Physica A 389 (2010) 4177–4187

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Detecting overlapping communities of weighted networks via a local algorithm

Duanbing Chen [a], Mingsheng Shang [a], Zehua Lv [b,*], Yan Fu [a]

[a] Web Sciences Center, School of Computer Science, University of Electronic Science and Technology of China, Chengdu 611731, PR China

[b] School of Computer Science, Huazhong University of Science and Technology, Wuhan 430074, PR China

ARTICLE INFO

Article history:
Received 15 January 2010
Received in revised form 20 May 2010
Available online 4 June 2010

Keywords:
Weighted networks
Overlapping community
Local algorithm
Node strength

ABSTRACT

Identification of communities is significant in understanding the structures and functions of networks. Since some nodes naturally belong to several communities, the study of overlapping communities has attracted increasing attention recently, and many algorithms have been designed to detect overlapping communities. In this paper, an overlapping communities detecting algorithm is proposed whose main strategies are finding an initial partial community from a node with maximal node strength and adding tight nodes to expand the partial community. Seven real-world complex networks and one synthetic network are used to evaluate the algorithm. Experimental results demonstrate that the algorithm proposed is efficient for detecting overlapping communities in weighted networks.

1. Introduction

Many complex systems in nature and society can be described as graphs or networks [1–3]. A community detecting problem is to detect community structures, which may overlap each other, in a weighted or unweighted complex network with n nodes and m edges. The edge weight can be set to 1 if the network is unweighted.

In recent years, the detection and analysis of community structures in complex networks has attracted a great deal of attention in many applications [4]. Community structure, whose nodes often cluster into tightly knit groups with high density of within-group edges and low density of between-group edges, has been paid much attention because of its significance in analyzing complex networks.

So far, many algorithms have been proposed to detect communities. Two classical algorithms are the spectral bisection algorithm on the basis of eigenvectors of the Laplacian matrix of a graph [5] and the Kernighan–Lin algorithm which improves on an initial division of network by optimization of the number of within- and between-community edges using a greedy strategy [6]. In recent years, various community detecting algorithms on the basis of modularity [7] have been presented. In 2004, Newman proposed a fast algorithm to detect community structures [8] based on modularity. Excellent results are obtained, especially for sparse networks, and it is typically thousands of times faster than the algorithm in [7]. Ruan and Zhang [9] proposed an efficient heuristic algorithm *QCUT*, which combines spectral graph partitioning and local searching to optimize the modularity Q. They applied *QCUT* to study a protein–protein interaction network and reveal some interesting biological results. Duch and Arenas [10] presented a method to find communities in complex networks based on

* Corresponding author. Tel.: +86 153 2726 0979.
E-mail address: hustcdb@yahoo.com.cn (Z. Lv).

0378-4371/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.05.046

extremal optimization of modularity. Experimental results show that it is feasible to be used for the accurate identification of communities in large complex networks. Wang et al. [11] proposed a very fast algorithm for detecting the community. The algorithm used local information and local modularity to analyze community structures in complex networks. It is based on a table that describes a network and a virtual cache similar to the cache in a computer structure. Other community structure detecting algorithms based on different strategies also have been presented. For example, Clauset et al. [12] presented a hierarchical agglomeration algorithm to detect the community in very large networks. Its running time on a network with n nodes and m edges is $O(md \log n)$, where d is the depth of the dendrogram describing the community structure. Clauset [13] proposed a local community detection algorithm based on local modularity defined by the author. The algorithm adds a node into partial community C and updates the neighbors of C at each step. In 2006, Newman proposed a community structure detecting algorithm using eigenvectors of matrices [14]. Based on the previous benefit function known as modularity, the author presented another benefit function–modularity matrix which plays a role in community detection similar to the Laplacian matrix in graph partitioning calculations. Chen et al. [15] presented a fast and efficient algorithm by adding a node into a partial community recursively until obtaining a local optimal community. An even faster and more accurate algorithm based on subgraph similarity was proposed by Xiang et al. [16].

All the above-mentioned algorithms have a strong assumption that each node belongs to one and only one community. However, a node may belong to several communities in real-world networks, and furthermore, some networks are weighted. Because of this, researchers pay much attention to studying weighted networks and detecting overlapping communities. Barrat et al. [17] studied two real-world weighted networks, i.e., a scientific collaboration network and a world-wide air-transportation network. They defined appropriate metrics combining weighted and topological observables that enable one to characterize the complex statistical properties and heterogeneity of the actual strength of the edges and vertices. The weights characterizing the various connections exhibit complex statistical features with highly varying distributions and power-law behavior. Their study showed that the analysis of the weighted quantities and the study of the correlations between weights and topology provide a complementary perspective on the structural organization of the network that might be undetected by quantities based only on topological information. Ou et al. [18] proposed a model for resource-allocation dynamics to investigate the dynamic behavior of resource/traffic flow on weighted scale-free networks. They found that the dynamical system will evolve into a kinetic equilibrium state, where the strength, defined by the amount of resource or traffic load, is correlated with the degree in a power-law form with tunable exponent. In recent years, overlapping community structures have been widely studied [19–27]. Baumes et al. [19] proposed two efficient local algorithms, *RaRe* and *IS*, to detect the overlapping community structures in networks, and both synthetic and real-world networks were used to test the performance of algorithms. Lancichinetti et al. [20] presented a local algorithm based on local optimization of a fitness function to find overlapping communities which are revealed by peaks in the fitness histogram. In [21], the community structure is detected by k-clique percolation and overlaps between communities are guaranteed by the fact that one node can participate in more than one clique. Evans and Lambiotte [22] used a partition of the links of a network to uncover its overlapping community structure in the complex network. Shen et al. [23] proposed a new measure named Q_c, which is based on a maximal clique view of the original network, to quantify the overlapping community structure and presented an identifying method to detect the overlapping community by finding an optimal cover, i.e., the one with the maximal Q_c. Zhang et al. [24] designed a novel algorithm to identify overlapping communities in complex networks by first mapping the network nodes into Euclidean space and then applying fuzzy c-means clustering. Chen et al. [25] presented another algorithm to detect the overlapping community structure in the network by expanding a partial community which is started from a special single node. Gregory [26] proposed a two-phase method of detecting overlapping communities. In the first phase, a network is transformed to a new one by splitting nodes, using the idea of split betweenness; in the second phase, the transformed network is processed by a disjoint community detection algorithm. Airoldi et al. [27] presented a new overlapping community detecting method, i.e., a Bayesian model, to investigate the protein–protein interactions, including their typical interaction patterns, and the degree of membership of objects to groups. Very recently, Shang et al. [28] proposed an efficient overlapping community detection algorithm based on merging the community cores and expanding the community.

