The Tenth International Symposium on Operations Research and Its Applications (ISORA 2011)
Dunhuang, China, August 28–31, 2011

Multiple Resolution Community Structure Analysis

Shoujun Xu[1] Jiguang Wang[2] Junhua Zhang[2]
Xiang-Sun Zhang[2,*]

[1]School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, Gansu Province
[2]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190

Abstract Detecting community structure is fundamental for uncovering the links between topological structure and function in complex networks. A popular method now widely used relies on the optimization of a quantity called modularity. But modularity optimization has a fundamental drawback: there exists a resolution limit which depends on the total size of the network and on the degree of interconnectedness of the modules. Resultantly there are tendencies to merge smaller modules. Intuitively the considered module is relative to the local structure, and is not relative to the total size of the network. In this paper, firstly, we give an simple explanation from the theoretic point of view that why the resolution limit depends on the total size of the network which seems to violate the intuition. Secondly, in order to address the issue of the resolution limit we propose a quantitative function MQ_λ with a tuning parameter λ $(-1 \leqslant \lambda \leqslant 1/2)$ which contains modularity as a special case at $\lambda = 0$. In particular, $MQ_{\frac{1}{2}}$ optimization ensures the partition not to merge the smaller modules in any complex network. Numerical results show that the proposed method has been validated using synthetic networks, discovering the predefined structures at all scales and its applications to some real social networks can correctly identify the number of communities.

Keywords Complex network; Community structure; Modularity; Module; Multiple resolution

1 Introduction

The study of complex networks has received an enormous amount of attention from the scientific community in recently years. Networks consisting of nodes and links are an efficient way to represent and study a large variety of technological, biological and social complex systems [1, 20]. A property that seems to be common to many complex networks is *community structure* (or *modular structure*), the division of network nodes into groups (called *community* or *module*) within which the network connections are dense, but between which they are sparser. Detecting communities can be a way to identify substructures which could correspond to important functions. For example, communities may correspond to groups of Web pages of the World Wide Web dealing with related topics [7], to functional modules such as cycles and pathways in metabolic networks [9, 22], to groups of related individuals in social networks [10, 16], to compartments in food webs

*Corresponding to zxs@amt.ac.cn

[23, 12], and so on. Separating the network into such communities could simplify functional analysis considerably.

The problem of community detection is quite challenging and has been the subject of discussion in various disciplines, such as computer science, social science. Several methods and algorithms have been developed for community detection in the literature. A survey of different approaches can be found in Refs. [5, 19]. A decisive advance in community detection was made by Newman and Girvan [21], who introduced a quantitative measure for the quality of a partition of a network into communities, the modularity (or Q value). This measure essentially compares the number of links inside a given module with the expected value for a randomized graph (null model) of the same size and same degree sequence. If one chooses modularity as the relevant quality function, the problem of community detection becomes equivalent to modularity optimization, it is untractable for large networks due to the non-deterministic polynomial-time (NP)-hard nature of the problem [3]. Heuristics for modularity optimization has become the only feasible (in computational time) method to detect community structure up to now, including spectral bisection [24] from computer scientists, hierarchical clustering developed by sociologists [29], and those based on physical intuition [26] and information-theoretic measures [11, 28].

Modularity optimization seems to be an effective method to detect communities both in real and artificially generated networks. However, the modularity has been exposed to resolution limits by Fortunato and Barthelemy [8]. They claimed that modularity contains an intrinsic scale that depends on the total size of links in the network. Modules smaller than this scale may not be resolved even in the extreme case that they are complete graphs connected by single bridges. Resultantly there are tendencies to merge smaller modules into larger communities in the process of modularity optimization.

In this paper, we address the issue of community detection and propose a quantitative measure for evaluating the partition of the complex network into communities by adding an tuning parameter into the modularity. We call this measure modified modularity with parameter λ $(-1 \leq \lambda \leq 1/2)$, denoted by MQ_λ. Firstly, we give an simple explanation why the resolution limit of modularity optimization depends on the whole size of links. Secondly, we show theoretically that the proposed criterion improves the resolution limit in community detection. In addition to the above theoretical results, the proposed method allows us to explore the community structure at variable resolutions by tuning the parameter λ and we show that it is validated on some synthatic and real complex networks. Finally, in discussion we compare it with recent approaches for detecting community structure in complex networks.

2 Multiple resolution method

We start by briefly reviewing the concept of modularity. The *modularity*, also *Q value*, of a partition of a network proposed by Newman and Girvan [21] can be written as

$$Q = \sum_{s=1}^{m} [\frac{l_s}{L} - (\frac{d_s}{2L})^2], \tag{2.1}$$

where the sum is over the m modules of the partition, l_s is the number of links inside module s, L is the total number of links in the network, d_s is the total degree of nodes in module s. The first term of the summand in Eq. (2.1) is the fraction of links inside module s, the second term, in contrast, represents the expected fraction of links in that module, if links were located at random in the network (under the only constraint that the degree sequence coincides with the one of the original graph).

Fortunato and Barthelemy [8] pointed out that modularity optimization may favor network partitions, with groups of modules combined into larger communities. In other words, the modularity optimization has the tendency of changing the inter-community links into the intra-community links. In order to counteract the tendency, we introduce a tunable term with parameter λ into the original definition of modularity, called *modified modularity with parameter* λ, denoted by MQ_λ, as follows:

$$MQ_\lambda = \sum_{s=1}^{m} \frac{\frac{l_s}{L} - (\frac{d_s}{2L})^2}{l_s^\lambda}. \tag{2.2}$$

In particular, MQ_0 is equal to Q.

