Generalized modularity measure for evaluating community structure in complex networks

Xiaoke Ma1,∗, Lin Gao1, Xuerong Yong2, Lidong Fu1 Sheng Wu1
1. School of Computer Science and Technology, Xidian University, 710071, P. R. China
2. Department of Mathematical Sciences, University of Puerto Rico at Mayaguez, P.O.Box 9018, PR 00681, USA

Abstract

Discovering community structure is fundamental for uncovering the links between structure and function in complex networks and modularity optimization is the widely accepted method for this issue. However, there is no consensus criteria for measuring the community structure. In this paper, we propose a new quantitative function for community partition—i.e., generalized modularity or $M$ value. We demonstrate that this quantitative function is superior to the widely used modularity and prove its equivalence with the objective functions of the weighted kernel $k$-means, symmetric matrix factorization and spectral clustering. Both theoretical and numerical results show that optimizing the new criterion not only can tolerate the resolution limit that modularity optimization approaches cannot achieve, but also can extract the number of communities and discovery reasonable communities with different sizes.

keyword: complex networks; community structure; modularity; resolution limit

1 Introduction

Various complex networks, such as social networks [1], technological networks [2], and biological networks [3], can be effectively modeled as graphs by regarding each entity as a vertex and each link as an edge. It has been shown that many real world networks have a structure of modules or communities which are characterized by groups of densely connected nodes. Generally, a community is a subgraph whose nodes are more tightly connected with each other than with nodes outside the subgraph. However, the intuitive meaning of a community differs greatly in different networks. This is, for example, confirmed in the case of the social networks, where communities correspond to social groups with similar interest or background. In protein-protein interaction networks, it is widely believed that the modular structure can be a functional unit.

Identifying community structure in special networks has considerable merit of practice because it gives us insights into the structure-functionality relationship. Over the years, researchers have introduced a large number of algorithms, for example, spectral clustering algorithm [4, 15], betweenness-based method [1], NMF-based algorithm [6, 7], fuzzy clustering approach [8], simulated annealing [10] etc. More algorithms for community detection can be referred to [9, 11]. However, designing of an efficient algorithm for community is still highly nontrivial largely due to the following two reasons.

Measure for community there is no consensus criteria for measuring the community structure, which is a main drawback in many algorithms. To tackle this difficulty, Newman et al. [5] introduced a modularity function $Q$ which measures the quality of a given partition of a network, and it can also be utilized to select automatically optimal number of communities based on the maximum $Q$ value. Many algorithms based on such strategy are proposed in [15–17], although finding optimal $Q$ value is

*Corresponding author. Communication Address: P.O.Box 171, School of computer science and technology, Xidian University, No.2 South TaiBai Road ,Xi’an,Shaanxi,710071,P.R.China. E-mail addresses: maxiaoke8218@163.com (X. Ma).
a NP-hard problem [16, 34]. Recently, Fortunato et al. [30] pointed out the serious resolution limit of widely used $Q$ function and claimed that the size of a detected community depends on the size of the whole network. Specifically, modularity maximization algorithms for community detection may fail to resolve communities with fewer than $\sqrt{L/2}$, where $L$ is the number of edges in the entire network.

**Structure of community** communities may be in complicated shapes. Palla et al. [12] revealed that complex network models exhibit an overlapping community structure, also called fuzzy community. Furthermore, Ravasz et al. [13] proved the existence of the hierarchical organization of modularity in metabolic networks. These overlapping and hierarchical communities are more realistic than average ones. For example, a person in social networks belongs to more than one community at the same time. But, these complicated structures actually make it harder to contrive appropriately algorithms for overlapping or hierarchical communities. Now, only a few efficient algorithms [8, 9, 11, 14] can uncover such realistic structures.

