Eur. Phys. J. B (2012) 85: 244
DOI: 10.1140/epjb/e2012-20898-3

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

Maximum modular graphs

S. Trajanovski[a], H. Wang, and P. Van Mieghem

Delft University of Technology, Faculty of EEMCS, P.O. Box 5031, 2600 GA Delft, The Netherlands

Received 3 November 2011 / Received in final form 7 May 2012
Published online 18 July 2012 –

Abstract. Modularity has been explored as an important quantitative metric for community and cluster detection in networks. Finding the maximum modularity of a given graph has been proven to be NP-complete and therefore, several heuristic algorithms have been proposed. We investigate the problem of finding the maximum modularity of classes of graphs that have the same number of links and/or nodes and determine analytical upper bounds. Moreover, from the set of all connected graphs with a fixed number of links and/or number of nodes, we construct graphs that can attain maximum modularity, named *maximum modular graphs*. The maximum modularity is shown to depend on the residue obtained when the number of links is divided by the number of communities. Two applications in transportation networks and data-centers design that can benefit of maximum modular partitioning are proposed.

1 Introduction

Real-world networks are composed of hierarchically connected communities. For instance, people in online social networks tend to connect with their friends, forming cohesive groups of schoolmates, colleagues and others with similar interests. Web pages covering the same topic tend to link to each other rather than to web pages covering different topics.

Quantitative measures for detecting the presence of communities can aid network exploration by dividing the analysis of interactions into inter-community and intra-community analyses. In this paper, we focus on one such measure: Newman's modularity [1]. Determining a partitioning of nodes in a network that maximizes the modularity is known to be an NP-complete problem [2] and several heuristic algorithms have been proposed [3–6] for calculating the maximum modularity.

The modularity is a multi-variable metric depending on the graph's topology, the number of nodes N and the number of links L, the degree distribution, and the number of communities c. We target the problem of finding the maximum modularity of networks subject to certain constraints, such as a fixed number of links L. We rewrite the common modularity expression in equation (2) [7] in a different form given in equation (4), which is suitable for finding the maximum modularity [8]. Our results determine:

– the number of links between the communities (*interconnecting links*) contributes [9] by an order of $O(L^{-1})$ and the differences of the *cumulative degrees* contribute by an order of $O(L^{-2})$ in the modularity;
– an upper bound for the modularity of the class of graphs with a given number of links L and given number of communities c;
– an upper bound for the modularity of the class of graphs with a given number of links L and nodes N, divided into 2 communities;
– graphs that attain maximum modularity are constructed, named *maximum modular graphs*.

Our results provide a step forward in the problem of finding the maximum modularity in general. Finally, we give two possible applications in transportation planning and data-center design, where maximum modular graphs could be used.

The remaining part of the paper is outlined as follows. Section 2 gives an overview of the current literature on modularity. In Section 3, we cover the notation and the theoretical basis needed for this paper. Our upper bound for the modularity is derived for the class of graphs with a given number of links L and communities c in Section 4. Section 5 places an additional constraint, namely that, besides a given L and c, also the number of nodes N is fixed. Moreover, in both Sections 4 and 5, we construct the graphs that attain the maximum modularity. In Section 6, some practical applications are proposed and the paper is concluded in Section 7.

2 Related work

Newman and Girvan's introduction of the modularity [1] immediately attracted attention in the fields of community detection and clustering. Unfortunately, finding the maximum modularity for a given graph has been proven to be an NP-complete problem [2]. A thorough survey of

[a] e-mail: S.Trajanovski@tudelft.nl

the modularity definition, proposed algorithms and close variations is given by Fortunato [7].

Newman [3] proposed a *greedy* algorithm for finding the maximum modularity, whose main drawback is that in some cases only local maxima are found. *Simulated annealing*, a technique that randomly explores the spaces of all the possible solutions, has been applied to this problem [4,5]. *External optimization*, which is a variation of *simulated annealing*, has also been used for modularity maximization [6]. A spectral approach, based on Fiedler's eigenvector partitioning [10], was also used for modularity [11], where the modularity matrix [12] is used instead of the Laplacian. An algorithm for calculating modularity of very large networks is given in [13].

A more theoretical approach was followed by Brandes et al. [2], who formally proved the NP-hardness of modularity maximization. Like our approach, they focus on the problem of maximum modular structures. Unlike our analytical upper bounds for the modularity, they proposed an *integer linear programming* formulation. Finding the maximum modularity as a mathematical programming problem was also suggested by Agarwal and Kempe [14]. Regarding the modularity, a measure for quality of the communities has been proposed by Delvenne et al. [15], who also found that "balanced" communities leads to high modularity. Unlike this work, we re-write modularity in a suitable form, such that, the property of "balanced" communities could be observed and furthermore, analytically determine the *maximum modular graphs*. Finally, an upper bound, which is only based on the number of links between the communities by an order of $\mathrm{O}(L^{-1})$ has been determined by Fortunato and Barthélemy [9]. However, we have extended this result by deducing a tighter upper bound and an additional dependence on the *cumulative degree* differences by an order of $\mathrm{O}(L^{-2})$. Furthermore, we have determined the structure of *maximum modular graphs*. Unlike our work, Fortunato and Barthélemy [9] have pointed to the weakness of algorithms for detecting small communities and the effect on modularity by merging communities. To determine the limit size of a "detectable" community, they have introduced the concept of *resolution limit*.

