PRAMANA — journal of physics
© Indian Academy of Sciences
Vol. 80, No. 1
January 2013
pp. 173–185

Community detection using global and local structural information

HAI-LONG YAN[1], JU XIANG[2,*], XIAO-YU ZHANG[2], JUN-FENG FAN[2], FANG CHEN[2], GEN-YI FU[2], ER-MIN GUO[2], XIN-GUANG HU[3], KE HU[4] and RU-MIN WANG[1]

[1]College of Physics and Electronic Engineering, Xinyang Normal University, Xinyang 464000, Henan, China
[2]Department of Basic Sciences, The First Aeronautical Institute of the Air Force, Xinyang 464000, Henan, China
[3]School of Information Engineering, HuangShan University, HuangShan 245021, Anhui, China
[4]Department of Physics, Xiangtan University, Xiangtan 411105, Hunan, China
*Corresponding author. E-mail: xiangju0208@yahoo.com.cn

Abstract. Community detection is of considerable importance for understanding both the structure and function of complex networks. In this paper, we introduced the general procedure of the community detection algorithms using global and local structural information, where the edge betweenness and the local similarity measures respectively based on local random walk dynamics and local cyclic structures were used. The algorithms were tested on artificial and real-world networks. The results clearly show that all the algorithms have excellent performance in the tests and the local similarity measure based on local random walk dynamics is superior to that based on local cyclic structures.

Keywords. Complex network; community structure; edge betweenness; local random walk.

PACS Nos 89.75.–k; 89.75.Fb; 89.75.Hc

1. Introduction

Community structure in complex networks – groups of vertices within which connections are dense while between which they are sparser – is an important topological property common to many real-world networks such as social, biological, as well as technological networks which has attracted much attention in physics and many interdisciplinary fields [1,2]. The communities or modules in complex networks generally correspond to such functional units as real social groupings in social networks, sets of pages on related topics on the WWW, sets of papers on related research subfields in citation networks, or cycles and pathways in metabolic networks [1,3–5]. Detecting such communities in complex networks will be of considerable interest, since it can provide a useful coarse-grained

DOI: 10.1007/s12043-012-0359-5; ***e*Publication: 20 December 2012**

representation of complex networks and can help in analysing the structure and function of the networks [1,2,6–8].

In the past decade, many community detection algorithms have been proposed based on various approaches such as edge clustering coefficients [9], dissimilarity and similarity measures [10,11], spectral analysis methods [12,13], physical models [14–16] and optimization [17–23], since the work of Girvan and Newman [24] (see refs [1,23,25,26] for reviews). Moreover, an important quality function for community division of network known as modularity, was also proposed by Newman and Girvan, which measures the probability of edges falling inside the communities in the network minus the expected value in the null model [27]. In general, the larger the modularity, the better the division is, as the more it deviates from the null model. Further, to make use of potential information contained in the weights and directions of the edges, it has also been generalized to the weighted and directed cases [28,29]. The modularity measure provides an objective way to evaluate the quality of community divisions of networks, and it also suggests a kind of alternative strategy for community detection – the modularity optimization [1].

In recent years, the modularity optimization becomes a very popular strategy in community detection, while it also encounters some difficulties such as the resolution limit – the embedded modules below some scales will be undetectable [30–32]. Thus, its improved versions as well as many new methods for community detection, such as the multiresolution methods [33–35] and the Bayesian-theory-based method [36] are being proposed continually. Interestingly, by redefining the weight of edges with local similarities and then feeding the weighted network to the weighted-modularity-based algorithms [37,38], one can find the embedded communities in the network better. As we know, the vertices in the same communities are usually of higher similarity, and thus the similarity measures between the vertices will have higher values. Redefining the weight of the edges with similarity measures can help in increasing the difference between the intra- and intercommunity edges, further making the community structure in the network becoming clearer and easier to be detected than before. In this paper, we shall study the community detection algorithms based on global and local structural information, using the 'trick'. Specifically, the edge betweenness and local similarity measures based respectively on local random walk dynamics and local cyclic structures will be used and compared. We believe that it will be worthwhile refocussing on some typical community detection methods, since they may be able to give exciting results, especially when some popular methods are suffering from certain intrinsic limits [30–32,39]. In the following sections, we show the general procedure of the new community detection algorithms, and then test them on artificial and real networks. Finally we come to our conclusions.