Compared with the second challenge, much fewer works focus on the resolution limit problem of modularity. As far as we known, only few papers are devoted to the resolution limit problem: Li et al. [29] addressed the problem of resolution by defining a modularity alternative called modularity density, shows it to be equivalent to $k$-means; Medus et al. [40] also addressed this issue by presenting new merit factors based on the weak and strong community definitions and they showed this local definitions can also avoid the resolution limit; Arenas et al. also [41] addressed this issue by providing the user with a parameter $r$ that directs modularity maximization algorithms to search for communities of certain natural sizes; The newest work for this topic is presented by Berry et al [42]. They extended Fortunato and Barthélemy’s argument with weighted edges. They concluded that weighted modularity algorithms may fail to resolve communities with fewer than $\sqrt{h\epsilon/2}$ total edge weight, where $h$ is the total edge weight in the network and $\epsilon$ is the maximum weighted of an inter-community edge.

In this paper, we propose a new quantitative measure for evaluating the clustering of a network into communities. We call this quantitative measure generalized modularity or $M$ value. We show that these measures, presented in [29] and [40], are special cases of $M$ value. Further, we also demonstrate that this quantitative function is superior to the widely used $Q$ function and prove its equivalence with the objective functions of the weighted kernel $k$-means, symmetric nonnegative matrix factorization (SNMF) and spectral clustering. Such an equivalence has an immediate implication: we may use the weighted kernel $k$-means, SNMF or spectral clustering to locally optimize the presented generalized modularity, and conversely, optimization $M$ value may be employed for these ones. Both theoretical and numerical results show that maximizing the new criterion not only can resolved detailed modules that existing approaches cannot achieve, but also can recognize and extract the number of communities.

The paper is organized as follows. In section 2, related works is discussed. In section 3, we present the new measure and the analysis for resolution limits problem. Section 4 is devoted to the equivalence of the objectives of some well known algorithms, such as weighted kernel $k$-means, symmetric nonnegative matrix factorization (SNMF) and spectral clustering. Experimental results appear in Section 5.

### 2 Related works

In this section, we briefly review some widely used measures for community structure. Let’s begin with the definition of modularity function $Q$. Given a network $G = (V,E)$ consisting of a vertex set $V = \{v_1,v_2,\cdots,v_{|V|}\}$ (where $|V|$ is the cardinality of $V$) and an edge set $E$, an $|V| \times |V|$ adjacent matrix $A = (A_{ij})$ is constructed with $A_{ij} = 1$ if vertex $i$ is connected to vertex $j$, 0 otherwise. If $V_1$ and $V_2$ are two disjoint subsets of $V$, we define $L(V_1,V_2) = \sum_{i \in V_1,j \in V_2} A_{ij}$, $L(V_1,V_1) = \sum_{i \in V_1,j \in V_1} A_{ij}$ and $L(V_1,\bar{V}_1) = \sum_{i \in V_1,j \in \bar{V}_1} A_{ij}$, where $\bar{V}_1 = V \setminus V_1$. Given a hard partition $\{V_c\}_{c=1}^m$ of $G$ where $V_c$ is the node set of the $c$-th cluster and $m$ is the number of clusters, the well-known modularity function $Q$ is defined as [5]:

$$Q((V_c)_{c=1}^m) = \sum_{c=1}^{m} \frac{L(V_c,V_c)}{L(V,V)} - \left( \frac{L(V_c,V)}{L(V,V)} \right)^2.$$ (2.1)
The modularity $Q$ is a global measure because the $\frac{L(V_c, V_c)}{L(V)}$ and $(\frac{L(V_c, V)}{L(V)})^2$ assume that connections between all pairs of nodes are equally probable which reflects connectivity among all clusters. However, in many complex networks most clusters are connected to only a small fraction of the remaining clusters. To take into account local cluster-connectivity and overcome global network dependency, a local modularity function $LD$ was presented by Muff et al [?]. It can be formulated as:

$$LQ(\{V_c\}_{c=1}^m) = \sum_{c=1}^m L(V_c, V_c) - \frac{L(V_c, V)}{L_c}(2.2)$$

where $L_c$ is the number of links between the $c$-th cluster and its neighborhood. Furthermore, to overcome the limit of modularity for large scale network, a new measure for local community was provided by Bagrow [39].