Inspired by the above approaches, in this paper, a new local algorithm based on node strength is proposed to detect the overlapping community structures. The main strategies are to find an initial community from a node with maximal node strength and to expand the partial community from the initial one by adding nodes that are tight with the community. Some real-world and synthetic networks are used to evaluate the presented algorithm and acceptable results are obtained.

2. Fundamental concepts

2.1. Community

A community consists of nodes and edges between these nodes, where nodes often cluster into tightly knit groups with high density of within-group edges and low density of between-group edges. It is noted that two communities may overlap each other since a node can join different communities. For example, as shown in Fig. 1, there are three communities in this network, denoted by circle, square and triangle, respectively. Node 6 is a common node since it should belong to the circle community as well as the square community.

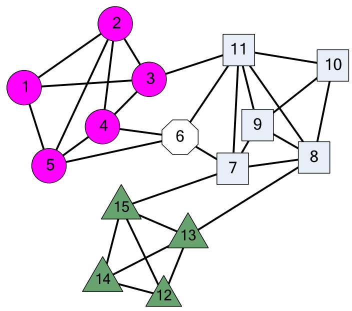

Fig. 1. Example of communities. There are three communities, and node 6 is the common node of both the circle and square communities.

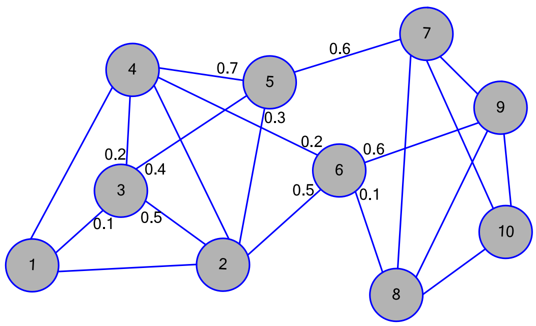

Fig. 2. A simple network and the node strength. The node strength of node 3 and 6 is 1.2 and 1.4, respectively.

2.2. Node strength

For a network $G(V, E)$ with n nodes and m edges, the weight of edge e_{uv} is w_{uv}; $w_{uv} = 0$ if nodes u and v are not connected by an edge. The node strength k_u of node u is defined as

$$k_u = \sum_{v \in V} w_{uv}. \tag{1}$$

For example, as shown in Fig. 2, the node strength of node 3 equals $0.1 + 0.2 + 0.4 + 0.5 = 1.2$, the node strength of node 6 is 1.4, etc. It is noted that if the network is unweighted, $w_{uv} = 1$ if nodes u and v are connected by an edge; otherwise, $w_{uv} = 0$. The node strength of a node is equal to the total number of edges connected with this node, correspondingly.

2.3. Belonging degree

For a community c and a node u, the belonging degree $B(u, c)$ between u and c is defined as

$$B(u, c) = \frac{\sum_{v \in c} w_{uv}}{k_u}. \tag{2}$$

From Eq. (2), $B(u, c)$ reflects how tight the node u is with community c. If all neighbors of node u are included in community c, $B(u, c) = 1$; otherwise, $B(u, c) < 1$. For example, as shown in Fig. 2, if community c includes nodes 1, 2, 3 and 4, the belonging degree $B(5, c)$ between node 5 and community c is $(0.3 + 0.4 + 0.7)/(0.3 + 0.4 + 0.7 + 0.6) = 0.7$, and the belonging degree $B(6, c)$ between node 6 and c is $(0.2 + 0.5)/(0.2 + 0.5 + 0.1 + 0.6) = 0.5$.

2.4. Modularity

In order to quantify the community structure of a network, Newman and Girvan [7] proposed the modularity Q as a measure of network partition:

$$Q = \sum_{i=1}^{k} \left(e_{ii} - a_i^2\right), \tag{3}$$

where e_{ii} is the fraction of weights of edges belonging to community i in the total weights of all edges and a_i is the fraction of weights of edges connecting community i with other communities in the total weights of all edges. If the network is unweighted, the weight of each edge equals 1. Despite the wide acceptance and applications in scientific society, the modularity proposed by Newman and Girvan [7] faces several problems. For example, it suffers a resolution limit problem [29], and cannot tackle overlapping community structure. Therefore, Shen et al. [23] proposed an extended measure based on a clique, Q_c, to quantify the overlapping community structure. Nicosia et al. [30] presented a more general measure, Q_{ov}, to quantify the overlapping community structure in a directed and weighted network.