3 Another expression for the modularity

Newman's original expression [1] for the modularity m of a graph with N nodes and L links states that

$$m = \frac{1}{2L}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(a_{ij} - \frac{d_i d_j}{2L}\right) 1_{\{i,j\in \text{ the same community}\}} \quad (1)$$

where a_{ij} is adjacency matrix element and d_j is the degree of node j.

Definition 1. *In a certain community C_i $(i = 1, 2, \ldots, c)$*

(i) *The cumulative degree D_{C_i} is the sum of the degrees of all nodes that belong to the community.*

(ii) *The total number of links within the community is denoted by L_{C_i}.*

By the Definition 1, D_{C_i} counts L_{C_i} (twice), but also the links that connect the community C_i with another community C_j $(i \neq j)$.

Definition 2. *The links whose end-nodes belong to different communities are named inter-connecting links. The sum of all inter-connecting links is denoted by L_{inter}.*

If we sum over all pairs of nodes within the community C_k, we have

$$\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}\cdot 1_{i,j\in C_k} = 2L_{C_k}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} d_i d_j\cdot 1_{i,j\in C_k} = \sum_{i=1}^{N} d_i \cdot \sum_{j=1}^{N} d_j \cdot 1_{i,j\in C_k} = D_{C_k}^2.$$

Substituted into equation (1), we deduce the widely used [7,8], alternative form of the modularity

$$m = \sum_{k=1}^{c}\left(\frac{L_{C_k}}{L} - \left(\frac{D_{C_k}}{2L}\right)^2\right). \quad (2)$$

Using the Cauchy identity [16]

$$\sum_{j=1}^{c} x_j^2 \sum_{k=1}^{c} y_k^2 = \left(\sum_{j=1}^{c} x_j y_j\right)^2 + \frac{1}{2}\sum_{j=1}^{c}\sum_{k=1}^{c}(x_j y_k - x_k y_j)^2$$

with $x_j = D_{C_j}$, $y_j = \frac{1}{\sqrt{c}}$ for each $j \in \{1, 2, \ldots, c\}$ and the fact that each link is counted twice in the sum of the *cumulative degrees* $(\sum_{k=1}^{c} D_{C_k} = 2L)$, we obtain

$$\sum_{k=1}^{c} D_{C_k}^2 = \frac{(2L)^2}{c} + \frac{1}{2c}\sum_{j=1}^{c}\sum_{k=1}^{c}\left(D_{C_j} - D_{C_k}\right)^2. \quad (3)$$

Additionally using the fact that $\sum_{k=1}^{c} L_k = L - L_{\text{inter}}$ and equation (3) in equation (2), we arrive at yet another expression for the modularity

$$m = 1 - \frac{1}{c} - \frac{L_{\text{inter}}}{L} - \frac{1}{2c}\sum_{j=1}^{c}\sum_{k=1}^{c}\left(\frac{D_{C_j} - D_{C_k}}{2L}\right)^2. \quad (4)$$

From equation (4), we immediately obtain the upper bound $1 - \frac{1}{c}$ that is only achieved for disconnected graphs $(L_{\text{inter}} = 0)$ with all equal *cumulative degrees*. Moreover, equation (4) clearly suggests that graphs that attain maximum modularity pose following properties: (1) have small number [9] of *inter-connecting* links and their contribution in the modularity is by an order of $\mathrm{O}(L^{-1})$ and, (2) all the communities are "balanced" i.e. have similar *cumulative degrees* (D_{C_i}), which has been also inspected in [15]. Additionally, equation (4) shows that the differences of *cumulative degrees* contribute by an order of $\mathrm{O}(L^{-2})$ in the modularity. This observation is related to the *resolution limit* [9] dependence on the number of links per community. Nevertheless, we shade new quantitative light by finding that *maximum modular graphs* tend to have the smallest number of *inter-connecting* links as a first priority and "balanced" communities as a second priority as $O(L^{-1}) > O(L^{-2})$. However, for a given numbers of links and communities, it is not always possible to construct completely "balanced" communities (equal D_{C_i}). In the sequel of this paper, we based upon mentioned facts and analytically determine the structure and details of the *maximum modular graphs*.