If one chooses $Q$ as the target function, the community detection problem becomes equivalent to modularity optimization. However, Fortunato and Barthélemy [30] recently pointed out the fatal resolution limit of $Q$ function. To attack this issue, Li et al., [29] introduced a new measure, named by modularity density $D$ value, based on the concept of average modularity density. It is formulated as:

$$D(\{V_c\}_{c=1}^m) = \sum_{c=1}^m \frac{L(V_c, V_c)}{|V_c|} - \frac{L(V_c, \bar{V}_c)}{|V_c|}(2.3)$$

In [40], authors use the weak and strong community definitions, proposed by Radicchi et al [36], to introduce new merit factors to evaluate the quality of a given partition $\{V_c\}_{c=1}^m$ of a network. Let $d_j$ is the degree of vertex $j$. Then, the merit factor $Q_{weak}$ for the weak community definition can be formulated as:

$$Q_{weak} = \sum_{c=1}^m \frac{L(V_c, V_c) - L(V_c, \bar{V}_c)}{\sum_j d_j} \sum_j d_j(2.4)$$

with the constraint that each cluster $V_i$ must satisfy the weak community definition as

$$\frac{L(V_c, V_c) - L(V_c, \bar{V}_c)}{\sum_j d_j} > 0, \forall c \in \{1, 2, \cdots, m\}(2.5)$$

In the same spirit, the merit factor $Q_{strong}$ according to community definition is constructed as:

$$Q_{strong} = \sum_{c=1}^m \frac{L(V_c, V_c) - L(V_c, \bar{V}_c)}{\sum_j d_j} \sum_j d_j(2.6)$$

with constraint

$$d_j^{in} - d_j^{out} > 0, \forall j \in V_c(1 \leq c \leq m), (2.7)$$

where $d_j^{in}$ and $d_j^{out}$ are the number of internal and external links for node $j$.

All these measures discussed above has their own limitations. In next section, we present the generalized modularity.

3 Generalized modularity

From the definitions of $Q$, $D$ and $LQ$, we can conclude that all these measures put emphasize on in-degree and out-degree from global or local perspective without taking into account the importance of each vertex. Actually, it is unreasonable to assumption that each vertex has identical importance in many real world network. In the generalized modularity or $M$ value, which is related to the importance of each vertex. Suppose that vertex $i$ is assigned a weight $w_i(1 \leq i \leq n)$, then the $M$ is defined as

$$M(\{V_c\}_{c=1}^m) = \sum_{c=1}^m \frac{L(V_c, V_c) - L(V_c, \bar{V}_c)}{w(V_c)}(3.1)$$
where \( w(V_c) = \sum_{v_i \in V_c} w_i \) is the total weight for the \( c \)-th component \( V_c \). Similar to \( D \) function, \( M \) function also provides a way to determine if a certain mesoscopic description of the graph is accurate in terms of communities. The larger the value of \( M \), the more accurate a partition is. Then, it is possible to implement the optimization algorithms developed for \( M \).

As the \( Q, D, Q_{weak} \) and \( Q_{strong} \), there is a big issue that \( M \) attains its maximum value \( M_{max} \) when all the network constitutes a single community. That means the best partition of the graph corresponds to no partition. However, it is possible that one could gain valid communities for a given \( m \)-subgraphs partition, with \( m > 1 \) and \( M < M_{max} \). The resulting community structure of network would correspond to a suboptimal solution.