In this paper, another simple form measure, Q_o, is proposed to quantify the overlapping community structure in a weighted network. In order to describe Q_o clearly, we rewrite the modularity in Eq. (3) as

$$Q = \frac{1}{2m} \sum_{c \in C} \sum_{u,v \in V} \delta_{cu} \delta_{cv} \left(A_{uv} - \frac{k_u k_v}{2m} \right), \tag{4}$$

where A is the adjacency matrix, m is the total number of edges, C is the set of communities corresponding to a partition, and k_u and k_v are the node strengths of nodes u and v defined in Eq. (1), respectively. If the network is unweighted, k_u and k_v are the degrees of node u and v, respectively. δ_{cu} denotes whether node u belongs to community c, which equals 1 if u belongs to community c and 0 otherwise; δ_{cv} is similar.

In the non-overlapping case, a node belongs to only one community. However, a node might be assigned to several communities in the overlapping case, for example, node 6 in Fig. 1. Thus δ_{cu} should be replaced by a belonging coefficient α_{cu} [30], which reflects how much the node u belongs to community c. With the help of the belonging coefficient, the modularity Q_o can be defined as

$$Q_o = \frac{1}{2m} \sum_{c \in C} \sum_{u,v \in V} \alpha_{cu} \alpha_{cv} \left(A_{uv} - \frac{k_u k_v}{2m} \right), \tag{5}$$

where

$$\alpha_{cu} = \frac{k_{cu}}{\sum\limits_{c \in C} k_{cu}} \tag{6}$$

and $k_{cu} = \Sigma_{v \in c} w_{uv}$.

Obviously, the definition in Eq. (6) satisfies the following conditions [30]:

$$0 \le \alpha_{cu} \le 1, \qquad \forall c \in C, u \in V \tag{7}$$

and

$$\sum_{c \in C} \alpha_{cu} = 1. \tag{8}$$

Furthermore, if node u belongs to only one community c, α_{cu} is equal to 1; if node u does not belong to community c, $\alpha_{cu} = 0$. That is to say, Eq. (6) is consistent with the definition of modularity for a non-overlapping community in [7].

3. Description of the method

The method proposed consists of two main components: (i) finding the initial community and (ii) expanding the community. In order to detect the overlapping community structures, we mark the nodes in the finding community with the label "T", denoted by VT, and $V_F = V - V_T$. All nodes are marked with the label "F" initially. We just consider the nodes in V_F while finding an initial community and consider the nodes in V while expanding.

3.1. Finding the initial community

For a given network, we calculate the node strength for each node at first, and then mine an initial community from the node with the largest node strength. The finding processes are described in the following four steps.

(i) Calculate the node strength k_u for node u with label "F" by $k_u \Sigma_{v \in V_F} w_{uv}$.
(ii) Select a node u with the largest node strength and find its neighbors with label "F". These nodes compose an initial community c.
(iii) For each node v in community c, if the belonging degree $B(v, c)$ is less than the threshold B^C ($B^C = 0.5$ in this paper, since we do not consider node v to be tight enough with community c if the belonging degree is less than 0.5), remove node v from c.
(iv) Repeat step (iii) until $\forall v \in c$, $B(v, c) \ge B^C$, and thus obtain the initial partial community, also denoted by c.

For example, as shown in Fig. 3, we calculate the node strength of each node, as shown in Table 1.

So, we select node 2 and its neighbors $\{1, 3, 4, 6, 8\}$ as the initial community $\{1, 2, 3, 4, 6, 8\}$. Remove node 8 from the initial community since the belonging degree ($=0.3/1.2 = 0.25$) is less than B^C ($=0.5$); the other nodes need not be removed. So, the initial community c is $\{1, 2, 3, 4, 6\}$ which is marked with pink circle.

Table 1
The node strength of each node as shown in Fig. 3.

Node index	1	2	3	4	5	6	7	8	9	10
Node strength	1.7	2.4	2.0	1.6	1.6	1.4	1.8	1.2	2.0	1.5

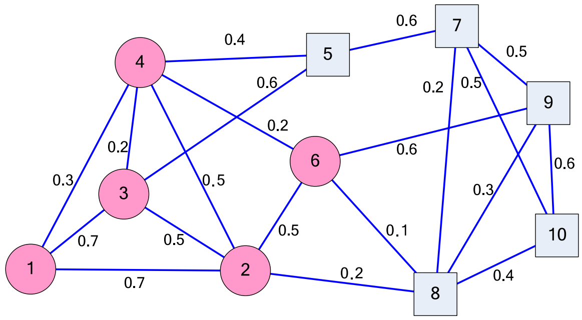

Fig. 3. Example network for finding the initial community. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

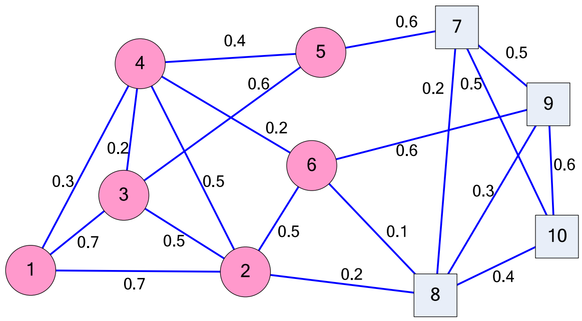

Fig. 4. Example network for community expanding. Community $\{1, 2, 3, 4, 5, 6\}$ is expanded from $\{1, 2, 3, 4, 6\}$.

3.2. Expanding the community

Find all neighbors N_c of the initial community c. If $\exists u \in N_c, B(u, c) > B^U$ ($B^U = 0.5$ in this paper since a node u is tight enough with community c if the belonging degree is larger than 0.5), add all these type nodes into community c directly. If $\forall u \in N_c, B(u, c) < B^L$ ($B^L = 0.4$ in this paper since a node u is loose enough with community c if the belonging degree is less than 0.4, and the node needs checking further using the modularity Q_o for a belonging degree between 0.4 and 0.5), stop expanding the community; otherwise, add the nodes with $B^L \le B(u, c) \le B^U$ into the community if the modularity Q_o becomes larger after adding. The detailed processes are listed as follows.