When the parameter $1 > \lambda > 0$, it acts as penalizing the intra-community links compared to the Q value, we have access to the substructure underneath those at $\lambda = 0$ in optimizing MQ_λ, thus the resolution limit existed in modularity optimization can be improved; while the parameter $\lambda < 0$, it acts as rewarding the intra-community links, we have access to the superstructure, in this case, we can address the issue of splitting the real communities into smaller communities in optimizing modularity, for example, modularity optimization probe the modular structure consisting in four communities in Zachary's karate club network, while this club was observed to split into two communities. The minimum value of λ corresponding to not merging smaller communities into larger communities in all networks in MQ_λ optimization, denoted by λ_{max}, has an upper boundary $\frac{1}{2}$ (for details see next section), i.e., $\lambda_{max} \leq 1/2$. the maximum value of λ corresponding to detecting modular structure consisting of two communities, denoted by λ_{min}, has also a lower boundary -1 (for details see next section). Therefore λ can take positive value and also negative value. In the following, we will only consider the parameter λ in the interval

$$(-1 \leq)\lambda_{min} \leq \lambda \leq \frac{1}{2}. \tag{2.3}$$

3 Theoretical results

The above-mentioned concept of community is qualitatively intuitive. Recently Radicchi *et al.* [25] gave a quantitative "weak" definition of community. If we consider a subgraph $V \subset G$, to which node i belongs, we can split node i's degree into two contributions: $k_i(V) = k_i^{in}(V) + k_i^{out}(V)$, $k_i^{in}(V)$ is the number of edges connecting node i to other nodes belonging to V. $k_i^{out}(V)$ is the number of connections toward nodes in the rest of the network. The subgraph V is called *community in a weak sense* if

$$\sum_{i \in V} k_i^{in}(V) > \sum_{i \in V} k_i^{out}(V), \tag{3.1}$$

i.e., in a *weak* community the sum of all degrees within V is larger than the sum of all degrees toward the rest of the network.

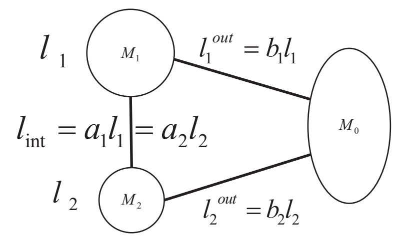

Fig. 1. Scheme of a network partition into three or more modules. The circles on the left represent two modules $\mathscr{M}_1$ and $\mathscr{M}_2$, the oval on the right represents the rest of the network $\mathscr{M}_0$, whose structure is arbitrary.

Theorem 1. *Any two modules satisfying weak definition can not be merged in $MQ_{\frac{1}{2}}$ optimization.*

Proof. We analyze a network with L links and with at least two modules, each of which satisfies the conditions of the weak definition. we focus on a pair of modules $\mathscr{M}_1$ and $\mathscr{M}_2$, and distinguish three types of links: those internal to each of the communities (l_1 and l_2, respectively), between $\mathscr{M}_1$ and $\mathscr{M}_2$ (l_{int}) and between the two communities and the rest of the network (l_1^{out} and l_2^{out}) (See Fig. 1). To simplify the calculations, we express the numbers of external links in terms of l_1 and l_2, so $l_{int} = a_1 l_1 = a_2 l_2, l_1^{out} = b_1 l_1$ and $l_2^{out} = b_2 l_2$, with $a_1, a_2, b_1, b_2 \geq 0$. Because $\mathscr{M}_1$ and $\mathscr{M}_2$ are communities in a weak sense, we also have $a_1 + b_1 \leq 2, a_2 + b_2 \leq 2$. We now consider two partitions A and B of the network. In partition A, $\mathscr{M}_1$ and $\mathscr{M}_2$ are taken as separate modules, and in partition B they are considered as a single community. The subdivision of the rest of the network, $\mathscr{M}_0$, is arbitrary but identical in both partitions. We want to prove MQ_λ^A of the partition A is more than MQ_λ^B of the partition B when $\lambda = \frac{1}{2}$. Because MQ_λ is a sum over the modules, the contribution of $\mathscr{M}_0$ is the same in both partitions and is denoted by M_0, From Eq. (2.2), we obtain

$$MQ_\lambda^A = \frac{\frac{l_1}{L} - (\frac{(a_1+b_1+2)l_1}{2L})^2}{l_1^\lambda} + \frac{\frac{l_2}{L} - (\frac{(a_2+b_2+2)l_2}{2L})^2}{l_2^\lambda} + M_0, \tag{3.2}$$

$$MQ_\lambda^B = \frac{\frac{l_1+l_2+a_1 l_1}{L} - (\frac{(a_1+b_1+2)l_1+(a_2+b_2+2)l_2}{2L})^2}{(l_1+l_2+a_1 l_1)^\lambda} + M_0. \tag{3.3}$$

The difference $\triangle MQ_\lambda = MQ_\lambda^A - MQ_\lambda^B$ is

$$\begin{aligned}\triangle MQ_\lambda =& 4L(l_1^{(1-\lambda)} + l_2^{(1-\lambda)} - (l_1 + l_2 + a_1 l_1)^{(1-\lambda)}) - (a_1 + b_1 + 2)^2 l_1^{(2-\lambda)} \\ &- (a_2 + b_2 + 2)^2 l_2^{(2-\lambda)} + \frac{((a_1 + b_1 + 2)l_1 + (a_2 + b_2 + 2)l_2)^2}{(l_1 + l_2 + a_1 l_1)^\lambda}.\end{aligned} \tag{3.4}$$

Because $\mathscr{M}_1$ and $\mathscr{M}_2$ are both modules by construction, we expect a larger MQ_λ for the partition where the two modules are separated, i.e., $MQ_\lambda^A > MQ_\lambda^B$, which in turn implies

$\triangle MQ_\lambda > 0$. From Eq. (3.4), we see that $\triangle MQ_\lambda$ is positive only if

$$4L(l_1^{(1-\lambda)}+l_2^{(1-\lambda)}-(l_1+l_2+a_1l_1)^{(1-\lambda)}) > (a_1+b_1+2)^2l_1^{(2-\lambda)}+(a_2+b_2+2)^2l_2^{(2-\lambda)} - \frac{((a_1+b_1+2)l_1+(a_2+b_2+2)l_2)^2}{(l_1+l_2+a_1l_1)^\lambda}. \quad (3.5)$$

If

$$l_1^{(1-\lambda)}+l_2^{(1-\lambda)}-(l_1+l_2+a_1l_1)^{(1-\lambda)} < 0, \quad (3.6)$$

the inequality (3.5) is not always satisfied with the increasing of L. Therefore, in order to prove the theorem, we must disprove the inequality (3.6) and prove the inequality (3.5) when $\lambda = 1/2$.