Since \( M \) value largely depends on the weight on each vertex, it is natural to ask how to assign the proper weight for each vertex. In the most intuitive sense, the weight on vertex should capture as much structural information as possible. One may associate the complicated network structural properties, such as the shortest path, spectrum of adjacent matrix or Laplacian matrix, or graph angles, with the weight of each vertex. It is, however, difficult to select proper structural properties to weight each vertex. Fortunately, the main task in this paper is to convey the idea of weighting vertices instead of choosing optimal structural properties. In this paper, for the sake of simplicity, the weight \( w_i \) assigned to vertex \( i \) is related to the degree \( d_i \), defined as

\[
w_i = f(d_i), 1 \leq i \leq |V|,
\]

where \( f \) is a function satisfying

\[
\begin{align*}
& f(t_1) \leq f(t_2) \quad \text{if } \forall \, t_1 \leq t_2 \\
& f(t) > 0 \quad \forall t.
\end{align*}
\]

The constraints on weighted function is easy to meet. Thus, there are many valid weighted functions, such as linear function, exponent function and logarithm function, can be chosen. A further interesting question is how to construct a weighted function for a given network which maximizes the \( M \) value. As we shall see in experiment section, there is no universal weighted function obtains good performance in all networks, i.e., different network requires different weighted function.

Now, we derive the relations among \( M \) and these measures listed in last section. From the Eq.(3.1), we can easily derive that \( D \) value is a special case of \( M \) value under the conditions \( w_i = 1(1 \leq i \leq |V|) \). Furthermore, casting \( w_i = d_i(1 \leq i \leq |V|) \), we can assert that \( Q_{weak} \) is equivalent to \( M \) and \( Q_{strong} \) is a constrained version of \( M \). Such relation immediately demonstrates the superiority of \( M \) function, since it is a generalized version of \( D, Q_{weak} \) and \( Q_{strong} \).

The search for optimal \( M \) is a NP-hard problem as the number of possible partitions of a network into clusters increases at least exponentially with the size of the network. Thus, making exhaustive optimization computationally is unfeasible even for moderated size network. In this paper, we will prove that the community detection problem based on optimizing \( M \) is equal to some famous algorithms. Such a theoretical results may be exploited to derive efficient computational approaches for optimizing \( M \). Before deriving such equivalence relation, we show that \( M \) value can tolerate resolution limit in the next section.

## 4 Resolution limits by generalized modularity

Fortunato et al., [30] proved that the widely used modularity \( Q \) has serious resolution limit in which community smaller than a threshold depends on the whole network may not be discovered. In this section, we perform the same tests as those example in [30] just to show that the \( M \) can tolerate this issue at large extent.

### 4.1 Clique

Given a clique \( K_n \) with \( n \) nodes and \( n(n-1)/2 \), maximizing weighted degree modularity or \( M \) should not divide it into two or more parts.
Figure 1: Schematic examples. (a) The ring of clique graph in left figure. Each community is a clique of \(n\) nodes, and two adjacent communities are connected by one edge. (b) A network with two pair of identical cliques in the right figure. One pair of cliques have \(n\) nodes and the other ones have \(p\) nodes.

We can prove this by contradiction. Suppose that \(G\) is partitioned into two parts \(\{V_c\}_{c=1}^{2}\). Let \(M_0\) be the \(M\) value of \(K_n\) and \(M_1\) be that of the partition \(\{V_c\}_{c=1}^{2}\). Provided that \(|V_1| = n_1\) and \(|V_2| = n_2\), we can

\[
M_0 = \frac{n-1}{f(n-1)},
\]

\[
M_1 = \frac{n_1(n_1-1)}{m_1 f(n-1)} + \frac{n_2(n_2-1)}{m_2 f(n-1)} = \frac{-2}{f(n-1)}.
\]

According to the second property, \(M_1 < M_0\) holds. Thus, maximizing \(M\) value does not separate clique into two parts.

### 4.2 Ring of cliques

The ring of cliques is a schematic example in [30], which is a network consisting of a ring of cliques connected through single links (shown in Fig.1(a)). Each clique is a complete graph \(K_n\) with \(n\) nodes and has \(n(n-1)/2\) edges. If we assume that there are \(m\) cliques, the network has a total of \(mn\) vertices and \(mn(n-1)/2 + m\) edges. For the sake of clarify and to simplify the mathematical expressions without affecting the final result, we assume that the \(d_{\text{max}} \leq n\) where \(d_{\text{max}}\) is the maximum degree of vertex.

