we have presented the method, combining fitness optimization clustering algorithm and community similarity, to uncover simultaneously both the hierarchical and the overlapping community structure of complex networks. This paper formulated the fitness function and community similarity formula, and on this basis, the algorithm is capable of accurately identifying hierarchy and overlapping structure.

The applications of our algorithm to the constructed and empirical network like Zachary's karate club [33] and Bottlenose dolphin network [35] and the American university football network compiled by Girvan and Newman [1], have given excellent results closing to the benchmark in the original reference, indicating the algorithm can find meaningful partitions and its validity. However, this algorithm is just useful for the undirected and unweighted network for now. Further research of the method is to extend to the exploration of directed and weighted network.

Acknowledgments. This work was supported by the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (Grant No. 71221001), Major Program of National Social Science Foundation of China (Grant No. 12&ZD053), Social Science Foundation of Hunan Province, China (Grant No. 2012BZZ121), and Youth Fund of Humanities and Social Science Research of Ministry of Education of China (Grant No. 15YJC630126).

References

1. Girvan, M., Newman, M. E. J.: Community Structure in Social and Biological Networks. Proceedings of the National Academy of Sciences**,** Vol. 99, 7821-7826. (2002)
2. Ravasz E., Barabási A.-L. :Hierarchical Organization in Complex Network. Physical Review E, Vol. 62, 241-251. (1999)
3. Wu F., Huberman B.A.: Finding Communities in Linear Time: A Physics Approach. The European Physical Journal B – Condensed Matter and Complex Systems, Vol. 38, 331-338. (2004)
4. Boccaletti S., Latora V., Moreno Y., Chavea M., Hwang D.U.: Complex Networks: Structure and Dynamics. Physics Reports, Vol. 424, 175-308. (2006)
5. Fortunato, S.: Community Detection in Graphs. Physics Reports, Vol. 486, 75-174.(2010)
6. Borgs C., Jennifer T.C., Cohn H., Zhao Y.F.: An L^P Theory of Sparse Graph Convergence I: Limits, Sparse Random Graph Models, and Power Law Distributions. Eprint Arxiv, 1401.2906. (2014)
7. Newman M. E. J., Peixoto T. P.:Generalized Communities in Network. Physical Review Letters, Vol.115, 088701. (2015)
8. Decelle A., Krzakala F., Moore C., Zdeborová L.: Inference and Phase Transitions in the Detection of Modules in Sparse Networks. Physical Review Letters, Vol. 107, 065701. (2011)
9. Newman M. E. J.: Detecting Community Structure in Network. The European Physical Journal B - Condensed Matter, Vol. 38, 321-330. (2004)
10. Newman, M. E. J., Girvan M.: Finding and Evaluating Community Structure in Networks. Physical Review E, Vol. 69, 026113. (2004)
11. Danon L., Duch J., Diaz-Guilera A., Arenas A.: Comparing Community Structure Identification. Journal of Statistical Mechanics - Theory and Experiment, Vol. 9, P09008. (2005)