4 Maximum modular graphs for given numbers of links L and communities c

In this section we consider the set of all graphs G with a fix number of links L and we determine those graphs, called *maximum modular graphs* that attain a maximum value for the modularity. Indeed, the modularity of all other partitions and all other graphs (with fixed number of links L) is not larger than the *maximum modularity*.

We show below that the maximum modularity depends on the residue L modulo c. For each possible residue L *mod* $c = r$, we find the maximum value, but also at least one partitioning where this modularity is obtained.

4.1 Graph modifications

We denote by *community graph* the graph abstraction, where a node reflects one community and a link connects two nodes from the original graph from different communities. Particularly, the term *tree-configuration* is a *community graph*, which is a tree. It is important to highlight that the original graph is not necessarily a tree, but only the *community graph*, which is composed by the communities. A community with exactly one *inter-connecting* link is *leaf* community. A *star-configuration* is a *community graph*, which is a star. In the *star-configuration*, the community connected to all *leaf* communities is called *central*. Similarly, *star-configuration* is different from star graph as in the first a node is one community.

We define the following graph modifying steps:

Definition 3. *Replacing one* inter-connecting *link between two communities with an internal link in one of those two communities is* local rewiring.

Definition 4. *Replacing one internal link in one community with an internal link in another community is* shifting.

Definition 5. *If the number of communities $c > 2$, replacing one* inter-connecting *link between two communities with one internal link in a third (different) community is* general rewiring.

The states before and after each of the defined operations are given in Figure 1: *local re-wiring* ((a) and (b)), *shifting* ((c) and (d)) and *general rewiring* ((e) and (f)).

The following Lemma 1 shows that the modularity increases by decreasing the number of *inter-connecting* links, given that the number of links L is fixed.

Lemma 1. *The modularity increases by* local rewiring.

Proof. The graph G is partitioned into c communities and we use the definitions given in Section 3. We apply *local rewiring* between C_i and C_j, such that an *inter-connecting* link between C_i and C_j is deleted and a new link is added in the community C_i. The result is a new graph G' with corresponding partitioning c' and all its properties mentioned in Section 3 are denoted by an accent. It holds that

$$\begin{cases} D'_{C_k} = D_{C_k}, \text{ for } k \neq i \text{ and } k \neq j \\ D'_{C_i} = D_{C_i} + 1 \text{ and } D'_{C_j} = D_{C_j} - 1 \\ L'_{\text{inter}} = L_{\text{inter}} - 1 \text{ and } c' = c. \end{cases} \quad (5)$$

Using equation (4) and equation (5) the difference between the modularity of G and G' is

$$m' - m = \frac{2L - (D_{C_i} - D_{C_j} + 1)}{2L^2}.$$

Details are given in Appendix A.1. Because of the basic law of the degrees $\sum_{k=1}^{c} D_{C_k} = 2L$; $D_{C_i} \leq 2L - 1$ and $D_{C_j} \geq 1$ we have $D_{C_i} - D_{C_j} + 1 \leq 2L - 1 - 1 + 1 = 2L - 1$. The difference is lower bounded as

$$m' - m \geq \frac{2L - (2L - 1)}{2L^2} = \frac{1}{2L^2} > 0.$$

Hence the modularity m' of the graph G' is larger. □

The next Lemma 2 resolves the distribution of the community's cumulative degrees in the maximum modular graph.

Lemma 2. *In the maximum modular graph with a given number of links L and number of communities c, the absolute value of the difference between two cumulative degrees D_{C_i} and D_{C_j} is always 0, 1 or 2, thus $|D_{C_i} - D_{C_j}| \leq 2$.*