2. Methods

For identifying of community structures in networks or other purposes, researchers from various fields have proposed many approaches that can depict the properties of the edges and vertices in the networks. Sociologists developed a variety of ways of defining the similarity between vertices in the study of social networks [40]. They seem to prefer the property known as structural equivalence. Two vertices are said to be structurally equivalent if they have the same set of neighbours. For example, the Pearson correlation between columns (or rows) of the adjacency matrix is a commonly used similarity measure in sociology [41]. Moreover, the number of edge- (or vertex-) independent paths between vertices is another useful similarity measure, based on structural equivalence (two paths that connect the same pair of vertices are said to be node-(or edge-) independent if they share none of the same vertices (or edge) other than their initial and final vertices) [42]. These traditional measures may be very useful in some cases, but are unsatisfactory for the community structure analysis of many general real-world networks. So, in recent years, many new measures were developed to extract information about community structure in networks, such as the edge betweenness [24], the edge clustering coefficients [9], the information centrality based on the network efficiency [43,44], the dissimilarity and similarity measures [10,11], and so on. In general, these measures may be classified into two classes: the one based on the global structures of networks and the other based on only local structural information of networks under study, which can reflect or emphasize different aspects of network structures.

2.1 *Information extraction based on local network structure*

To increase the difference between weights of the inter- and intracommunity edges, any reasonable method, based on global or local information of networks, can be used to reweight the edges in the networks. In the literatures, many similarity measures have been proposed and discussed [40,45,46], which will be helpful to some extent for community detection. Here, we shall focus on two similarity measures based on local structural information, recently proposed by Lai *et al* [37] and Berry *et al* [38], which can provide important information on local structures in networks, and they will be compared in the following texts. Due to the local nature of these measures, they can be calculated very quickly, and they may give us some useful local information on communities beyond global measures such as edge betweenness.

Similarity based on local random walk dynamics. In general, the dynamic processes triggered on vertices in the same community possess similar behaviour patterns, but dissimilar on vertices in different communities. Based on the observation, Lai *et al* (LLN) proposed the similarity measure based on local random walk dynamics. The probability of a walker from one vertex to another in t-step random walk is determined by the matrix P^t. The parameter t is the random walk length, determining the range of the local structure that will be browsed. In general networks, good results can often be obtained by using a small t-value ($t = 2, 3, \ldots$) [37]. The element p_{ij} of the transition matrix P is the ratio between the weight of the edge (i,j) and the weighted degree of vertex i, i.e. $p_{ij} = w_{ij}/w_i$. The behaviour patterns of the random walk dynamics from each vertex can be quantified by an n-dimensional vector $v_i (i = 1–N)$ – the row of the matrix $\sum_{\tau=1}^{t} P^{\tau}$ where all the random walks whose steps vary from 1 to t are taken into account to reinforce the contributions from the vertices near the target vertices currently considered. The similarity of Lai *et al* (LLN), using the cosine of two vectors, is defined as [37]

$$S_{i,j}^{\mathrm{LLN}} = \frac{(v_i, v_j)}{\sqrt{(v_i, v_i)}\sqrt{(v_j, v_j)}}, \tag{1}$$

where (v_i, v_j) is the inner product of the vectors v_i and v_j. If the behaviour vectors v_i and v_j are highly consistent, then the similarity $S_{i,j}^{\mathrm{LLN}} \to 1$, 0 otherwise. Generally, the pairs of vertices in the same communities have higher values of similarity than those in different communities. So the similarity measure can be used as the probability of an edge to lie in a community.

Similarity based on local cyclic structure. To re-weight the inter- and intracommunity edges in the networks, Berry *et al* proposed a similarity measure based on local cyclic structures, called neighbourhood coherence (NC). It is defined as [38]

$$S_{i,j}^{\mathrm{NC}} = \frac{G_e}{W_e}, \tag{2}$$

where G_e is the sum of the weight of the edges that reside on at most four order cycles containing the edge $e = (i, j)$ and are incident on vertices i and j, W_e is the total weight of the edges incident on the two end-points of the edge $e = (i, j)$. In general, the edges between communities are only included in few local cyclic structures, while there are a number of local cycles inside the communities. So the intercommunity edges have a low value of the neighbourhood coherence index, while the intracommunity edges have a high value.

2.2 *Information extraction based on global network structures*

In the past decade, many measures based on the global structure of networks have been proposed and discussed, which have been used to evaluate the properties of the edges (or vertices) in certain aspects from the viewpoints of the whole networks. For example, the information centrality and 'betweenness' centrality of the edges have been proposed to identify the intra- or intercommunity edges in networks, and further to detect the communities in the networks. The 'betweenness' centrality was first proposed by Freeman as a measure of the centrality and influence of vertices in the networks, which is defined as the number of the shortest paths on the network that run through a given vertex [47]. To identify which edges in a network are most 'between' communities, Girvan and Newman directly extended the Freeman's betweenness centrality to the edges [24]. So, the betweenness of edges (EB), as a natural generalization to edges of the vertex betweenness, is defined as the number of all the shortest paths on a network that run along a given edge. Clearly, if the network consists of communities only loosely connected by a few intercommunity edges, all shortest paths between vertices in different communities have to go through few intercommunity edges. Then, the edges connecting communities will have higher edge betweenness score than those inside the communities and by recursively removing these edges, one can obtain a good community division of the network. The information centrality has its property similar to the betweenness. In this paper, the edge betweenness will be used in the following.