Firstly, we disprove (3.6) in the case $\lambda = 1/2$, i.e.,

$$\sqrt{l_1}+\sqrt{l_2}-\sqrt{l_1+l_2+a_1l_1} \geq 0, \quad (3.7)$$

it is sufficient to prove that $2\sqrt{l_1l_2} \geq a_1l_1$ i.e., note that $a_1l_1 = a_2l_2$, $4l_1l_2 \geq (a_1l_1)^2 = a_1a_2l_1l_2$, since $a_1 \leq 2, a_2 \leq 2$, it holds. Further we disprove (3.6).

Secondly, combining the inequality (3.7), we prove (3.5) in the case $\lambda = 1/2$, it is sufficient to show (note that $L \geq l_1+l_2+a_1l_1$)

$$\begin{aligned}&4(l_1+l_2+a_1l_1)(\sqrt{l_1}+\sqrt{l_2}-\sqrt{l_1+l_2+a_1l_1})\\ &\geq (a_1+b_1+2)^2l_1^{\frac{3}{2}}+(a_2+b_2+2)^2l_2^{\frac{3}{2}}-\frac{((a_1+b_1+2)l_1+(a_2+b_2+2)l_2)^2}{\sqrt{l_1+l_2+a_1l_1}}. \quad (3.8)\end{aligned}$$

Substituting $(a_1/a_2)l_1$ for l_2 in the inequality (3.8), we obtain

$$\begin{aligned}&4(a_1+a_2+a_1a_2)^{\frac{3}{2}}(\sqrt{a_1}+\sqrt{a_2}-\sqrt{a_1+a_2+a_1a_2})\\ &\geq (a_1+a_2+a_1a_2)^{\frac{1}{2}}(a_1+b_1+2)^2a_2^{\frac{3}{2}}+(a_1+a_2+a_1a_2)^{\frac{1}{2}}(a_2+b_2+2)^2a_1^{\frac{3}{2}}\\ &-((a_1+b_1+2)a_2+(a_2+b_2+2)a_1)^2. \quad (3.9)\end{aligned}$$

Let $a = a_1+b_1+2$ and $b = a_2+b_2+2$, then $a_1+2 \leq a \leq 4$ and $a_2+2 \leq b \leq 4$, and substituting them into (3.9), we obtain

$$\begin{aligned}&4(a_1+a_2+a_1a_2)^{\frac{3}{2}}(\sqrt{a_1}+\sqrt{a_2}-\sqrt{a_1+a_2+a_1a_2})\\ &\geq (a_1+a_2+a_1a_2)^{\frac{1}{2}}a^2a_2^{\frac{3}{2}}+(a_1+a_2+a_1a_2)^{\frac{1}{2}}b^2a_1^{\frac{3}{2}}-(aa_2+ba_1)^2. \quad (3.10)\end{aligned}$$

The right-hand side of the inequality (3.10) is a quartic function of positive first-termed coefficient with parameters a_1 and a_2 in terms of variables a and b, denoted by $h(a,b)$, so $h(a,b)$ would reach the maximum value only at the nodes of the fields, i.e.,

$$h(a,b) \leq \max\{h(2+a_1,4), h(2+a_1,2+a_2), h(4,2+a_2), h(4,4)\}. \quad (3.11)$$

In order to prove (3.10), it is sufficient to prove that

$$\begin{aligned}&4(a_1+a_2+a_1a_2)^{\frac{3}{2}}(\sqrt{a_1}+\sqrt{a_2}-\sqrt{a_1+a_2+a_1a_2})\\ &\geq \max\{h(2+a_1,4), h(2+a_1,2+a_2), h(4,2+a_2), h(4,4)\}. \quad (3.12)\end{aligned}$$

We only need prove the inequality holds for each term in right-hand side of (3.12), i.e., (note that by the symmetry of $h(a_1+2,4)$ and $h(4,a_2+2)$, we need prove one of them.) we need prove the following three inequalities, respectively,

$$\begin{aligned}
&4(a_1+a_2+a_1a_2)^{\frac{3}{2}}(\sqrt{a_1}+\sqrt{a_2}-\sqrt{a_1+a_2+a_1a_2})\\
&\geq h(2+a_1,4)\\
&= a_2^{\frac{3}{2}}\sqrt{a_1+a_2+a_1a_2}(2+a_1)^2+16a_1^{\frac{3}{2}}\sqrt{a_1+a_2+a_1a_2}\\
&\quad -(a_2(a_1+2)+4a_1)^2, \qquad (3.13)\\
&4(a_1+a_2+a_1a_2)^{\frac{3}{2}}(\sqrt{a_1}+\sqrt{a_2}-\sqrt{a_1+a_2+a_1a_2})\\
&\geq h(2+a_1,2+a_2)\\
&= (a_1+a_2+a_1a_2)^{\frac{1}{2}}(a_1+2)^2a_2^{\frac{3}{2}}+(a_1+a_2+a_1a_2)^{\frac{1}{2}}(a_2+2)^2a_1^{\frac{3}{2}}\\
&\quad -((a_1+2)a_2+(a_2+2)a_1)^2, \qquad (3.14)\\
&4(a_1+a_2+a_1a_2)^{\frac{3}{2}}(\sqrt{a_1}+\sqrt{a_2}-\sqrt{a_1+a_2+a_1a_2})\\
&\geq h(4,4)\\
&= 16(a_1^{\frac{3}{2}}+a_2^{\frac{3}{2}})\sqrt{a_1+a_2+a_1a_2}-16(a_1+a_2)^2. \qquad (3.15)
\end{aligned}$$