Such a network has a clear community structure where the communities correspond to single cliques. However, not all algorithms can be able to recognize and discovery the communities, i.e. algorithms based on modularity optimization [30]. The weighted degree modularity \(M_{\text{single}}\) of this natural partition can be easily calculated as follows

\[
M_{\text{single}} = m \frac{n(n-1)-2}{(n-2)f(n-1) + 2f(n)}.
\]

On the other hand, the weighted degree modularity \(M_{\text{pairs}}\) of the partition in which pairs of consecutive cliques are considered as single communities is

\[
M_{\text{pairs}} = \frac{m}{2} \frac{2n(n-1)}{(2n-4)f(n-1) + 4f(n)},
\]

where \(m\) is supposed to be even.

\[
M_{\text{single}} - M_{\text{pairs}} = m \frac{n(n-1)-2}{(n-2)f(n-1) + 2f(n)} - m \frac{n(n-1)-2}{(n-2)f(n-1) + 2f(n)} = m \frac{n^2 - n - 4}{2(n-2)f(n-1) + 2f(n)} \quad (4.1)
\]
The condition $M_{\text{single}} > M_{\text{pairs}}$ is satisfied only if $n \geq 3$. The above analysis is conducted for the special cases where the 2 consecutive cliques are merged into single communities, by the same argument we can also prove that such a result is valid for any kind of grouping cliques as shown in [29]. For $n = 2$, $K_2$ can never be a community since its outer degree is equal to the inner degree. Such analysis indicates that optimization $M$ can result in the correct partition.

4.3 Cliques with different size

Given a network with four cliques, two of them are $K_n$, the others are $K_p$, for $3 \leq p \leq n$ (see Fig.1 (b)), Fortunato et al. [30] also proved that communities identified through modularity optimization may fail to recognize small ones. In the following, we can prove that $M$ based on optimization can suffer from this risk.

Let $M_{\text{separate}}$ denotes the $M$ value of the partition in which the two $K_p$ are separated and $M_{\text{merge}}$ represents that of the partition where two small cliques are merged. So,

$$M_{\text{separate}} = \frac{n(n-1) - 1}{(n-1)f(n-1) + f(n)} + \frac{n(n-1) - 3}{(n-3)f(n-1) + 3f(n)} + \frac{2}{(p-2)f(p-1) + 2p}.$$  

$$M_{\text{merge}} = \frac{n(n-1) - 1}{(n-1)f(n-1) + f(n)} + \frac{n(n-1) - 3}{(n-3)f(n-1) + 3f(n)} + \frac{2p(p-1)}{(2p-4)f(p-1) + 4p}.$$  

Then,

$$M_{\text{separate}} - M_{\text{merge}} = 2\frac{p(p-1) - 2}{(p-2)f(p-1) + 2p} - \frac{2p(p-1)}{(2p-4)f(p-1) + 4p} = \frac{p^2 - p - 4}{(p-2)f(p-1) + 2p}.$$  

The condition $M_{\text{separate}} > M_{\text{merge}}$ is satisfied only if $p \geq 3$. For $p = 2$, $K_p$ cannot be a community by itself. Such a analysis demonstrates that algorithm based on optimization $M$ can discovery communities with different size.

From the above analysis, we can conclude that the new measure $M$ value can tolerate the resolution limit at large extent. Such a reliability demonstrates the effectiveness of the $M$ value as a alternative for $Q$ function with the purpose to get ride off the resolution limit problem. Another interesting problem is how to design efficient and effective algorithms to maximize $M$ value arised immediately. We will deal with such a problem by showing the equivalence of objectives of $M$ value and other well-known algorithms in the next section.