(i) Find all neighbors N_c of community c, and calculate the belonging degree $B(u, c)$ for each neighbor u.
(ii) Find all nodes with $B(u, c) > B^U$ and $B^L \le B(u, c) \le B^U$, denoted by $N_u = \{u|B(u, c) > B^U\}$ and $N_{lu} = \{u|B^L \le B(u, c) \le B^U\}$, respectively.
(iii) If $|N_u| > 0$, add all nodes of N_u into the community and obtain a larger partial community, also denoted by c, and return to step (i).
(iv) If $|N_{lu}| > 0$, for each node u in N_{lu}, add u into the community if the modularity Q_o becomes larger after adding it, obtain a larger partial community, also denoted by c, and return to step (i).
(v) If $|N_u| = 0$ and $|N_{lu}| = 0$, stop expanding and mine a community ultimately.

For example, as shown in Fig. 3, the neighbors of community $c = \{1, 2, 3, 4, 6\}$ are $\{5, 8, 9\}$. Since $B\,(5, c) = 5/8$, $B(8, c) = 1/4$ and $B(9, c) = 3/10$, add node 5 into community c and expand to a larger community, also denoted by c. After adding node 5, all neighbor nodes of community c satisfy $B(., c) < 0.4$, so stop the expanding process and obtain an ultimate community $\{1, 2, 3, 4, 5, 6\}$, as shown in Fig. 4.

After extracting a community c, mark the nodes in c with the label "T". Repeat the processes described in Sections 3.1 and 3.2 to find all other communities.

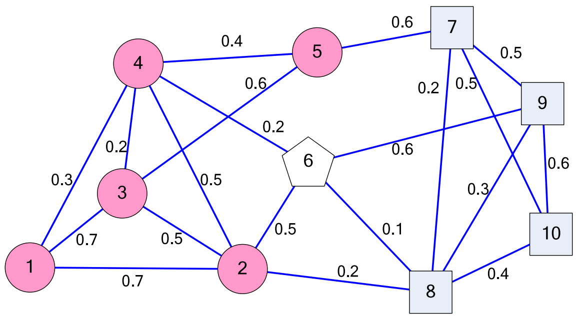

Fig. 5. Final partition of the network. Node 6 is a common node.

For example, as shown in Fig. 4, after extracting the first community $c = \{1, 2, 3, 4, 5, 6\}$, start from node 10 to find another community, i.e. {6, 7, 8, 9, 10}. Node 6 is a common node of the two communities found, as shown in Fig. 5.

3.3. Computational complexity

In order to find an initial community, $O(n)$ nodes need investigating. At the expanding step, all neighbors of a partial community, i.e., $O(n)$ nodes, should be investigated while adding some proper nodes. After adding new nodes into the community, a larger community will be obtained. In the worst case, $O(n)$ nodes will be added into the community, so this process will be repeated $O(n)$ times in order to obtain an ultimate community. And the computational complexity is $O(n + n * n) = O(n^2)$ while finding a community. There are $O(n)$ communities in the worst case, so the complexity of the method is $O(n^3)$. However, in the general case, a community contains $O(1)$ nodes, so the computational complexity is $O(n)$ while finding a community, and it is $O(n^2)$ to find all communities.

4. Experimental results

The overlapping community detecting algorithm proposed in this paper is implemented by C#.net programming language running on a PC with a 3.0 GHz processor and 3.0 GB memory.

A synthetic weighted network built by the method in [31,32] and seven classic real-world networks taken from the literature were used to evaluate the proposed algorithm. That is, the friendship network from Zachary's karate club study [33], college football [34], collaboration network [14], dolphin's associations [35], cond-mat-2003 [36], email, and PGP [26]. The first two real-world network communities are already known from other sources. These networks can be downloaded from http://www-personal.umich.edu/~mejn/netdata/ and http://www.cs.bris.ac.uk/~steve/networks/peacockpaper/.

For some large-scale networks, such as the collaboration network [14], cond-mat-2003 [36], email [26] and PGP [26], the comparisons of execution time and modularity between some well-known algorithms and our approach are listed in Table 2. From Table 2, we can see that the execution time of our approach is fairly short, the division is rather good, and the Peacock + CNM algorithm gives better division of overlapping communities but with longer execution time.

4.1. Synthetic weighted network

In order to evaluate our method, we use the method presented in [31,32] to build a weighted network with 50 000 nodes and 163 042 edges which has 543 communities. Some main parameters are listed in Table 3. 614 overlapping communities with modularity 0.3133 and execution time 792 s are detected by our method.

4.2. Friendship network

The well-known "karate club" by Zachary [33] is widely used as a benchmark for community detection. This network describes the social interactions between the individuals in a karate club in an American university; it contains 34 nodes and 78 edges. By chance, a dispute arose between the club's administrator and the club's instructor, and as a result, the club eventually split into two smaller clubs, centered on the administrator and the instructor, respectively. The network and its fission are depicted in Fig. 6. The administrator and the instructor are represented by node 1 and node 33, respectively.

As shown in Fig. 6, two overlapping communities with node 10 in common are detected by our method, with Q_o being 0.4214. This network is also discussed in [22–25]. In [22,23], four communities were found by the authors. In Zhang et al. [24], three communities and four common nodes are found. In [25], there are two communities and five common nodes, that is, nodes 3, 9, 10, 14 and 31. Our algorithm finds two communities, which is in accordance with the real split. Since node 10 has one connection with each community, it is reasonable to consider it as a common node of two communities.