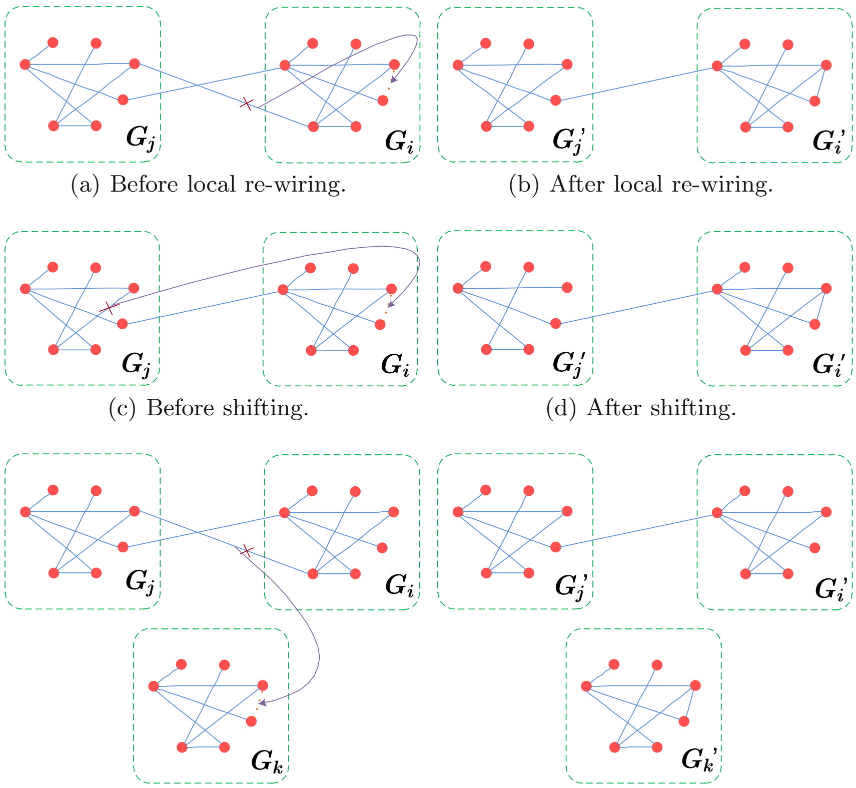
Proof. Let us consider a maximum modular graph with cumulative degrees D_{C_k}, $k = 1, 2, \ldots, c$. Assume that there are two communities C_i and C_j such that $D_{C_i} - D_{C_j} > 2$. If we *shift* one link from the community C_i into the community C_j, we obtain a new graph G' and then the corresponding partitioning has modularity m'. Now, if we label the cumulative degrees by D'_{C_k}, $k = 1, 2, \ldots, c$, it holds that

$$\begin{aligned} &D'_{C_k} = D_{C_k}, \text{ for } k \neq i \text{ and } k \neq j \\ &D'_{C_i} = D_{C_i} - 2 \text{ and } D'_{C_j} = D_{C_j} + 2. \end{aligned} \quad (6)$$

Using equation (6), we show in Appendix A.2 that

$$m' - m = \frac{D_{C_i} - D_{C_j} - 2}{L^2} > 0.$$

Thus we obtain a graph G' with higher modularity, which is a contradiction. The last proves that $|D_{C_i} - D_{C_j}| \leq 2$ or every two cumulative degrees D_{C_i} and D_{C_j} are equal, consecutive or differ by 2. □



(a) Before local re-wiring. (b) After local re-wiring.

(c) Before shifting. (d) After shifting.

(e) Before global re-wiring. (f) After global re-wiring.

Fig. 1. (Color online) Graph modifying steps.

The next Lemma 3 shows that increasing/decreasing the modularity by a *general rewiring* is conditional.

Lemma 3. *The modularity does not necessarily increase if one* inter-connecting *link is changed by an internal link.*

Proof. The graph G is partitioned into c communities with cumulative degrees D_{C_k}, $k = 1, 2, \ldots, c$ and the number of *inter-connecting* links is L_{inter}. If we *globally rewire,* i.e. we delete one *inter-connecting* link between C_i and C_j and add a new link in the community C_k, then we obtain a graph G' with corresponding partitioning $c' = c$. It holds that

$$\begin{cases} D'_{C_i} = D_{C_i} - 1 \text{ and } D'_{C_j} = D_{C_j} - 1 \\ D'_{C_k} = D_{C_k} + 2 \text{ and } L'_{\text{inter}} = L_{\text{inter}} - 1 \end{cases} \quad (7)$$

and the modularity is

$$m' = 1 - \frac{L'_{\text{inter}}}{L} - \frac{1}{c} - \frac{1}{2c}\sum_{p=2}^{c}\sum_{k=1}^{c}\left(\frac{D'_{C_p} - D'_{C_k}}{2L}\right)^2. \quad (8)$$

Using equation (4) and equation (7) the difference between the modularity is

$$m' - m = \frac{2L + D_{C_i} + D_{C_j} - 2D_{C_k} - 3}{2L^2}. \quad (9)$$

Details are given in Appendix A.3. If $D_{C_i} = D_{C_j} = 1$ and $D_{C_k} = 2L - 2$ (for $c = 3$), then $m' < m$ such that the modularity decreases. If $D_{C_k} = 1$, then $m' > m$ in which case the modularity increases. □

Corollary 1. *In a connected graph, the modularity increases if one* inter-connecting *link is changed by an internal link in the community C_k, under the condition $D_{C_k} \leq L - 1$.*

Proof. According to the relation (9) and the fact that the graph is connected $D_{C_i} \geq 1$ and $D_{C_j} \geq 1$

$$m' - m \geq \frac{2L + 1 + 1 - 2(L-1) - 3}{2L^2} = \frac{1}{2L^2} > 0. \quad (10)$$

This implies that the modularity increases. □

4.2 A maximum modular connected graph

In this section, we focus on the value of the modularity of a maximum modular connected graph and the existence of the corresponding partitioning.