First of all, we prove the inequality (3.13). After tiding up, we obtain

$$\begin{aligned}
&\sqrt{a_2}\sqrt{a_1+a_2+a_1a_2}a_1(4-a_1a_2)+4\sqrt{a_2}\sqrt{a_1+a_2+a_1a_2}(a_2+a_1a_2-3a_1)\\
&\geq a_1(a_2-2)(6a_1+3a_1a_2+4a_2). \qquad (3.16)
\end{aligned}$$

Obviously, the first term of the left-hand side of the inequality (3.16) is nonnegative and the right-hand side of it is not positive. Under the condition that $a_2 \geq 3a_1/(1+a_1)$, the second term of the left-hand side of (3.16) is nonnegative, so (3.16) holds. So we need prove (3.16) in the case $0 \leq a_2 \leq 3a_1/(1+a_1)$, it is sufficient to prove

$$4\sqrt{a_2}\sqrt{a_1+a_2+a_1a_2}(a_2+a_1a_2-3a_1) \geq a_1(a_2-2)(6a_1+3a_1a_2+4a_2). \qquad (3.17)$$

Since both sides of the inequality (3.17) are all negative, it is sufficient to prove that the difference of the square of the left-hand side and one of the right-hand side, denoted by $f(a1,a2)$ associated with two parameters a_1 and a_2, is less than 0. Since the equations

$$\begin{cases} \frac{\partial f(a_1,a_2)}{\partial a_1} = 0 \\ \frac{\partial f(a_1,a_2)}{\partial a_2} = 0 \end{cases} \qquad (3.18)$$

has no zero in the fields of $0 < a_1 < 2, 0 < a_2 < 3a_1/(1+a_1)$, so $f(a_1,a_2)$ would reach optimal values in the boundary of the fields. Since $f(a_1,0) = 0, f(2,a_2) = 16(a_2-2)^2a_2(25a_2+6) \geq 0, f(a_1,3a_1/(1+a_1)) = 9a_1^4(12+4a_1-5a_1^2)/(a_1+1)^4 \geq 0$, $f(a_1,a_2)$ is nonnegative in the fields. So we accomplish the proof of (3.13).

Secondly, we prove the inequality (3.14), it is relevantly simple. Tiding up (3.14), we obtain

$$\sqrt{a_1}\sqrt{a_1+a_2+a_1a_2}a_2(4-a_1a_2)+\sqrt{a_2}\sqrt{a_1+a_2+a_1a_2}a_1(4-a_1a_2) \geq 0. \qquad (3.19)$$

Obviously, under the conditions $0 \le a_1 \le 2, 0 \le a_2 \le 2$, the inequality (3.19) holds, so we complete the proof of (3.14).

Finally, we prove the inequality (3.15). Tiding up, we get

$$\begin{aligned}&(a_1+a_2+a_1a_2)^{\frac{1}{2}}(\sqrt{a_1}+\sqrt{a_2})(-3a_1-3a_2+a_1a_2+4\sqrt{a_1a_2})\\ &\ge (3a_1+3a_2+a_1a_2)(a_1a_2-a_1-a_2).\end{aligned} \tag{3.20}$$

Since $(2-a_1)(2-a_2) \le 4$, $a_1a_2-a_1-a_2 \le 0$, thus

$$-3a_1-3a_2+a_1a_2+4\sqrt{a_1a_2} = a_1a_2-a_1-a_2-2(\sqrt{a_1}-\sqrt{a_2})^2 \le a_1a_2-a_1-a_2 \le 0. \tag{3.21}$$

So both sides of Eq. (3.20) are negative.

It is sufficient to prove that the difference of the square of the right-hand side of the inequality (3.20) from the square of the left-hand side of (3.20) is more than 0. i.e., (if let $c = a_1+a_2, d = \sqrt{a_1a_2}$, then $0 \le c \le 4, 0 \le b \le c/2$, substituting them)

$$g(c,d) = (3c+d^2)^2(d^2-c)^2-(c+d^2)(c+2d)(-3c+d^2+4d)^2 \ge 0. \tag{3.22}$$

Since

$$\begin{aligned}&g(4,d) = (d-2)^2d(d^3+6d^2+20d+24) \ge 0\\ &g(c,0) = 0\\ &g(c,\frac{c}{2}) = \frac{1}{256}(c-4)^2c^4(c^2+16c+112) \ge 0\end{aligned}$$

and the equations

$$\begin{cases}\frac{\partial g(c,d)}{\partial c} = 0\\ \frac{\partial g(c,d)}{\partial d} = 0\end{cases} \tag{3.23}$$

have only zero $c \doteq 3.5007, d \doteq 0.820443$ in the fields of $0 \le c \le 4, 0 \le d \le c/2$, and its value is equal to 78.5776, so $g(c,d) \ge 0$. So we complete the proof of the inequality (3.15). Thus, we complete all proofs of the inequality (3.5). So the result holds. □

Corollary 2. *In circle-like clique networks, $MQ_{\frac{1}{2}}$ optimization discovers the predefined modular structures in the networks.*

Here $1/2$ is the upper boundary of λ that ensures that there exists no community emerging several communities in $MQ_{\frac{1}{2}}$ optimization, but in fact, in experiments it is s-maller than $1/2$. For example, in circle-like clique networks, on matter how change about the size m of cliques and the number of cliques, when $\lambda = \lg_2(1+\frac{1}{m(m-1)}) < \frac{1}{m(m-1)\ln 2} \doteq \frac{1}{0.69m(m-1)}$, for example, $m = 10$, $\lambda \doteq 0.016.$, MQ_λ optimization detects the community structure as expected. (For proof, see Appendix A.)