Table 2
Execution time and modularity on some real-world networks.

Instance	Nodes	Edges	CONGA[a] [37]		CFinder [38]		Peacock + CNM [26]		Peacock + WT [26]		Peacock + PL [26]		Peacock + BGLL [26]		Our approach	
			Time (s)	Q_{ov}	Time (s)	Q_{ov}	Time (s)	Q_{ov}	Time (s)	Q_{ov}	Time (s)	Q_{ov}	Time (s)	Q_{ov}	Time (s)[b]	Q_{ov}[c]
Netscience	379	914	1.3	0.75	0.25	0.42	0.95	0.93	1.25	0.6	0.38	0.86	0.41	0.71	0.08	0.8492
cond-mat-2003[d]	27 519	116 181	1127	0.41	1134	0.57	1171	0.47	1099	0.4	1112	0.57	1091.7	0.45	175	0.4521
Email	1 133	5 451	30.2	0.24	4	0.46	36.6	0.53	39.21	0.1	33.05	0.37	32.84	0.16	0.67	0.4172
PGP	10 680	24 316	83	0.56	34 745	0.57	108.4	0.71	94.42	0.6	92.46	0.72	90.7	0.66	19.52	0.6428

[a] The execution time and modularity of CONGA, CFinder, Peacock + CNM, Peacock+WT, Peacock + PL and Peacock + BGLL are taken from Table 1 and Fig. 10 of [26]. These algorithms are to run on an AMD Opteron 250 with 2.4 GHz.
[b] Our approach is to run on a PC with 3.0 GHz and 3.0 GB memory.
[c] The modularity formula using here is taken from [26,30].
[d] In order to compare with other algorithms, cond-mat-2003 used here is unweighted.

Table 3
Parameters of synthetic weighted network.

N	μ_t	μ_w	$\langle k \rangle$	k_{max}	s_{min}	s_{max}	Overlapping nodes	Memberships of the overlapping nodes
50 000	0.5	0.1	5	100	50	150	20	2

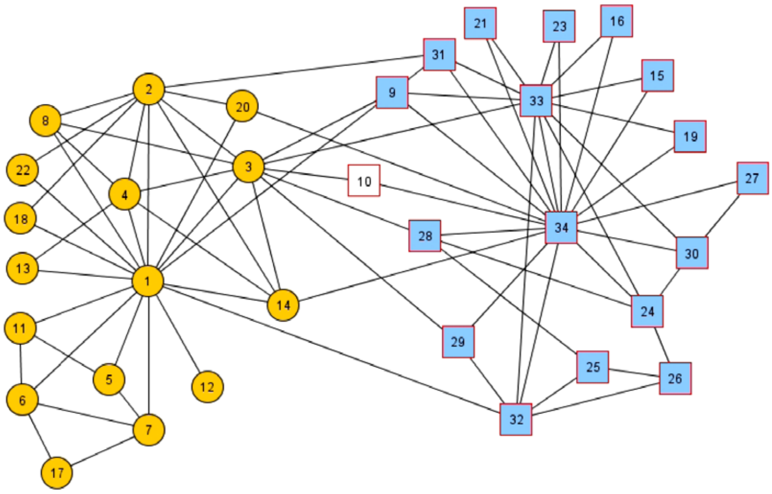

Fig. 6. Two overlapping communities in the friendship network from Zachary's karate club study.

4.3. *College football*

The network of college football studied in [34] is also used as a benchmark for the community detecting. This network describes an American college football games between Division I colleges during the regular season in fall 2000, where nodes denote football teams and edges represent regular season games. It consists of 115 nodes and 616 edges, and can be divided into 12 communities according to athletic conferences. Each community contains 8 to 12 teams.

13 non-overlapping communities with Q_o being 0.5868 are detected by our method, as shown in Fig. 7. Six communities are detected correctly by our method, that is, *Atlantic Coast*, *Big* 10, *Big* 12, *Mountain West*, *Pacific* 10 and *SEC*. Communities *Western athletic*, *Conference USA* and *Big East* are nearly correct, with one or two teams being assigned to the incorrect community. All teams in *Independents* are assigned to other groups since sparse connections exist among these teams while they have density connections with other groups. It is found that the misclassified teams are hard to be correctly identified. For example, team *Notre Dame* in community *Independents* is assigned to community *Big East* by our method because *Notre Dame* has three connections with *Big East* yet no connection with *Independents*. Similarly, *Boise State* in community *Western Athletic* is assigned to *Sunbelt* by our method. But this assignment is reasonably since *Boise State* has four connections with *Sunbelt* and only one connection with five other groups, respectively. Community *Mid American* is nearly split into two subgroups by our method although the teams are tightly connected. This needs to be studied further in the future.

4.4. *Dolphin's associations*

This data set is taken from [35]. It describes the associations between 62 dolphins living in Doubtful Sound, New Zealand, with ties between dolphin pairs being the statistically significant frequent association. This network can be split naturally into two groups. For this network, four overlapping communities are detected by our algorithm with modularity Q_o 0.5478. Three common nodes, i.e., 3(Bumper), 40(SN89), 62(Zipfel), connect three communities, as shown in Fig. 8. This network has also been studied in [23,25], and the modularity Q_o is 0.5481 and 0.3286, respectively. In [25], two communities with 9 common nodes are found and four communities with no common nodes are found in [23].