Corollary 2. *The maximum modular connected graph with L links and divided into c communities has a tree-*configuration *and $L_{inter} = c - 1$.*

Proof. Assume that the number of *inter-connecting* links in the graph obeys

$$L_{\text{inter}} \leq c - 2. \quad (11)$$

Inequality (11) implies that the number of connected communities is at most $c-1$. Consequently, at least one community will be isolated from the remaining part, such that the graph is not connected. Suppose next a maximum modular connected graph, where

$$L_{\text{inter}} \geq c. \quad (12)$$

There are two possibilities in this case:

- There are two communities C_i and C_j with at least 2 *inter-connecting* links between them. *Local rewiring* can be applied between C_i and C_j.
- The communities are arranged in a *circle-configuration*. This is a *community graph*, where all the communities considered as a nodes form a circle graph. Exactly 2 *inter-connecting* links start from each community. It is again possible to apply *local rewiring*.

In both cases, the communities C_i and C_j will remain connected and also the graph in general. According to the Lemma 1 the modularity of the modified graph after local rewiring is higher. Hence, we constructed a graph with the same number of links L, but with a higher modularity, which is in contradiction with inequality (12). As a result of inequalities (11) and (12), we obtain $L_{\text{inter}} = c-1$ for a maximum modular graph. The last is possible when all the communities are arranged in a *tree-configuration*. □

Based on Corollary 1 and equation (4) for the modularity, a maximum modular graph obeys

$$m = 1 - \frac{1}{c} - \frac{c-1}{L} - \frac{1}{c}\sum_{j=2}^{c}\sum_{k=1}^{j-1}\left(\frac{D_{C_j} - D_{C_k}}{2L}\right)^2. \quad (13)$$

4.2.1 Implication of the residue L mod c on the maximum modularity

The double sum in equation (13) can be rewritten as

$$S_m = \frac{1}{8cL^2}\sum_{j=1}^{c}\sum_{k=1}^{c}\left(D_{C_j} - D_{C_k}\right)^2. \quad (14)$$

A community connected to exactly one other community is called a *leaf community*. Let $\mathcal{P}_k$ be the set of communities connected to k other communities and $n_k = |\mathcal{P}_k|$. Hence, $\mathcal{P}_1$ is the set of *leaf communities*. If all the communities in $\mathcal{P}_k$ have the same cumulative degree equal to D, the numbers of internal links are also the same and equal $l_k = (D-k)/2$.

Lemma 4. *In a* tree-configuration $S_m = 0$ *if and only if* c *is even and* $L \equiv \frac{c}{2} \pmod{c}$

Proof. Suppose that the graph is partitioned in a *tree-configuration* with $L_{\text{inter}} = c-1$ (Corollary 2) and that $S_m = 0$. Based on equation (14), we have that $D_{C_i} = D_{C_j}$ for all $i, j \in \{1, 2, \ldots, c\}$. There are c communities in total

$$\sum_{k=1}^{c-1} n_k = c. \quad (15)$$

If we sum over all *inter-connecting* links $L_{\text{inter}} = c-1$ per community, considering that each link is counted twice, we have

$$\sum_{k=1}^{c-1} k n_k = 2(c-1). \quad (16)$$

If the number of internal links in the *leaf community* C_p of a *tree-configuration* is l, then the cumulative degree is

$$D_{C_p} = 2l + 1.$$

Hence, all *leaf communities* have an odd cumulative degree. Because all cumulative degrees are equal, all communities have an odd number of *inter-connecting links*. The last results in $n_{2i} = 0$, for $i = 1, 2, \ldots, (t-1)$ and $t = \left[\frac{c}{2}\right]$ such that equation (15) becomes

$$\sum_{i=1}^{t} n_{2i-1} = c \quad (17)$$

and equation (16) is transformed into

$$\sum_{i=1}^{t} (2i-1)\, n_{2i-1} = 2\,(c-1)\,. \quad (18)$$

If we subtract equation (17) from equation (18), we obtain

$$2\sum_{i=2}^{t} (i-1)\, n_{2i-1} = c - 2. \quad (19)$$

Equation (19) implies that c must be even, hence that $t = \left\lfloor\frac{c}{2}\right\rfloor = \frac{c}{2}$. Thus, it can be rewritten as

$$\sum_{i=2}^{\frac{c}{2}} (i-1)\, n_{2i-1} = \frac{c}{2} - 1. \quad (20)$$

Consider the total number of links in the graph. Because $n_{2i} = 0$, there exist only communities connected to k other communities, where k is odd and we can use the variables l_{2i-1} for the number of internal links in each community in $\mathcal{P}_{2i-1}$. The cumulative degree of the communities in $\mathcal{P}_{2i-1}$ is $(2l_i + 2i - 1)$. If we compare the cumulative degrees of the communities in $\mathcal{P}_{2i-1}$ and the *leaf communities* $(\mathcal{P}_1)$ we obtain