From the process of the proof of Theorem 1, we can also give a simple explanation why the resolution limit depends on the total size of links: $\mathscr{M}_1$ and $\mathscr{M}_2$ are two communities in the weak sense. In MQ_λ optimization, in order to avoid to merge $\mathscr{M}_1$ and $\mathscr{M}_2$

as a larger community, inequality (3.5) must hold, i.e., MQ_λ^A for the partition where $\mathscr{M}_1$ and $\mathscr{M}_2$ are separated is larger than MQ_λ^B for the partition where $\mathscr{M}_1$ and $\mathscr{M}_2$ are merged. But in the case $\lambda = 0$ there exists a large number L, the total size of the network, such that inequality (3.5) does not hold since the inequality (3.6) holds. So the existence of the resolution limit in Q (MQ_0) optimization is obvious.

In the following we prove

Theorem 3. *A clique of order n can not be split in $MQ_{\frac{1}{2}}$ optimization.*

Proof. We need prove

$$\frac{4L\frac{n_1(n_1-1)}{2}-(n_1(n_1-1)+n_1n_2)^2}{4L^2\sqrt{\frac{n_1(n_1-1)}{2}}}+\frac{4L\frac{n_2(n_2-1)}{2}-(n_2(n_2-1)+n_1n_2)^2}{4L^2\sqrt{\frac{n_2(n_2-1)}{2}}}$$
$$\leq \frac{4L\frac{(n_1+n_2)(n_1+n_2-1)}{2}-((n_1+n_2)(n_1+n_2-1))^2}{4L^2\sqrt{\frac{(n_1+n_2)(n_1+n_2-1)}{2}}},$$

where n_1 and n_2 are positive integers such that $n_1+n_2=n$. Since

$$4L\sqrt{\frac{n_1(n_1-1)}{2}}+4L\sqrt{\frac{n_2(n_2-1)}{2}}\leq 4L\sqrt{\frac{(n_1+n_2)(n_1+n_2-1)}{2}},$$

we only need prove

$$\frac{(n_1(n_1-1)+n_1n_2)^2}{\sqrt{\frac{n_1(n_1-1)}{2}}}+\frac{(n_2(n_2-1)+n_1n_2)^2}{\sqrt{\frac{n_2(n_2-1)}{2}}}-\frac{((n_1+n_2)(n_1+n_2-1))^2}{\sqrt{\frac{(n_1+n_2)(n_1+n_2-1)}{2}}}>0.$$

Using the Software Mathematics 5.2, we easily prove the above inequality for any n_1,n_2.

□

In order to find the lower boundary of λ, we give another expression on the contribution of a special module M to Q value. We distinguish the links into 3 classes in terms of module M: internal links l_i, links l_o joining nodes of M to the rest of the network and rest links l_r (see Fig. 2).

Lemma 4. *The contribution $Q(M)$ of Module M to Q value is* $\frac{4l_il_r-l_o^2}{4L^2}$.

Proof.

$$\begin{aligned}Q(M)&=\frac{4Ll_i-(2l_i+l_o)^2}{4L^2}\\&=\frac{4(l_i+l_o+l_r)l_i-(2l_i+l_o)^2}{4L^2}\\&=\frac{4l_il_r-l_o^2}{4L^2}.\end{aligned}$$

□

Theorem 5. *$MQ_{-1-\varepsilon}$ optimization detect a partition consisting of two modules, ε is infinitesimal.*

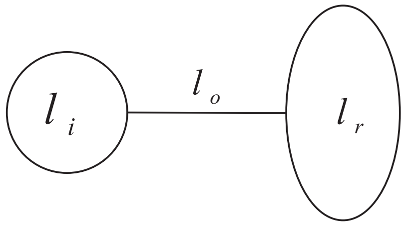

Fig. 2. Scheme for single module with the number l_i of the internal links, the number l_o of the external links connecting it to the rest network, and the number l_r of the rest links.

Proof. In $MQ_{-1-\varepsilon}$ optimization, we assume on the contrary that we obtain a partition consisting of communities $M_1, M_2, \cdots, M_k (k \geq 3)$. There exists a community, say M_1, such that the contribution of M_1 to Q value is positive, otherwise $MQ_{-1-\varepsilon}$ is non-positive, obviously detecting modular structure in such a network is not meaningful. We choose the community M_1 such that the contribution is maximum. Construct a new community by merging communities $M_2, M_3, \cdots, M_k$ into one group, denoted by $M^{'}$, then a new partition $\{M_1, M^{'}\}$ is constructed. Let l_{ij}, l_{oj}, l_{rj} $(1 \leq j \leq k)$ associated with the module M_j correspond to internal links, links joining between nodes of M_j and the rest of the network and rest links l_r. Then from Lemma 4 we have

$$\begin{aligned}
& MQ_{-1-\varepsilon}(M_1, M_2, \cdots, M_k) \\
= & \frac{4l_{i1}l_{r1} - l_{o1}^2}{4L^2} l_{i1}^{1+\varepsilon} + \frac{4l_{i2}l_{r2} - l_{o2}^2}{4L^2} l_{i2}^{1+\varepsilon} + \cdots + \frac{4l_{ik}l_{rk} - l_{ok}^2}{4L^2} l_{ik}^{1+\varepsilon} \\
\leq & \frac{4l_{i1}l_{r1} - l_{o1}^2}{4L^2} l_{i1}^{1+\varepsilon} + \frac{4l_{i1}l_{r1} - l_{o1}^2}{4L^2} (l_{i2}^{1+\varepsilon} + \cdots + l_{ik}^{1+\varepsilon}) \qquad \text{(by the choice of module } M_1\text{)} \\
< & \frac{4l_{i1}l_{r1} - l_{o1}^2}{4L^2} l_{i1}^{1+\varepsilon} + \frac{4l_{i1}l_{r1} - l_{o1}^2}{4L^2} (l_{i2} + \cdots + l_{ik})^{1+\varepsilon} \\
\leq & MQ_{-1-\varepsilon}(M_1, M^{'}),
\end{aligned}$$

which contradicts the choice of the partition $\{M_1, M_2, \cdots, M_k\}$. So $k \leqslant 2$. If we obtain a unique community formed by the whole network, in which case $MQ_{-1-\varepsilon}$ is equal to 0, obviousely detecting modular structure in such a network discussed is not meaningful. Therefore the theorem holds. □

4 Experimental results

In this section, we conduct experiments on both artificial networks and well-studied real networks. We implement the MQ_λ optimization by using simulated annealing and we adopt the same recipe introduced in [9], which makes the optimization procedure very effective.