5 Equivalence of objectives

In this section, we derive the equivalence of objectives of $M$ value, weighted kernel $k$-means algorithm, SNMF and spectral clustering. At first glance, the objectives of these four approaches appear to be unrelated. But, we first express the $M$ value as a trace maximization problem. Then, we rewrite objectives of the weighted kernel $k$-means, SNMF and spectral clustering identically as trace maximization, thus, showing that these objectives are mathematically equivalent.

5.1 Generalized modularity as trace maximization

Let $X_c = (x_c(1), x_c(2), \cdots, x_c(|V|))^T$ is a indicator vector for the $c$-th cluster, where

$$x_c(i) = \begin{cases} 
1 & \text{if } v_i \in V_c \\
0 & \text{otherwise}.
\end{cases}$$
It is easy to see that $X^T_j X_j = |V_i|$ if $i = j$, 0 otherwise. Let $W = \text{diag}(w_1, w_2, \cdots, w_n)$ and $D = \text{diag}(d_1, d_2, \cdots, d_m)$ be two diagonal matrices. Then, we can deduce

$$M(\{V_c\}_{c=1}^m) = \sum_{c=1}^m \frac{L(V_c, V_c) - L(V_c, \bar{V}_c)}{w(V_c)}$$

$$= \sum_{c=1}^m \frac{2L(V_c, V_c) - L(V_c, V)}{w(V_c)}$$

$$= \sum_{c=1}^m \frac{X^T_c (2A - D) X_c}{X^T_c W X_c}$$

$$= \sum_{c=1}^m \bar{X}^T_c (2A - D) \bar{X}_c,$$  

(5.1)

where $\bar{X}_c = X_c/((X^T_c W X_c)^{1/2})$.

Casting the last term of Eq.(5.1) as the trace maximization, we obtain

$$\max_M(\{V_c\}_{c=1}^m) = \max_{\bar{X}} \text{trace}(\bar{X}^T (2A - D) \bar{X}),$$

(5.2)

where $\bar{X}$ is the matrix of all the $\bar{X}_c (1 \leq c \leq m)$ vectors, i.e., $\bar{X} = (\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_m)$.

Let $Z = W^{-1/2} X$, it is easy to validate that $Z^T Z = I_m$. Thus, the target function of $M$ value can be rewritten as the trace maximization

$$\max_M(\{V_c\}_{c=1}^m) = \max_{\bar{Z}} \text{trace}(Z^T W^{1/2} (2A - D) W^{1/2} Z).$$

(5.3)

### 5.2 Generalized modularity and Weighted kernel $k$-means

The weighted Kernel $k$-means is a generalized version of the kernel $k$-means algorithm [43, 44]. Its objective function is expressed as

$$F(\{V_c\}_{c=1}^m) = \sum_{c=1}^m \sum_{v_i \in V_c} w_i \|\phi(v_i) - m_c\|^2,$$

(5.4)

where $m_c = \sum_{v_i \in V_c} \phi(v_i) / \sum_{v_i \in V_c} w_i$ is the $c$-th centroid or the mean of cluster $V_c$ and $\phi(v_i)$ is a function mapping vertex $v_i$ onto a generally higher dimensional space.

The squared distance $\|\phi(v_i) - m_c\|^2$ equals

$$\phi(v_i) \cdot \phi(v_i) - 2 \sum_{v_j \in V_c} \phi(v_i) \cdot \phi(v_j) / (\sum_{v_j \in V_c} \phi(v_j) \cdot \phi(v_j)),$$

(5.5)

Given a kernel matrix $K$ with entry $K_{ij} = \phi(v_i) \cdot \phi(v_j)$, the above expression can be expressed as

$$K_{ii} = 2 \sum_{v_j \in V_c} K_{ij} / (\sum_{v_j \in V_c} w_j)^2.$$

(5.6)