$$\begin{aligned} & 2l_{2i-1} + 2i - 1 = 2l + 1 \\ \Longleftrightarrow\ & l_{2i-1} = l - i + 1 \end{aligned}$$

for all $i = 1, 2, \ldots, \frac{c}{2}$ and $l_1 = l$ number of internal links in a *leaf community*. If we sum over all those links and add the $L_{\text{inter}} = c-1$ *inter-connecting* links, using the relation for l_{2i-1} and $n_{2i} = 0$ we find

$$\begin{aligned} & c - 1 + \sum_{i=1}^{c-1} n_i l_i = L \\ \Longleftrightarrow\ & c - 1 + \sum_{i=1}^{\frac{c}{2}} n_{2i-1} l_{2i-1} + \sum_{i=1}^{\frac{c}{2}-1} n_{2i} l_{2i} = L \\ \Longleftrightarrow\ & c - 1 + \sum_{i=1}^{\frac{c}{2}} n_{2i-1}\,(l - i + 1) = L \end{aligned}$$

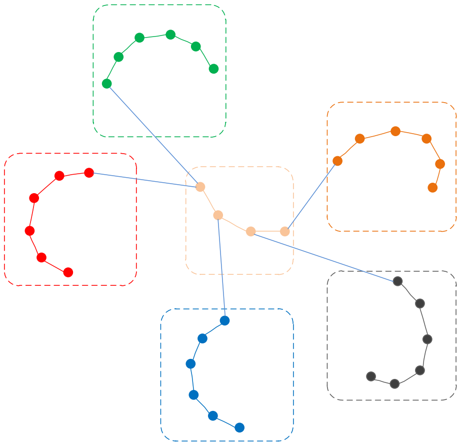


Fig. 2. (Color online) One example of maximum modular graph for $L = 33$ and $c = 6$.

and

$$l\sum_{i=1}^{\frac{c}{2}} n_{2i-1} - \sum_{i=2}^{\frac{c}{2}} (i-1)\, n_{2i-1} + c - 1 = L. \tag{21}$$

Substituting equation (17) and equation (20) into equation (21) yields

$$\begin{aligned} & cl - \left(\tfrac{c}{2} - 1\right) + c - 1 = L \\ \Longleftrightarrow\quad & L = cl + \tfrac{c}{2}. \end{aligned} \tag{22}$$

From equation (22), it follows that the assumption is possible if $L \equiv \frac{c}{2} \pmod{c}$.

For the proof in another direction, if c is even and $L \equiv \frac{c}{2} \pmod{c}$, $L = lc + \frac{c}{2}$ there is a graph with corresponding partitioning, where $S_m = 0$. A *star-configuration*, where the *central community* has $l - \frac{c}{2} + 1$ internal links and is connected with all others $(c-1)$ *leaf communities* each with l internal links, satisfies the condition. Figure 2 exemplifies the case for $L = 33$ and $c = 6$ with modularity $m = \frac{15}{22} \approx 0.682$.

The total number of links is $(c-1)\, l + \left(l - \frac{c}{2} + 1\right) + (c-1) = cl + \frac{c}{2} = L$. The cumulative degree of the *central community* is $2\left(l - \frac{c}{2} + 1\right) + c - 1 = 2l + 1$ equal to the cumulative degree of each of the leaf communities which leads to $D_{C_i} = D_{C_j}$ for all $i, j \in \{1, 2, \ldots, c\}$ and $S_m = 0$. Hence, Lemma 4 is completed. □

Another example when $L \equiv \frac{c}{2} \pmod{c}$ and $S_m = 0$ is given in Figure 3, particularly when $r = \frac{c}{2}$.

Lemma 4, combined with equations (14) and (13), shows, when c is even and $L \equiv \frac{c}{2} \pmod{c}$, that the maximum modular graph has modularity

$$m = 1 - \frac{1}{c} - \frac{c-1}{L}$$

and if c is odd or $L \equiv r \pmod{c}$, $r \neq \frac{c}{2}$ the modularity of the maximum modular graph is

$$m < 1 - \frac{1}{c} - \frac{c-1}{L}.$$

Let us denote the set of communities with equal cumulative degree k by Q_k and $|Q_k| = q_k$. Based on the Lemma 2, all communities in the graph with maximum modularity might be divided into three sets Q_i, with $i \in \{D-1, D, D+1\}$ (consecutive numbers). Clearly,

$$|Q_{D-1}| + |Q_D| + |Q_{D+1}| = q_{D-1} + q_D + q_{D+1} = c.$$

Let us consider all the values $|D_{C_i} - D_{C_j}|$, where $i, j \in \{1, 2, \ldots, c\}$ in the sum S_m in equation (14). All the differences between elements in the same set Q_{D-1}, Q_D or Q_{D+1} are 0. In the other cases, we have

$$|D_{C_i} - D_{C_j}| = \begin{cases} 1, & C_i \in Q_{D-1}, C_j \in Q_D \\ 1, & C_i \in Q_D, C_j \in Q_{D+1} \\ 2, & C_i \in Q_{D-1}, C_j \in Q_{D+1}. \end{cases}$$