2.3 *General procedure*

Here, we shall try to detect communities in networks by combining global and local structural measures, i.e. the edge betweenness (EB) and the local similarity measure (LLN or

NC). First of all, for each edge of the original network, we shall redefine the edge weight by the similarity measure (LLN or NC) between the two vertices connected by the edge, and then a new weighted network can be obtained. Repeating the strategy on the new weighted network, we can iteratively enlarge the differences between weights of inter- and intracommunity edges. After re-weighting the edges in the network, we shall divide the virtual-weighted network. More formally, the general procedure of the new community detection algorithm is the following (for brevity, let LL-EB and NC-EB denote the procedures with LLN and NC respectively).

(1) Redefine the weight $w_{i,j}$ of the edges using LLN or NC.
(2) Calculate the betweenness $B_{i,j}$ for all edges in the original un-weighted network.
(3) Remove the edge with the highest value of $B_{i,j}/w_{i,j}$.
(4) Recalculate the betweenness $B_{i,j}$ for all edges affected by the removal, ignoring the weights.
(5) Repeat step (3) until no edges remain.

The output of the new algorithm is in the form of a dendrogram, which represents an entire hierarchy of possible community divisions of the network found by the algorithm. To get a suitable community division of the network from the dendrogram, there are at least two choices, e.g. by using the un-weighted modularity Q or the weighted modularity Q_{w}, both of which will be tried in the following sections. Given a weighted network partitioned into communities, the definition of the weighted modularity [28] can be expressed as

$$Q_{\mathrm{w}} = \sum_{r} \frac{w_{rr}}{W} - \left(\frac{w_r}{2W}\right)^2, \tag{3}$$

where W is the total weight of the edges in the network, w_{rr} is the total weight of edges within the group r, w_r is the weighted degree of group r, and the sum over all communities in the given network. For un-weighted case, w_{rr} becomes the number of edges within the group r, w_r the degree of group r, and W the total number of edges in the network. In order to avoid confusion with the weighted case in the following sections, the un-weighted modularity [27] is denoted by Q.

3. Application

In this section, we shall apply the new algorithms to the Girvan–Newman benchmark [24] – a set of computer-generated networks where the pre-defined communities are within the range of the resolution of modularity, the network with the pre-defined communities being out of the resolution of modularity – a generalization of the Girvan–Newman benchmark, the LFR networks with more realistic network properties, as well as the Zachary's karate club network [46], and then compare the results with those by the original GN algorithm.

3.1 *Girvan–Newman benchmark*

We have created a set of computer-generated networks, where pre-defined communities are clearly within the range of the resolution of modularity. Each of these networks consists of four communities of 32 vertices. Each vertex has on average k_{in} edges connecting it to vertices in the same community and k_{out} edges connecting it to vertices in other communities, with k_{in} and k_{out} chosen so as to keep the total expected degree $k_{\text{in}} + k_{\text{out}} = 16$. With k_{out} increasing from zero, the communities in the networks become more and more difficult to identify. Different algorithms, when applied to these networks, may give different results, reflecting their ability of detecting communities in the networks.

Since the 'real' community structure is well known in these networks, here we shall use the normalized mutual information measure (NMI) to evaluate the performance of the detection algorithms, which can measure the amount of information correctly extracted by the algorithms [25]. This measure is to estimate the similarity between the real and the found communities. When they are perfectly matched, the NMI is 1. Otherwise, the less is a match between the found and the real communities, the smaller is the value of the NMI. Figure 1 shows the NMI between the real and the found communities in the computer-generated networks by different algorithms as a function of the average number k_{out} of intercommunity edges per vertex.