4.1 Artificial networks

Let us first check our algorithm by applying it to two artificial test networks with known community structure.

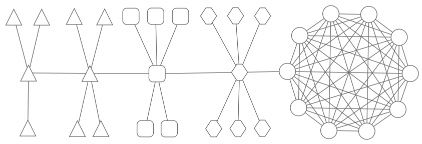

Fig. 3. Toy model network with communities of different sizes and densities. Modularity optimization detect the modular structure represented by different shapes of nodes.

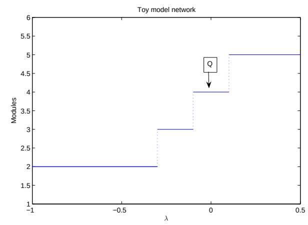

Fig. 4. Number of modules as a function of the parameter λ using our method on toy model network.

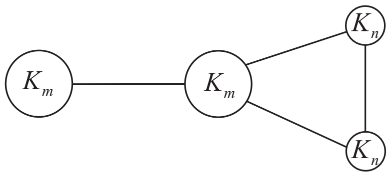

Fig. 5. 4-clique network with two cliques with m nodes and the other two cliques with n nodes.

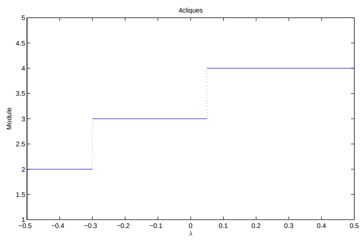

Fig. 6. Number of modules as a function of the parameter λ using our method on 4-clique network.

Toy model network. The first example is the toy model network analyzed by A. Arenas et al. [2]. It has a simple topology, but it is difficult for community-detection algorithms because it includes communities of different sizes, some of them sparse and others dense, see Fig. 3. When $\lambda \geq 0.1$, our method detects the community structure consisting of 5 modules as expected from the structure. But modularity optimization detects a community structure consisting of 4 modules as shown by the different shapes of nodes in Fig. 3. If $-0.2 \leq \lambda \leq -0.1$, we obtain a partition consisting of 3 communities, and if $-0.3 \leq \lambda$, the number of communities is 2. The number of modules as a function of the parameter λ is shown in Fig. 3.

4-clique network. Another example is shown in Fig. 5, where the circles again represent cliques (i.e., complete graphs): the two on the left have m nodes each, the other two have $n < m$ nodes. When $\lambda \geq 0.05$, if we take $m = 20$ and $n = 5$, the method can correctly split the network into 4 classes. When $-0.3 \leq \lambda \leq 0.05$, the number of communities is 3, especially, in the case $\lambda = 0$ MQ_0 optimization (modularity optimization) detects the community structure of 3 modules consisting of the community merging two cliques with 5 nodes and two cliques with 20 nodes as communities. $\lambda \leq -0.3$, the number of communities is 2.

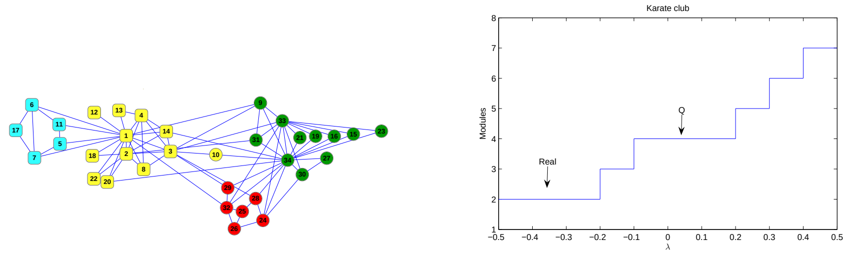

Fig. 7. Left: Zachary's karate club network. The administrator and the instructor are represented by nodes 1 and 33 respectively. Square nodes and circle nodes represent the administrator's faction and the instructor's faction, respectively. The partition consisting of distinct color nodes is obtained by modularity optimization. Right: The number of modules in the optimal partitions at different scales.

4.2 Real-world networks

Although artificial networks provide a reproducible and well controlled test for our community-structure algorithm, it is clearly desirable to test the algorithm on data from real-world networks as well. To this end, we have selected two datasets representing real-world networks for which the community structure is already known from other sources. Now we do the test on real networks.

Zachary's karate club network. The first network is drawn from the well known karate club study of Zachary [30]. In this study, Zachary observed 34 members of a karate club at a US university over a period of 2 years. During the course of the study, a disagreement over whether to raise club fees developed between the administrator (node 1) of the club and the club's instructor (node 33), which ultimately resulted in the instructor's leaving there and starting a new club, taking about a half of the original club's members with him. See Fig. 7 for separations represented by different shapes of nodes.

By using our method, when $\lambda = -0.2$, the network was partitioned into two communities exactly consistent with real partition except the node 10, in fact node 10 belonging to any community is reasonable from the topology structure of the network (see Fig. 7). However, maximizing the Q value, we obtained the "optimal" partition consisting of 4 communities, which seems to be reasonable from the topology of the network.

Journal index network. The journal index network constructed by Rosvall and Bergstrom [28] consists of 40 journals as nodes from 4 different fields: physics, chemistry, biology, and ecology and 189 links connecting nodes if at least one article from one journal cities an article in the other journal during 2004. Ten journals with the highest impact factor in the 4 different fields were selected. Using our method, we can partition the network into 4 communities correctly (Fig. 8).