4.5. *Collaboration network*

This benchmark is taken from a weighted collaboration network of coauthorships between scientists who are themselves publishing on the topic of networks [14]. There are in total 1589 scientists in this collaboration network. We just take the largest connected component with 397 scientists as the test benchmark. We find the tight communities in this network including some small communities with modularity Q_o being 0.8531, as shown in Fig. 9. In Fig. 9, the white parallelograms are common nodes between communities, many of these common nodes are cross-disciplinary researchers, such as node 106(Kleinberg, J), 131(Crucitti, P) and 214(Schnitzler, A), etc.

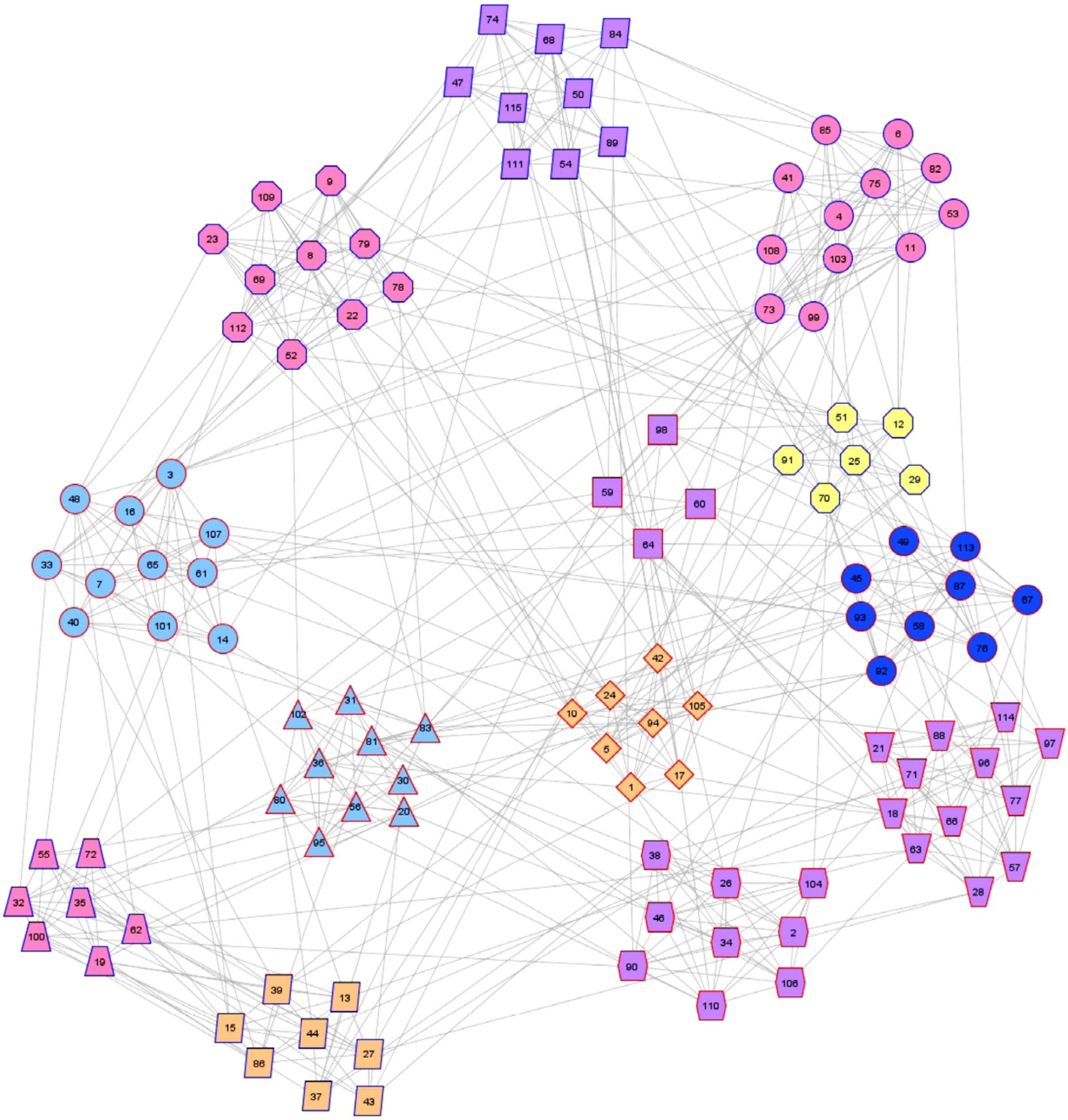

Fig. 7. The 13 non-overlapping communities in the college football network.

4.6. Condensed matter collaboration 2003

This benchmark is another classic weighted network of coauthorships between scientists posting preprints on the Condensed Matter E-Print Archive. It includes all preprints post between January 1, 1995 and June 30, 2003. The largest component of this network, which contains 27 519 scientists, has been used by several authors as a test-bed for community finding algorithms for large networks.

978 overlapping communities with modularity 0.4651 and execution time is 48 s are found by our method.

5. Conclusions

An efficient local algorithm to detect overlapping communities in weighted networks is proposed in this paper. The key aspect of the proposed algorithm is selecting a node with maximal node strength and extracting a partial community. Then, the partial community is expanded to an ultimate one. Seven classical real-world networks and a synthetic network were