Thus, according to equation (14) and the cardinalities of Q_{D-1}, Q_D and Q_{D+1}

$$\begin{aligned} S_m &= \frac{2q_{D-1}\cdot q_D \cdot 1 + 2q_D \cdot q_{D+1} \cdot 1 + 2q_{D-1}\cdot q_{D+1}\cdot 2^2}{8cL^2} \\ &= \frac{q_D\,(q_{D-1} + q_{D+1}) + 4q_{D-1}\cdot q_{D+1}}{4cL^2} \\ &= \frac{q_D\,(c - q_D) + 4q_{D-1}\cdot q_{D+1}}{4cL^2}. \end{aligned} \tag{23}$$

Because $q_{D-1}, q_{D+1} \geq 0$ we have

$$S_m \geq \frac{q_D\,(c - q_D)}{4cL^2}. \tag{24}$$

With the exception that $c \mid L$, we will prove that it is possible to construct the case, where equality in inequality (24) is achieved. Equality is possible only if ($q_{D-1} = 0$ or $q_{D+1} = 0$) and q_D is an integer from the set $\{1, 2, \ldots, c-1\}$. Without loss of generality we can consider the case $q_{D-1} = 0$, because the equality in inequality (24) is achieved when q_i are two consecutive integers with sum c.

Lemma 5. *If $L \equiv r \pmod{c}$, and $r \neq 0$, the minimum value of S_m is*

$$S_m = \begin{cases} \frac{r(c-2r)}{2cL^2}, & 1 \leq r \leq \left[\frac{c}{2}\right] \\ \frac{(c-r)(2r-c)}{2cL^2}, & \left[\frac{c}{2}\right] < r \leq c-1. \end{cases}$$

Proof. Suppose that $q_{D-1} = 0$ and $1 \leq q_{D+1} \leq c-1$. Using $L = ck + r$, $r \neq 0$ and $q_D = c - q_{D+1}$ we derive

$$(c - q_{D+1}) \cdot D + q_{D+1} \cdot (D+1) = 2L = 2ck + 2r$$

from which

$$D = 2k + \frac{2r - q_{D+1}}{c}. \tag{25}$$

Because D is integer, we have

$$c \mid (2r - q_{D+1}). \tag{26}$$

We distinguish two cases: $1 \leq r \leq \left\lfloor \frac{c}{2} \right\rfloor$ and $\left\lfloor \frac{c}{2} \right\rfloor + 1 \leq r \leq c-1$.

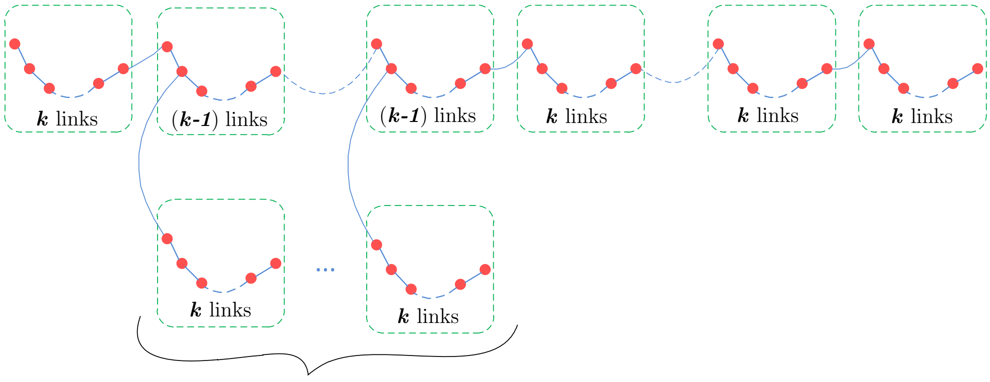


Fig. 3. (Color online) Maximum modular partitioning for $1 \leq r \leq \lfloor \frac{c}{2} \rfloor$.