We can also partition the network into two, three, or five modules. When we partition the network into two components, physical journals cluster together with chemical journals and biological journals cluster together with ecological journals. When we split it into three component, ecological journals and biological journals separate, but physical journals and chemical journals remain together in a single module. When we intend to

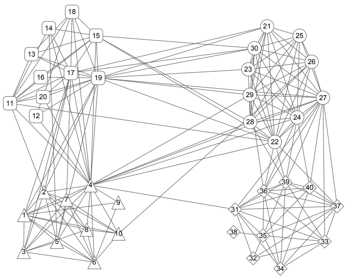

Fig. 8. Journal index network with 4 field journals represented by different shapes of nodes.

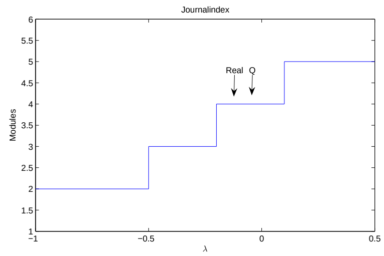

Fig. 9. Number of communities in the partition obtained from MQ_λ optimization on function of λ

split the network into five modules, we get essentially the same partition as with four, only with the singly connected journal Conservation Biology split off by itself as a community. This result is consistent with that in [28]. One might not even consider that singleton to be a valid module.

MQ_λ optimization can detect all possible partitions as above mentioned, for details and the number of communities on the function of λ see Fig. 9.

5 Discussion

Considered that modularity optimization contains an intrinsic scale that depends on the total size of links in the network and communities smaller than this scale may not be resolved, recently some multi-resolution methods have been proposed, one of which is the Hamiltonian-based method introduced by Reichardt and Bornholdt (RB)[26], this method contains a tuning parameter $\gamma\,(\gamma \geqslant 0)$ which can be used to study communities of different sizes, in fact, the parameter γ is to tune the null model in Q value (i.e., γ is the prefactor that multiplies the null model). Arenas et al. (AFG) proposed another multiresolution method [2] i.e., modifying the modularity by adding an additional self-link of weight r $(r \geqslant 0)$ to each node as a new quality function of the partitions which can be used to probe the community structure at different resolutions by tuning r. The two methods are all contained modularity as a special case, $\gamma = 1$ and $r = 0$, respectively.

RB and AFG methods are somewhat related, although not equal. The tuning parameter, γ and r, behave qualitatively in the same way: small values yield large communities, and large parameter values allow finding small communities. The number of communities in RB method changes from 1 (i.e., the case the whole network constitutes one unique community) to the size of the network (i.e., the case each node constitutes one community). And the number of communities in AFG method changes from the same value as Q optimization method to the size of the network. Obviously, AFG method fail to detect community structure in some networks in which the number of physical communities is less than the number of communities obtained by the optimization of Q value no matter how the tuning parameter change, for example in Zachary's karate club network, modularity optimization detects 4 communities, while the number of physical communities is 2. In our method the number of communities change from 2 to some value in that case any community can not be obtained by merging smaller communities while the parameter λ increase gradually.

On the other hand, in the RB method, the effect of γ can be interpreted as the "effective" number of links in the network equals L/γ, which is not equivalent to tuning the resolution of modularity optimization, RB method has also a resolution limit due to similar underlying reasons [13]. In the AFG method, the effect of r is to add a self-link of weight r (or r self-loops if r is integer) for every node in the original network, in fact the whole topology structure of network are changed although the new network presents the same characteristics as the original network in terms of connectivity. In our method, the effect of λ is adjusting the ratio of the importance between inter-community edges and intra-community edges, i.e., contribution of inter-community edges is more or less important than contribution of intra-community edges to MQ_λ by tuning parameter λ, but the whole topology structure of the network is not changed.

6 Conclusions

In conclusion, motivated by the recent finding that the optimization of modularity has a resolution limit, we propose a multiple resolution prodedure containing a tuning parameter. The main idea consists of adjusting the importance between intra-community edges and inter-community edges other than the topology structure of the network, which allows the search of modules at different topology scales. We have provided examples of the modular substructure found in synthetic and real complex networks. The results are sets of partitions, in which the number of modules screen from 2 to the maximum reasonable number, i.e., in this case there is no community which is obtained by merging some smaller modules.

We have also provided some theoretical results. Firstly, we easily showed why the resolution limit of modularity optimization depends on the total size of the network by the explicit expression. Secondly, we also prove any clique can not be separated into two or more modules whatever the tuning parameter λ are changed.

From the theoretical view, when the parameter λ range from -1 to $1/2$, by optimizing MQ_λ partitions vary from one with 2 communities to one in which any community cannot be obtained by emerging two modules satisfying the weak definition. But in test networks, we can accomplish this aim in very small subinterval of $[-1, 1/2]$.

Appendix A: the upper boundary of λ in circle-like clique network

In Fig 10, we show a network consisting of a ring of n cliques, connected through single links. Each clique is a complete graph K_m with m nodes and has $m(m-1)/2$ links.

The network has a clear modular structure where the communities correspond to single cliques, and we expect that any detection algorithm should be able to detect these communities. The modified modularity MQ_λ^{single} of this natural partition can be easily calculated and is equal to

$$MQ_\lambda^{single} = \frac{1 - \frac{2}{m(m-1)+2} - \frac{1}{n}}{\left(\frac{m(m-1)}{2}\right)\lambda}.$$

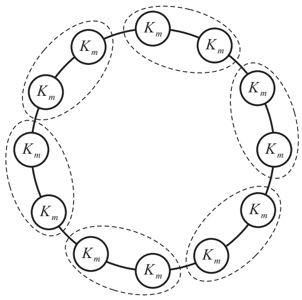

Fig. 10. The circle-like clique graph. Each module is a clique with m nodes, and two adjacent modules are connected by one edge.