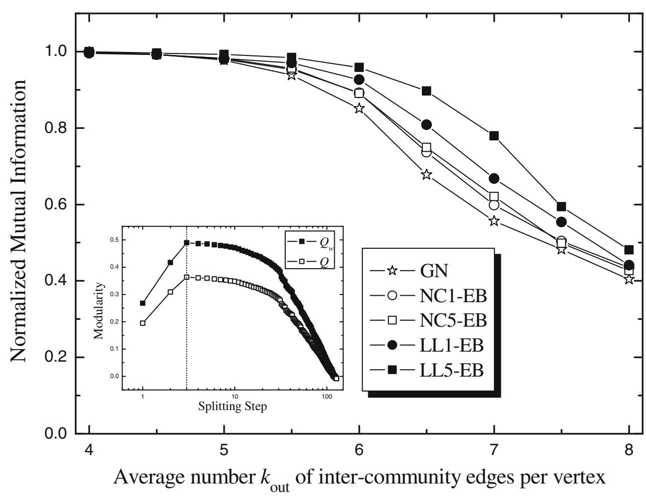


Figure 1. The normalized mutual information (NMI) between the real and the found communities in the networks described in the text. The five curves show results in these networks by five algorithms – as a function of the average number k_{out} of the intercommunity edges per vertex. LL5 (LL1) means that the LLN will be re-applied five times (one time) to re-weight the edges of the networks (the random walk length $t = 3$, though larger t-values can slightly improve the results of LLN, as shown in [37]), and NC5 (NC1) means that the NC will be re-applied five times (one time) to re-weight the edges of the networks. The inset in figure 1 shows the curves of modularity Q (and Q_{w}) vs. the splitting step of the network with $k_{\text{out}} = 6$. It clearly indicates that the third splitting of the network produces maximum modularity values for both Q and Q_{w}. We found that all the algorithms give similar results in the test, i.e. maximum modularity values for Q and Q_{w} will be obtained almost at the same splitting step of the networks.

Because the communities in the artificial networks are within the range of the resolution of modularity, similar results can be obtained by using the un-weighted modularity Q or the weighted modularity Q_w to choose the community divisions from the dendrograms by different algorithms (see the inset in figure 1). In figure 1, as we see, all the algorithms work very well, while for higher value of k_{out} the new algorithms will work better, and the results can be improved by more re-weighting iterations, especially for LLN. Moreover, the result shows that LLN can give better outputs than NC. The reason may be that except for the special dynamical processes and definition of LLN, NC only considers at most four-order cycles containing the edge under study, while LLN with t-step random walk is related to $(2t + 1)$-order cycles. For example, LLN with $t = 3$ may be related to 7-order cycles, which is clearly larger than NC's. This can thus help LLN to extract much information about communities more easily.

3.2 *Generalization of Girvan–Newman (GN) benchmark*

Here we generate a network where the pre-defined community structure is out of the resolution of modularity. The network, which is a generalization of the Girvan–Newman benchmark, contains two compartments (C1 and C2) (see the dashed line boxes in figure 2), and each compartment contains two communities of 32 vertices (C11 and C12 for C1, and C21 and C22 for C2) (see the solid line boxes in figure 2). Each vertex has on average k_{in1} edges connecting it to vertices in the same community, k_{in2} edges to vertices of another community in the same compartment and k_{out} edges to any other vertices at

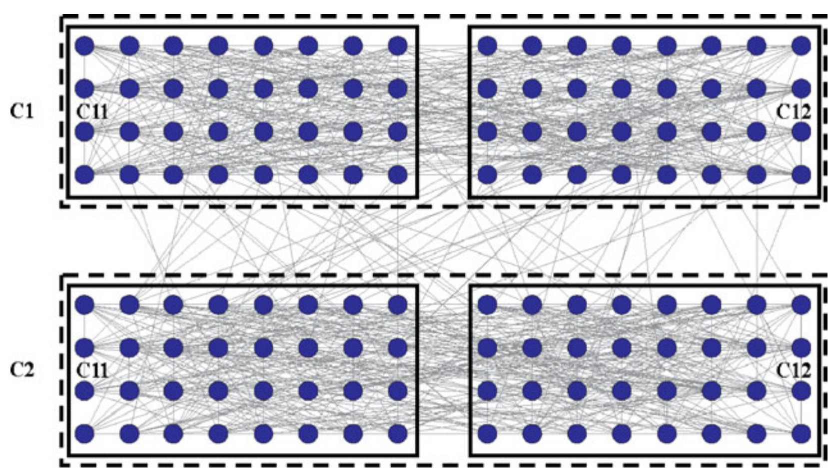


Figure 2. The network example with $k_{in1} = 11$ and $k_{in2} = 4$, described in the text, where the pre-defined community structure is out of the resolution of modularity (the network was drawn by Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek/)). The network consists of four communities (C11, C12, C21 and C22), where C11 and C12 belong to compartment C1 while C21 and C22 belong to C2. C11, C12, C21 and C22 contain vertices (1–32), (33–64), (65–96) and (97–128), respectively.

random in the network, with k_{in1}, k_{in2} and k_{out} chosen so as to keep the total expected degree $k_{\text{in1}} + k_{\text{in2}} + k_{\text{out}} = 16$.