In [44], authors showed that objective function of the weighted $k$-means can be transformed to

$$F(\{V_c\}_{c=1}^m) = \text{trace}(W^{1/2} \Phi^T \Phi W^{1/2}) - \text{trace}(Y^T W^{1/2} \Phi^T \Phi W^{1/2} Y),$$

(5.7)

where $\Phi = (\phi(v_1), \phi(v_2), \cdots, \phi(v_n))$, $Y = [Y_1, Y_2, \cdots, Y_n]$ is an orthonormal $n \times m$ data assignment matrix, i.e., $Y^T Y = I_m$, with entry $Y_c(i) = (w_{ij})^{1/2}$ if $v_i$ belongs to the cluster $V_c$ and 0 otherwise. Since the
first term in the left of Eq. (5.7) is a constant, objective function of the weighted \(k\)-means is equivalent to the maximization of the \(\text{trace}(Y^T W^{1/2} \Phi^T \Phi W^{1/2} Y)\). Further, we can validate that \(Y = \tilde{X}\) holds, thus, the following expression can be got

\[
\min F(\{V_c\}_{c=1}^m) \propto \text{trace}(Z^T W^{1/2} K W^{1/2} Z),
\]

(5.8)

where \(K = \Phi^T \Phi\) is a kernel function and \(Z\) has the same meaning in last subsection.

Based on Eqs. (5.3) and (5.8), we can assert the equivalence of objectives of the weighted kernel \(k\)-means and \(M\) value by setting \(K = 2A - D\),

\[
\max M(\{V_c\}_{c=1}^m) \propto \max F(\{V_c\}_{c=1}^m).
\]

(5.9)

However, it is important to note that the matrix \(2A - D\) should be positive semi-definite to guarantee convergence of the maximization \(M\). To attack this issue, a fluctuation factor is added to the kernel function as \(K = \sigma I + 2A - D\), where \(\sigma\) is a real number chosen to be sufficiently large so that \(K\) is positively definite. Such an operator does not change the equivalence relation in Eq. (5.9), since

\[
\text{trace}(Z^T W^{1/2}(\sigma I + 2A - D) W^{1/2} Z) = m \sigma + \text{trace}(Z^T W^{1/2}(2A - D) W^{1/2} Z).
\]

5.3 Generalized modularity and SNMF

NMF is a recently introduced method for multivariate data [24,25]. It has received considerable attention in several disciplines immediately, particularly in the complex network [6,7]. In this section, we show the equivalence of the objectives of SNMF and the weighted degree modularity.

From Eq. (5.3), we can obtain easily

\[
m \sigma + \max M = \max_Z \text{trace}(Z^T W^{1/2} K W^{1/2} Z)
\]

(5.10)

where \(K = \sigma I + 2A - D\). We show that the right term in above expression can be solved by the symmetric NMF

\[
K \approx (W^{1/2} Z)(W^{1/2} Z)^T, Z \geq 0.
\]

(5.11)

Casting it as an optimization form, the objective function can be approximated by minimizing the sum of squared errors, defined as

\[
\min_{Z \geq 0} \| K - (W^{1/2} Z)(W^{1/2} Z)^T \|^2,
\]

(5.12)

where \(\|B\|^2\) is the Frobenius norm of matrix \(B\). Further, we can rewrite the right term of Eq. (5.10) as

\[
\max_Z \text{trace}(Z^T W^{1/2} K W^{1/2} Z)
\]

\[
\propto \min_Z \text{trace}(Z^T W^{1/2} K W^{1/2} Z)
\]

\[
\propto \min_{(W^{1/2} Z)^T (W^{1/2} Z) = I_m, Z \geq 0} \| K - (W^{1/2} Z)(W^{1/2} Z)^T \|^2.
\]

(5.13)

Relaxing the orthogonality \((W^{1/2} Z)^T (W^{1/2} Z) = I_m\), we can assert that SNMF is equivalent to \(M\) value, i.e.,

\[
\max_M \propto \min_{Z \geq 0} \| K - (W^{1/2} Z)(W^{1/2} Z)^T \|^2.
\]

(5.14)
5.4 Generalized modularity and spectral clustering

There are a number of different graph clustering objectives have been proposed and studied, for example, ratio association [45], ratio cut [46], normalized cut [47], etc. Dillon et al. [44] generalized the ratio and cut problems to weighted variant. They also proved the equivalence of the target functions of the general weighted graph cuts/association and the weighted $k$-means. Here, we derive the relation between the general weighted graph cuts/association and $M$ value.