On the other hand, the modified modularity MQ_λ^{pairs} of the partition in which pairs of consecutive cliques are considered as single communities (as shown by the dotted lines in Fig. 10) is

$$MQ_\lambda^{pairs} = \frac{1 - \frac{1}{m(m-1)+2} - \frac{2}{n}}{(m(m-1))^\lambda}.$$

The condition $MQ_\lambda^{single} > MQ_\lambda^{pairs}$, i.e., $\frac{MQ_\lambda^{single}}{MQ_\lambda^{pairs}} > 1$ is satisfied only if

$$2^\lambda > \frac{1 - \frac{1}{m(m-1)+2} - \frac{2}{n}}{1 - \frac{2}{m(m-1)+2} - \frac{1}{n}},$$

it need

$$\lambda = \lg_2 \max_{m,n} \frac{1 - \frac{1}{m(m-1)+2} - \frac{2}{n}}{1 - \frac{2}{m(m-1)+2} - \frac{1}{n}} < \lg_2(1 + \frac{1}{m(m-1)}) < \frac{1}{m(m-1)ln2}.$$

For example, when $m = 10$, $\lambda = 0.016$.

Acknowledgements

This work is partially supported by the National Science Foundation for Post-doctoral Scientists of China(grant NO. 20080440071), the Fundamental Research Funds for the Central Universities (grant NO. lzujbky-2009-51) and NSFC (Grant NO. 11001113).

References

[1] Albert, R. and Barabási, A.L., Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74(1)** (2002) 47.

[2] Arenas A., Fernandez A. and Gomez S., Analysis of the structure of complex networks at different resolution levels, *New J. Phys.* **10** (2008) 053039.

[3] Branes U., Delling, D., Gaertler M., Gorke, Hoefer, M., Nikoloski, Z. and Wagner D., On modularity clustering, *IEEE transactions on knowledge and data engineering* **20(2)** (2008) 172.

[4] Clauset A., Newman M. E. J. and Moore C., Finding community structure in very large networks, *phys. Rev. E* **70** (2004) 006611.

[5] Danon L., Diaz-Guilera A., Duch J. and Arenas A., Comparing community structure identification, *J. Stat. Meth.:Theorem and Experiment* **9** (2005) P09008.

[6] Duch J. and Arenas A., Community detection in complex networks using extremal optimization, *phys. Rev. E* **72** (2005) 027104.

[7] Flake G.W., Lawrence S., Lee Giles C. and Coetzee F.M., Self-organization and identification of Web communities, *IEEE Computer* **35** (2002) 66.

[8] Fortunato S. and Barthélemy M., Resolution limit in community detection, *Proc. Natl Acad. Sci. USA* **104** (2007) 36.

[9] Guimera R. and Amaral Nunes L. A., Functional cartography of complex metabolic networks, *Nature* **433** (2005) 895.

[10] Girvan M. and Newman M., Community strucuture in social and biological networks, *Proc. Natl Acad. Sci. USA* **99** (2002) 7821.

[11] Hofman J. M. and Wiggins C. H., Bayesian approach to network modularity, *Phys. Rev. Lett.* **100** (2008) 258701.

[12] Krause A. E., Frank K. A., Mason D. M., Ulanowicz R. E., Taylor W. W., Compartments exposed in food-web structure. *Nature* **426** (2003) 282.

[13] Kumpula, J. M., Saramaki, J., Kaski, K. and Kertesz, J., Limited resolution in complex network community detection with potts model approach, *Eur. Phys. J. B* **56** (2007) 41.

[14] Leicht E. A. and Newman M. E. J., Community structure in directed networks, *phys. Rev. Lett.* **100** (2008) 118703.

[15] Lusseau D., The emergent properties of a dolphin social network, *Proc. R. Soc. London B (suppl.)* **270** (2003) S186.

[16] Lusseau D. and Newman M. E. J., Identifying the role that animals play in their social networks, *Proceedings of the Royal Society of Landon B* **271** (2004) S477.

[17] Lusseau D., Schneider K., Boisseau O. J., Haase P., Slooten E., and Dawson S. M., The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait? *Behavioral Ecology and Sociobiology* **54** (2003) 396.

[18] Muff S., Rao F. and Caflisch A., Local modularity measure foe network clusterizations, *phys. Rev. E.* **72** (2005) 056107.

[19] Newman M. E. J., Detecting community structure in networks, *Eur. Phys. J. B* **38** (2004) 321.

[20] Newman M. E. J., The structure and function of complex networks, *SIAM Review* **45(2)** (2003) 167.

[21] Newman M. E. J. and Girvan M., Finding and evaluating community structure in networks, *Phys. Rev. E.* **69** (2004) 026113.

[22] Palla G., Dere'nyi I., Farkas K., Vicsek T., Uncovering the overlapping community structure of complex networks in nature and society, *Nature* **435** (2005) 814.

[23] Pimm S. L., The structure of food webs, *Theoretical Population Biology* **16** (1979) 144.

[24] Pothen A., Simon H. and Liou K. P., Partitioning sparse matrices with eigenvectors of graphs, *SIAM J. Matrix Anal. Appl.* **11** (1990) 430.

[25] Radicci F., Castellano C., Cecconi F., Loreto V. and Parisi D., Defining and identifying communities in networks, *Proc. Natl. Acad. Sci. USA* **101** (2004) 2658.

[26] Reichardt J. and Bornholdt S., Detecting fuzzy community structures in complex networks with a Potts model, *Phys. Rev. Lett.* **93** (2004) 218701.

[27] Reichardt J. and Bornholdt S., Statistical mechanics of community detection, *Phys. Rev. E* **74** (2006) 016110.

[28] Rosvall M. and Bergstrom C. T., An information-theoretic framework for resolving community srructure in complex networks, *Proc. Natl Acad. Sci. USA* **104** (2007) 7327.

[29] Wasserman S. and Faust K., Social Networks Analysis (Cambridge:Cambridge University Press), 1994.

[30] Zachary W. W., An information flow model for conflict and fission in small groups, *J. Anthropol. Res.* **33** (1977) 452.