Feeding the network described above (see figure 2) to GN algorithm and the new algorithm respectively, we obtain the dendrograms as shown in figure 3. As we see, GN can only identify two compartments in the test, each of which contains two communities (see figure 3a), due to which it is very difficult to detect the communities within the compartments. In this case, it is clearly impossible to find four pre-defined communities from the dendrogram in figure 3a by using the un-weighted or weighted modularity, while the new algorithm can correctly reveal the four pre-defined communities. As shown in figure 3b,

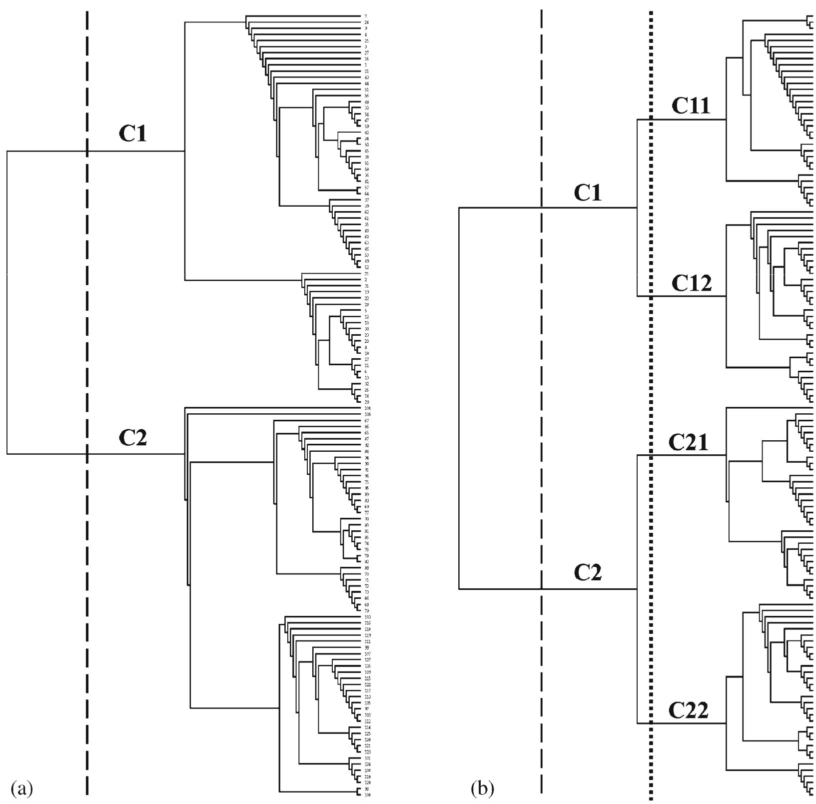


Figure 3. The dendrograms obtained in the network with $k_{\text{in1}} = 11$ and $k_{\text{in2}} = 4$ by (**a**) GN, (**b**) the new algorithm. The dashed line in (**a**) corresponds to the split into two compartments (C1 and C2) at the second level, obtained by the un-weighted modularity, where C1 and C2 contain vertices (1–64) and (65–128), respectively and the dashed line in (**b**) corresponds to the split into two compartments (C1 and C2) at the second level, obtained by the un-weighted modularity. The dotted line corresponds to the split into four communities (C11, C12, C21 and C22) at the first level, obtained by the weighted modularity, where C11, C12, C21 and C22 contain vertices (1–32), (33–64), (65–96) and (97–128), respectively.

the network is first split into two compartments, and then is divided into four communities. In this case, the community division of the four communities can be found, if the weighted modularity Q_w is used to select the community division from the dendrogram, while the un-weighted modularity Q can only see the two compartments due to its resolution limit problem (note that the optimization of un-weighted modularity will fail in the test too, due to the same reason). It is clear that, re-weighting the edge weight with the local similarities can help in improving the resolution of modularity.

3.3 *LFR network*

The Girvan–Newman benchmark is a class of classical and simple test networks for community detection, but it lacks some important properties in real networks. In [48], Lancichinetti, Fortunato and Rachicchi (LFR) proposed a class of more general community detection benchmarks, which have more realistic network properties such as power-law distributions of vertex degree and community size. The vertex degrees and community sizes of the benchmark network are determined by power-law distributions with exponents γ and β respectively, and the network contains a common mixing parameter μ that controls the ratio between the external degree of each vertex with respect to its community and the total degree of the vertex. The larger the value μ of a network is, the harder it is to detect communities in the network.