The weighted association problem can be expressed as a trace maximization form

$$W_{Assoc}(\{V_c\}_{c=1}^m) = \sum_{c=1}^m \frac{L(V_c, V_c)}{w(V_c)} = \sum_{c=1}^m X_c^T A X_c = \sum_{c=1}^m \tilde{X}_c^T A \tilde{X}_c,$$

(5.15)

where $\tilde{X}_c = X_c / (X_c^T W X_c)^{1/2}$. Then, the above expression can be transformed to trace maximization

$$\max_{\tilde{X}} W_{Assoc}(\{V_c\}_{c=1}^m) = \max_{\tilde{X}} \text{trace}(\tilde{X}^T A \tilde{X}),$$

(5.16)

where $\tilde{X}$ is the matrix of vectors $\tilde{X}_c$. Hence, optimizing $W_{Assoc}$ on the matrix $A$ is equivalent to optimizing $M$ value on $2A - D$ with $W = I$.

Similar analysis applied to general weighted cuts problem, we can obtain

$$\max_{\tilde{X}} W_{Cut}(\{V_c\}_{c=1}^m) = \min_{\tilde{X}} \text{trace}(\tilde{X}^T (D - A) \tilde{X}).$$

(5.17)

Hence, optimizing $W_{Assoc}$ on the matrix $D - A$ is equivalent to optimizing $M$ value on $2A - D$ with $W = I$.

6 Experimental analysis

In this section, the new measures $GQ$ is applied to a class of widely used artificial networks and well studied real-world networks with an immediate purpose to test its performance. First, the ability to discovery reasonable community structure and the correct number of clusters is investigated; Second, the ability to tolerate resolution limit problem is also studied. The simulated annealing algorithm, presented in [10], is adopted here because of its simplicity and good performance on community detection. The optimization algorithm can be used in two ways. If the number of communities is known, we use it to discovery the communities. Otherwise, we start with the number of clusters $m = 2$ and then keep increasing $m$ until the $GQ$ reaches its peak. After fixing $m$, we can discovery reasonable communities. And, the algorithm is coded using the MATLAB version 6.5.

It is worth mentioning that the main goal of this experiment is to show the proposed measure $GQ$ can be used to guide community detection problem, instead of finding optimal weighting manner. Thus, we only consider three styles of weight strategy as following:

- $W_I: f(d_i) = ad_i + b, i = \{1, 2, \cdots, |V|\},$
- $W_{II}: f(d_i) = \ln(d_i + c) / \ln(a) + b, i = \{1, 2, \cdots, |V|\}$
- $W_{III}: f(d_i) = a \exp(d_i) + b, i = \{1, 2, \cdots, |V|\},$

where coefficients $a, b, c$ are chosen such that $f(d_i) > 0$. Obviously, $W_I, W_{II}$ and $W_{III}$ are linear function, logarithm and exponential function. Of course, there are maybe other complicated weighting function having much better performance.
7 Conclusion

In this paper, we proposed a measured, called weighted modularity density or $M$ function for tolerating the resolution limit problem. We showed that some former measures, such as $D$ function, $Q_{weak}$ and $Q_{strong}$ are special cases of $M$ function. Furthermore, we also proved that our criterion is equivalent to the objective function of weighted kernel $k$-means algorithm. The analysis for resolution limit was also presented. And, the experiment results also supported the theoretic analysis.

We would like to close our paper by posing two possible further researching directions related to weighted modularity density.

References