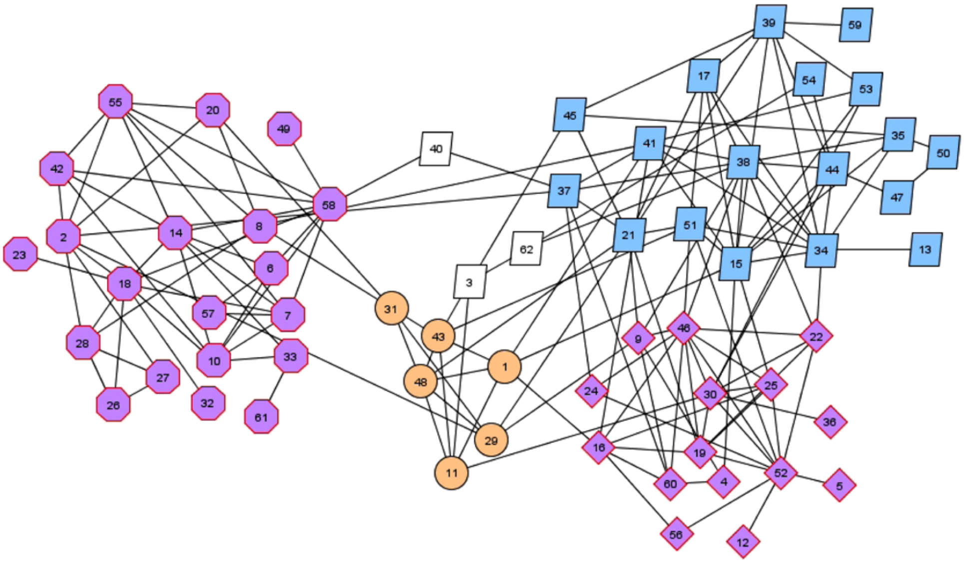

Fig. 8. The four overlapping communities in the dolphin association network.

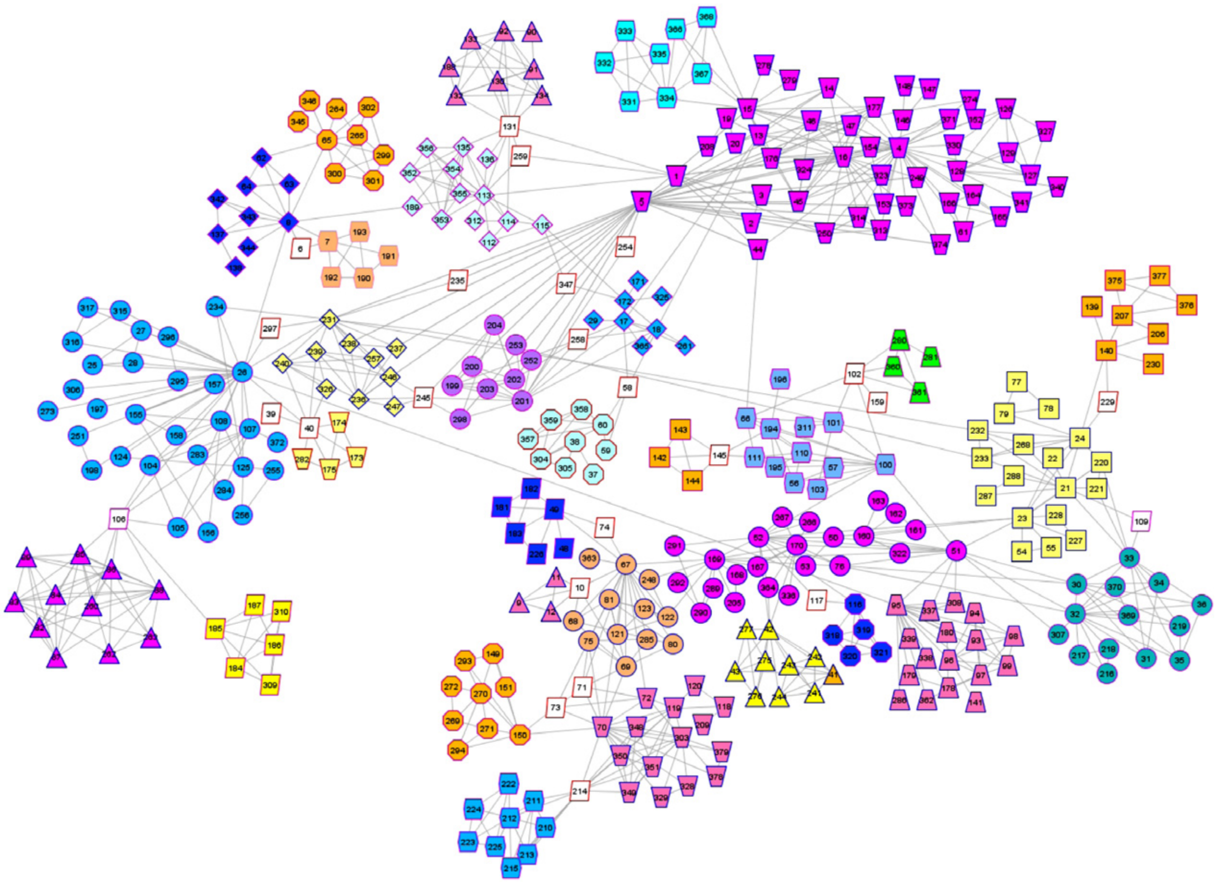

Fig. 9. The 31 overlapping communities from the collaboration network.

used to evaluate the algorithm presented. For the karate club's friendship network, two communities with one common node were found correctly by the algorithm. For the college football network, six communities were detected correctly, and most of the other teams were assigned to the correct community except *Mid American*. For the scientist collaborate network and dolphin association network, 31 and 4 overlapping communities, with modularity 0.8531 and 0.5478, were found, respectively. Some large networks such as cond-mat-2003 [36] and PGP [26] were also tested by our method and acceptable results were obtained. As compared with our method, the Peacock + CNM algorithm gives better division of overlapping communities but with longer execution time. Experimental results demonstrate that the proposed algorithm is rather efficient for detecting overlapping communities in weighted networks.