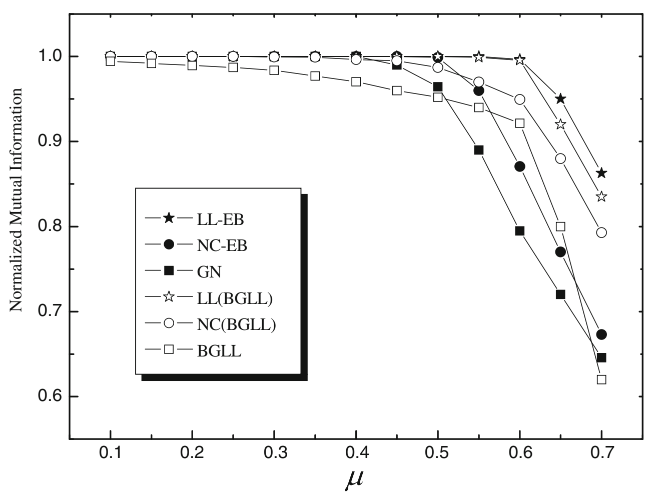


Figure 4. The normalized mutual information (NMI) between the real and the found communities in the networks described in the text. The six curves show results obtained by six algorithms in the networks – as a function of the mixing parameter μ. The LL-EB and NC-EB denote the new algorithms using the LLN and NC, respectively. BGLL denotes the modularity optimization algorithm introduced by Blondel *et al*. LL(BGLL) and NC(BGLL) denote the enhanced versions of BGLL by using LLN and NC to re-weight the edges in the networks, respectively. (The random walk length $t = 3$, and all the similarity measures are re-applied five times to re-weighting the edges in the networks.)

We create a set of undirected binary networks with 1000 vertices. The exponent of the degree distribution is −2, and the exponent of the community size distribution is −1. For other parameters, the default values provided by the algorithm are used. The mixing parameter varies from 0.1 to 0.7. With μ increasing, the communities in the networks become more and more difficult to identify. Figure 4 shows the performance comparison of five algorithms on the networks. As shown in figure 4, the results are similar to those in §3.1 – our new algorithms work better for higher value of μ and LLN clearly outperforms NC. On the same plot, we also show the results of the modularity optimization algorithm of Blondel *et al* [49] (BGLL) and its enhanced versions respectively by using LLN and NC to re-weight the edges in the networks. Similarly, the two enhanced-version algorithms outperform the original BGLL. More interestingly, the figure shows that our new algorithm with LLN clearly outperforms all the other algorithms in the test, including the BGLL and its enhanced versions.

3.4 *Zachary's karate club network*

The Zachary's karate network represents how the friendly relations between the 34 members in a karate club have split into two groups during Zachary's study, due to the disagreement between the administrator and the instructor of the club (see figure 5a) [46]. The network has become a well-known benchmark to test community detection algorithms. The new algorithm can correctly divide the club network into the two real-world groups in the first splitting (see the dashed line in figure 5b), while GN has one misclassified vertex, vertex 3, in the network. The community division of the network corresponding to the maximum of the weighted modularity is the same as in the unweighted case, consisting of five groups including one single vertex, vertex 10 (see the dotted line in figure 5b).

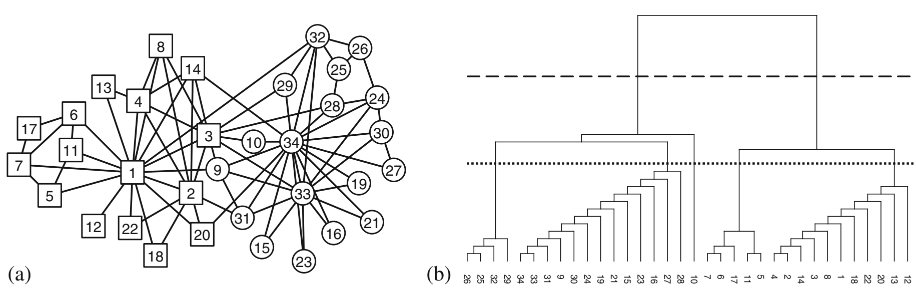
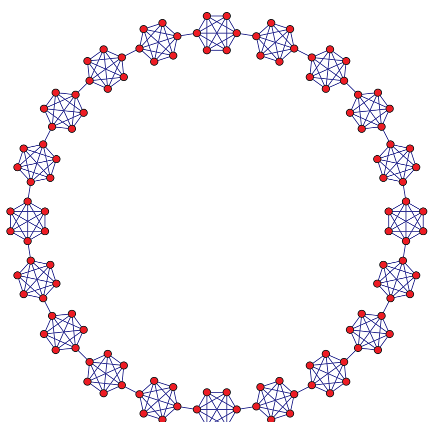


Figure 5. **(a)** The network of friendly relations between members in the karate club. The administrator and the instructor are represented by vertices 1 and 33, respectively. The squares represent individuals who aligned with the club's administrator after the split of the club, open circles represent those in the group of the instructor. **(b)** The dendrogram obtained by the new algorithm in the karate club network. The dashed line corresponds to the splitting into two communities that perfectly match the real-world split of the club, as denoted by the shapes of the vertices. The dotted line corresponds to the community division of the network containing five communities, according to the maximum of the weighted modularity.