12. Leicht E. A., Newman M. E. J.: Community Structure in Directed Network. Physical Review Letters, Vol. 100, 118703. (2008)
13. Ball B., Karrer B., Newman M. E. J.: An Efficient and Principled Method for Detecting Communities in Networks. Physical Review E, Vol. 84, 036103. (2011)
14. Yan B., Gregory S.: Detecting Community Structure in Network Using Edge Prediction Methods. Journal of Statistical Mechanics - Theory and Experiment, Vol. 9, P09008. (2012)
15. Zhang X., Newman M. E. J.: Multiway Spectral Community Detection in Networks. Physical Review E, Vol. 92, 052808. (2015)
16. Clauset A., Moore C., Newman M. E. J.: Hierarchical Structure and the Prediction of Missing Links in Network. Nature, Vol. 453, 98-101. (2008)
17. Cui Y.Z., Wang X. Y., Li J. Q.: Detecting Overlapping Communities in Network Using the Maximal Sub-graph and the Clustering Coefficient. Physica A: Statistical Mechanics and its Applications, Vol. 405, 85-91. (2014)
18. Mu C. H, Liu Y., Liu Y., Wu J. S., Jiao L. C.: Two-stage Algorithm Using Influence Coefficient for Detecting the Hierarchical, Non-overlapping and Overlapping Community Structure. Physica A: Statistical Mechanics and its Applications, Vol.408, 47-61. (2014)
19. Evans T. S., Lambiotte R.: Line Graphs, Link Partitions, and Overlapping Communities. Physical Review E, Vol.80, 016105. (2009)
20. Zhang X.S., Wang R.S., Wang Y., Wang J., Qiu Y., Wang L., Chen L.: Modularity Optimization in Community Detection of Complex Networks. Europhysics Letter, Vol.87, 38002. (2009)
21. Gregory S.: An Algorithm to Find Overlapping Community Structure in Network. In Knowledge Discovery in Database: PKDD 2007; Joost N.K., Jacek K., Ramon L .D. M., Stan M., Eds.; Springer-Verlag Berlin Heidelberg: Berlin, Germany, pp.91-102. (2007)
22. Tyler J.R.; Wilkinson D.M.; Huberman B.A. Email as spectroscopy: automated discovery of community structure within organizations. Communities and Technologies, Proceedings of the First International Conference on Communities and Technologies, Amsterdam, Netherlands, 2003;Huysman M.H., Wenger E., Wulf V., Eds.; Springer -Verlag Berlin Heidelberg: Berlin, Germany, pp.81-96. (2003)
23. Radicchi F., Castellano C., Cecconi F., Loreto V., Parisi D.: Defining and Identifying Communites in Network. Proceeding of the National Academy of Sciences of the Uinted States of America, Vol.101, 2658-2663. (2004)
24. Palla G., Derenyi I., Farkas I., Vicsek T.: Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. Nature, Vol.435, 814-818.(2005)
25. Lancichinetti A., Fortunato S., Kertész J.: Detecting the Overlapping and Hierarchical Community Structure of Complex Networks. New Journal of Physics, Vol.11, 033015.(2009)
26. Shen H. W., Cheng X. Q., Cai K., Hu M. B.: Detect Overlapping and Hierarchical Community Structure in Networks. Physica A: Statistical Mechanics and its Applications, Vol. 388, 1706-1712. (2009)
27. Zhang X. W., You H. B., Zhu W.; Qiao S. J., Li J. W., Gutierrez L. A., Zhang Z., Fan X. N.: Overlapping Community Identification Approach in Online Social Networks. Physica A: Statistical Mechanics and its Applications, Vol.421, 233-248. (2015)
28. Chen D.B., Shang M. S., Lv Z. H., Fu Y.: Detecting Overloapping Communite of Weighted Networks via a Local Algorithm. Physica A: Statistical Mechanics and its Applications, Vol.389, 4177-4187. (2010)
29. Psorakis I., Roberts S., Ebden M., Sheldon B.: Overlapping Community Detection Using Bayesian Non-negative Matrix Factorization. Physical Review E, Vol.83, 066114. (2011)
30. Zhang Z.W., Wang Z.Y.: Mining Overlapping and Hierarchical Communities in Complex Networks. Physica A: Statistical Mechanics and its Applications, Vol.421, 25-33. (2015)
31. Kim P., Kim S.: Detecting Overlapping and Hierarchical Communiteis in Complex Network Using Interaction-based Edge Clustering. Physica A: Statistical Mechanics and its Applications, Vol.417, 46-56. (2015)

32. H. W. Shen, X. Q. Cheng, J. F. Guo, Quantifying and identifying the overlapping community structure in networks, J. Stat. Mech. P07042 (2009)
33. Bron C., Kerbosch J.: Algorithm 457: Finding All Cliques of An Undirected Graph. Communication of the ACM-CACM,Vol.16, 575-577. (1973)
34. Zachary W. W.: An Information Flow Model for Conflict and Fission in Small Groups. Journal of Anthropological Research, Vol.33, 452-473. (1977)
35. D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, S. M. Dawson, The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations, Behav. Ecol. Sociobiol. Vol. 54, 396–405. (2003)
36. D. Lusseau, Evidence for social role in a dolphin social network, Preprint qbio/0607048 (http://arxiv.org/abs/q-bio.PE/0607048).
37. M. Girvan and M. E. J. Newman, PNAS, Vol.99, 7821 (2002).

Min Wang is an assistant professor at College of Finance & Statistics, Hunan University, China. She received the MA degrees from Hunan University and Nottingham University in 2004 and 2005, respectively. Currently, she is a PhD student in College of Finance & Statistics, Hunan University, China. Her research interests are mainly focused on risk management, financial market and financial network.

Shenggang Yang is a professor in Hunan University, China. His research interests are international financial management and credit management, currently focusing on systemic risk, social and economic networks. He received Ph.D degree in economics from Wuhan University, China and used to be a visiting scholar in National Taiwan University during 1999 and University of Miami during 2006.

Liyuan Wu is an associate professor at College of Business Administration, Huaqiao University, China. She received her Ph.D in Economics from Hunan University in 2007. Her research interests include internet finance and public-private partnership, currently focusing on social network and financial network.

Received: February 17, 2016; Accepted: May 26, 2016.