Acknowledgements

Thanks are due to the anonymous referees and editors for many instructive suggestions. We would like to thank Prof. S. Gregory, Prof. S. Fortunato and Prof. A. Lancichinetti for their enthusiastic help and Dr. Zhipeng Lü for checking our paper carefully. This work was partially supported by the National Natural Science Foundation of China under Grant Nos. 60973069, 90924011, 60973120 and 60903073, by the International Scientific Cooperation and Communication Project of Sichuan Province in China under Grant No. 2010HH0002, and by the China Postdoctoral Science Foundation under Grant No. 20080431273.

References

[1] M.E.J Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.
[2] R. Albert, A.L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[3] S.H. Strogatz, Exploring complex networks, Nature 410 (2001) 268–276.
[4] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (2010) 75–174.
[5] M. Fiedler, Algebraic connectivity of graphs, Czech Math. J. 23 (1973) 298–305.
[6] B.W. Kernighan, S. Lin, An efficient heuristic procedure for partition graphs, Bell. Syst. Tech. J. 49 (1970) 291–307.
[7] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113.
[8] M.E.J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (2004) 066133.
[9] J.H. Ruan, W.X. Zhang, Identifying network communities with a high resolution, Phys. Rev. E 77 (2008) 016104.
[10] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, Phys. Rev. E 72 (2005) 027104.
[11] X.T. Wang, G.R. Chen, H.T. Lu, A very fast algorithm for detecting community structures in complex networks, Physica A 384 (2007) 667–674.
[12] A. Clauset, M.E.J. Newman, C. Moore, Finding community structure in very large networks, Phys. Rev. E 70 (2004) 066111.
[13] A. Clauset, Finding local community structure in networks, Phys. Rev. E 72 (2005) 026132.
[14] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E 74 (2006) 036104.
[15] D.B. Chen, Y. Fu, M.S. Shang, A fast and efficient heuristic algorithm for detecting community structures in complex networks, Physica A 388 (2009) 2741–2749.
[16] B. Xiang, E.H. Chen, T. Zhou, Finding community structure based on subgraph similarity, Stud. Comput. Intell. 207 (2009) 73–81.
[17] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. USA 101 (2004) 3747–3752.
[18] Q. Ou, Y.D. Jin, T. Zhou, B.H. Wang, B.Q. Yin, Power-law strength–degree correlation from a resource-allocation dynamics on weighted networks, Phys. Rev. E 75 (2007) 021102.
[19] J. Baumes, M.K. Goldberg, M.S. Krishnamoorthy, M.M. Ismail, N. Preston, Finding communities by clustering a graph into overlapping subgraphs, in: N. Guimaraes, P.T. Isaias (Eds.), 2005 Proc. IADIS International Conference on Applied Computing, pp. 97–104.
[20] A. Lancichinetti, S. Fortunato, J. Kertesz, Detecting the overlapping and hierarchical community structures in complex networks, New. J. Phys. 11 (2009) 033015.
[21] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (2005) 814–818.
[22] T.S. Evans, R. Lambiotte, Line graphs, link partitions, and overlapping communities, Phys. Rev. E 80 (2009) 016105.
[23] H.W. Shen, X.Q. Cheng, J.F. Guo, Quantifying and identifying the overlapping community structure in networks, J. Stat. Mech. (2009) P07042.
[24] S. Zhang, R.S. Wang, X.S. Zhang, Identification of overlapping community structure in complex networks using fuzzy c-means clustering, Physica A 374 (2007) 483–490.
[25] D.B. Chen, Y. Fu, M.S. Shang, An efficient algorithm for overlapping community detection in complex networks, in: 2009 Proc. Global Congress on Intelligent Systems, 19–21 May, Xiamen China, pp. 244–247.
[26] S. Gregory, Finding overlapping communities using disjoint community detection algorithms, Stud. Comput. Intell. 207 (2009) 47–61.
[27] E.M. Airoldi, D.M. Blei, E.P. Xing, S.E. Fienberg, Mixed membership stochastic block models for relational data, with applications to protein–protein interactions, in: 2006, Proc. International Biometric Society-ENAR Annual Meetings, 26–29 March, Hyatt Regency Tampa.
[28] M.S. Shang, D.B. Chen, T. Zhou, Detecting overlapping communities based on community cores in complex networks, Chinese. Phys. Lett. 27 (2010) 058901.
[29] S. Fortunato, M. Barthélemy, Resolution limit in community detection, Proc. Natl. Acad. Sci. USA 104 (2007) 36–41.
[30] V. Nicosia, G. Mangioni, V. Carchiolo, M. Malgeri, Extending the definition of modularity to directed graphs with overlapping communities, J. Stat. Mech. (2009) P03024.
[31] A. Lancichinetti, S. Fortunato, F. Radicchi, Benchmark graphs for testing community detection algorithms, Phys. Rev. E 78 (2008) 046110.
[32] A. Lancichinetti, S. Fortunato, Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities, Phys. Rev. E 80 (2009) 016118.
[33] W.W. Zachary, An information flow model for conflict and fission in small groups, J. Anth. Res. 33 (1977) 452–473.
[34] M. Girvan, M.E.J Newman, Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99 (2002) 7821–7826.
[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. 54 (2003) 396–405.
[36] M.E.J. Newman, The structure of scientific collaboration networks, Proc. Nat. Acad. Sci. USA 98 (2001) 404–409.
[37] S. Gregory, A fast algorithm to find overlapping communities in networks, in: PKDD, LNAI, 5211, Springer, Heidelberg, 2008, pp. 408–423.
[38] B. Adamcsek, G. Palla, I. Farkas, I. Derényi, T. Vicsek, CFinder: locating cliques and overlapping modules in biological networks, Bioinformatics 22 (2006) 1021–1023.