4. Conclusions and discussions

In this paper, we studied the community detection algorithms using global and local structural information, where the edge betweenness and local similarity measures based respectively on local random walk dynamics and local cyclic structures were used. We tested the new algorithms on the Girvan–Newman benchmark networks and its generalization, the LFR benchmark networks, as well as the Zachary's karate club network. The results show that the new algorithms can work very well, and the local similarity measure based on local random walk dynamics (LLN) can give better results than based on local cyclic structures (NC). Interestingly, our results also show that the new algorithm with LLN can outperform the BGLL with LLN that has excellent performance [37].

In our methods, we have to first produce the output in the form of dendrogram (tree), and then we can select the final community division by modularity. In some sense, our methods can be regarded as a heuristic for finding the optimal modularity, if the optimum is consistent with the natural community division of the networks. But, they do not depend on modularity, and thus are not affected by the resolution limit problem. As we know, one can also make use of other quality functions or principles of communities, such as the local modularity [50] and the strong or weak definitions of communities [9], to determine the final output of the methods, and in some cases, they may be able to do better. Take the clique-ring network model as a simple example (shown in figure 6). Our methods will be able to produce a clear dendrogram in which all the cliques can be discovered. One



(a)

(b)

N_C	New	Mod
15	**15**	**15**
20	**20**	**20**
25	**25**	**25**
30	**30**	**30**
35	**35**	18
40	**40**	20
45	**45**	23
50	**50**	25
60	**60**	30

Figure 6. **(a)** An example of 6-vertex-clique-ring network model (the network was drawn by Pajek (http://vlado.fmf.uni-lj.si/pub/networks/pajek/)). **(b)** The number (N_c) of 6-vertices cliques in the test networks vs. the number of groups detected by different methods. Our new algorithms (New) can find out all the cliques in the networks and correctly recover the number N_c of the cliques by using the strong or weak definitions of communities [9]. The modularity-based algorithms (Mod) such as BGLL [49] can discover the cliques for small values of N_c, while they will fail for large values of N_c, because of the resolution limit problem of modularity. The conclusion of Mod can also be extended to the combination of our algorithms and modularity.

can easily obtain a reasonable community division from the dendrogram by the weak or strong definitions, while the modularity will fail, when the resolution limit problem of modularity appears in the networks of this type (see figure 6). Moreover, [51] recently show a phase transition in the spectral modularity optimization algorithms based on the eigenvectors of modularity matrix [51], which indicates a regime of detectability of such methods. This transition, which can be calculated theoretically in the stochastic block model, is of interest. The phenomena appear in the spectral methods, but may also exist in other community detection methods, such as the Bayesian maximum-likelihood method [36] and the community detection methods in this paper.

As we know, the vertices in the same communities are generally more similar, and thus the edges between them will have higher values of similarity measures. Redefining the weight of the edges with the similarity measures can help in increasing the difference between the intra- and intercommunity edges, making the community structure in the network clearer and easier to be detected. Many similarity measures are proposed and discussed in [40,45,46]. They will also be helpful to some extent for community detection. And, one can expect that any reasonable way to increase the difference between the intra- and intercommunity edges may be used to enhance the ability of the community detection. Further, one can take into account the combination of different community detection methods, by using the strategy of redefining the edge weight in networks. Finally, we hope that the study should help one to understand the community structure in the networks and to enrich the studies of community detection algorithms in the networks.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 11147121), the Scientific Research Fund of Education Department of Hunan Province of China (Grant No. 11B128), the project of Basic and Advanced Technology Research of Henan Province (Grant No. 112300410021), the Scientific Research Project of Huangshan University (Grant No. 2011xkj007), and partly by the Doctor Startup Project of Xiangtan University (Grant No. 10QDZ20).

References

[1] S Fortunato, *Phys. Rep.* **486**, 75 (2010)
[2] S Boccaletti, V Latora, Y Moreno, M Chavez and D U Hwang, *Phys. Rep.* **424**, 175 (2006)
[3] R Guimera and L A Nunes Amaral, *Nature* **433**, 895 (2005)
[4] E Ravasz, A L Somera, D A Mongru, Z N Oltvai and A L Barabasi, *Science* **297**, 1551 (2002)
[5] P Chen and S Redner, *J. Informetrics* **4**, 278 (2010)
[6] L Zemanová, G Zamora-López, C Zhou and J Kurths, *Pramana – J. Phys.* **70**, 1087 (2008)
[7] S Sinha and S Poria, *Pramana – J. Phys.* **77**, 833 (2011)
[8] R Pan and S Sinha, *Pramana – J. Phys.* **71**, 331 (2008)
[9] F Radicchi, C Castellano, F Cecconi, V Loreto and D Parisi, *Proc. Natl. Acad. Sci. USA* **101**, 2658 (2004)
[10] H Zhou, *Phys. Rev.* **E67**, 061901 (2003)
[11] Y Pan, D-H Li, J-G Liu and J-Z Liang, *Physica* **A389**, 2849 (2010)
[12] A Capocci, V D P Servedio, G Caldarelli and F Colaiori, *Physica* **A352**, 669 (2005)
[13] H-J Li, Y Wang, L-Y Wu, Z-P Liu, L Chen and X-S Zhang, *Europhys. Lett.* **97**, 48005 (2012)

[14] J Reichardt and S Bornholdt, *Phys. Rev. Lett.* **93**, 218701 (2004)
[15] M Rosvall and C T Bergstrom, *Proc. Natl. Acad. Sci. USA* **104**, 7327 (2007)
[16] F Wu and B A Huberman, *Eur. Phys. J.* **B38**, 331 (2004)
[17] A Medus, G Acuña and C O Dorso, *Physica* **A358**, 593 (2005)
[18] A Clauset, M E J Newman and C Moore, *Phys. Rev.* **E70**, 066111 (2004)
[19] J Duch and A Arenas, *Phys. Rev.* **E72**, 027104 (2005)
[20] J M Pujol, J Béjar and J Delgado, *Phys. Rev.* **E74**, 016107 (2006)
[21] K Wakita and T Tsurumi, Finding community structure in mega-scale social networks, in: *Proceedings of the 16th International Conference on World Wide Web*, ACM, Banff (Alberta, Canada, 2007) pp. 1275–1276
[22] X Wang, G Chen and H Lu, *Physica* **A384**, 667 (2007)
[23] L Donetti and M A Munoz, *J. Stat. Mech.: Theor. Exp.* **2004**, P10012 (2004)
[24] M Girvan and M E J Newman, *Proc. Natl. Acad. Sci. USA* **99**, 7821 (2002)
[25] L Danon, A Díaz-Guilera, J Duch and A Arenas, *J. Stat. Mech.: Theor. Exp.* **2005**, P09008 (2005)
[26] D Chen, Y Fu and M Shang, *Physica* **A388**, 2741 (2009)
[27] M E J Newman and M Girvan, *Phys. Rev.* **E69**, 026113 (2004)
[28] M E J Newman, *Phys. Rev.* **E70**, 056131 (2004)
[29] E A Leicht and M E J Newman, *Phys. Rev. Lett.* **100**, 118703 (2008)
[30] S Fortunato and M Barthélemy, *Proc. Natl. Acad. Sci. USA* **104**, 36 (2007)
[31] B H Good, Y-A de Montjoye and A Clauset, *Phys. Rev.* **E81**, 046106 (2010)
[32] X S Zhang, R S Wang, Y Wang, J Wang, Y Qiu, L Wang and L Chen, *Europhys. Lett.* **87**, 38002 (2009)
[33] A Arenas, A Fernández and S Gómez, *New J. Phys.* **10**, 053039 (2008)
[34] P Ronhovde and Z Nussinov, *Phys. Rev.* **E80**, 016109 (2009)
[35] V A Traag, P Van Dooren and Y Nesterov, *Phys. Rev.* **E84**, 016114 (2011)
[36] J M Hofman and C H Wiggins, *Phys. Rev. Lett.* **100**, 258701 (2008)
[37] D Lai, H Lu and C Nardini, *Phys. Rev.* **E81**, 066118 (2010)
[38] J W Berry, B Hendrickson, R A LaViolette and C A Phillips, *Phys. Rev.* **E83**, 056119 (2011)
[39] J Xiang and K Hu, *Physica A: Statistical Mechanics and its Applications* **391**, 4995 (2012)
[40] L Lü and T Zhou, *Physica A: Statistical Mechanics and its Applications* **390**, 1150 (2011)
[41] S Wasserman and K Faust, *Social network analysis* (Cambridge University Press, Cambridge, 1994)
[42] R K Ahuja, T L Magnanti and J B Orlin, *Network flows: Theory, algorithms, and applications* (Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA, 1993)
[43] V Latora and M Marchiori, *New J. Phys.* **9**, 188 (2007)
[44] S Fortunato, V Latora and M Marchiori, *Phys. Rev.* **E70**, 056104 (2004)
[45] E A Leicht, P Holme and M E J Newman, *Phys. Rev.* **E73**, 026120 (2006)
[46] W W Zachary, *J. Anthropol. Res.* **33**, 452 (1977)
[47] L C Freeman, *Sociometry* **40**, 35 (1977)
[48] A Lancichinetti, S Fortunato and F Radicchi, *Phys. Rev.* **E78**, 046110 (2008)
[49] V D Blondel, J-L Guillaume, R Lambiotte and E Lefebvre, *J. Stat. Mech.: Theor. Exp.* **2008**, P10008 (2008)
[50] S Muff, F Rao and A Caflisch, *Phys. Rev.* **E72**, 056107 (2005)
[51] R R Nadakuditi and M E J Newman, *Phys. Rev. Lett.* **108**, 188701 (2012)