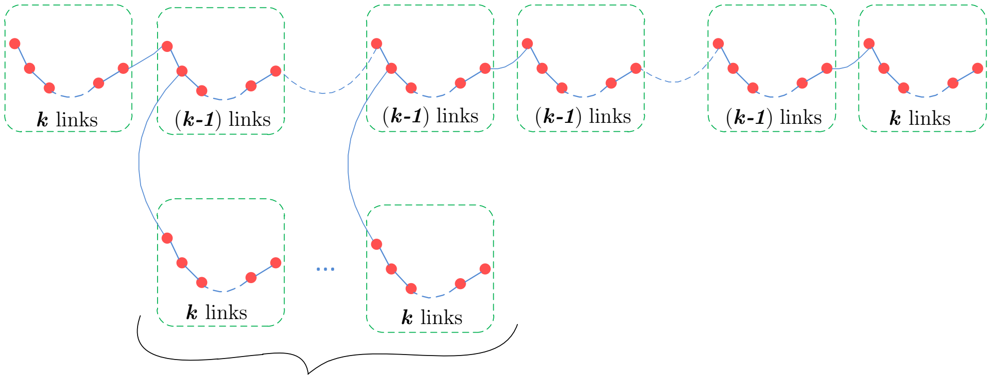


Fig. 4. (Color online) Maximum modular partitioning for $\lfloor \frac{c}{2} \rfloor < r \leq c-1$.

1. If $1 \leq r \leq \lfloor \frac{c}{2} \rfloor$, we have $1 \leq 2r \leq c$, hence
$$-(c-3) \leq 2r - q_{D+1} \leq c-1 \qquad (27)$$
and Inequality (27) holds only if $2r - q_{D+1} = 0$.
It is possible to construct a partitioning where $|Q_D| = c - 2r$ and $|Q_{D+1}| = 2r$. If we have the partitioning in Figure 3, where there is a *tree-configuration* such as $(r+1)$ communities are *leafs* (with 1 *inter-connecting* link) and having k internal links; $(r-1)$ communities have 3 *inter-connecting* and $(k-1)$ internal links and $(c-2r)$ communities have 2 *inter-connecting* and $(k-1)$ internal links. The cumulative degree of the first two groups $((r+1)+(r-1) = 2r$ in total-set $Q_{D+1})$ is $2(k-1)+3 = 2k+1$. The cumulative degree of the third group $((c-2r)$ in total-set $Q_D)$ is $2(k-1)+2 = 2k$. The total number of links in this *tree-configuration* is $(r+1)k + (r-1)(k-1) + (c-2r)(k-1) + (c-1) = ck + r = L$
In this case $S_m = \frac{r(c-2r)}{2cL^2}$.
2. If $\lfloor \frac{c}{2} \rfloor + 1 \leq r \leq c-1$, we have $2 \leq 2r - c \leq c-2$ and we can rewrite equation (25) as
$$D = 2k + 1 + \frac{2r - c - q_{D+1}}{c}. \qquad (28)$$

Thus, D is integer only in the case where $q_{D+1} = 2r - c$. It is possible to construct a partitioning where $|Q_D| = 2r$ and $|Q_{D+1}| = c - 2r$. The partitioning in Figure 4 is a *tree-configuration*, where $(c-r+1)$ communities are *leafs* (1 inter-connecting link) having k internal links; $(c-r-1)$ communities have 3 inter-connecting and $(k-1)$ internal links and $2r$ communities have 2 inter-connecting and k internal links. The cumulative degree of the first two groups $((c-r+1)+(c-r-1) = 2c-2r$ in total-set $Q_D)$ is $2(k-1)+3 = 2k+1$. The cumulative degree of the third group $((2r-c)$ in total-set $Q_{D+1})$ is $2k+2$. The total number of links in this *tree-configuration* is $(c-r+1)k + (c-r-1)(k-1) + (2r-c)k + (c-1) = ck + r = L$
In this case $S_m = \frac{(c-r)(2r-c)}{2cL^2}$. □

In the next Lemma we consider the remaining case, L is divisible by c $(r = 0)$.

Lemma 6. *If $c \mid L$ then the minimum value of S_m is $\frac{1}{2L^2}$.*

Proof. Let us assume that $q_{D-1} = 0$. Using $L = ck$, we derive
$$(c - q_{D+1})D + q_{D+1}(D+1) = 2L.$$

Hence,
$$D = 2k - \frac{q_{D+1}}{c}.$$

Because $1 \leq q_{D+1} \leq c-1$ it is not possible for D to be an integer. Hence, for $q_{D-1} = 0$ is not possible to construct a

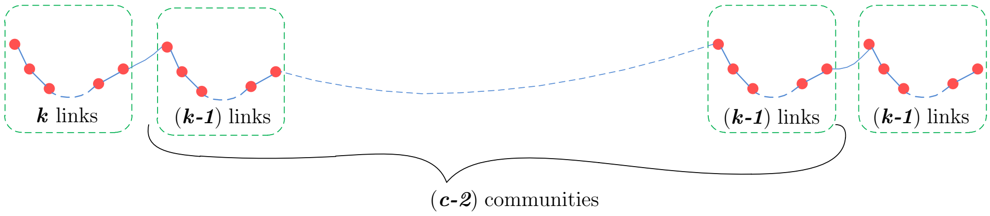


Fig. 5. (Color online) Maximum modular partitioning for $r = 0$.

partitioning. Thus $q_{D-1} \geq 1$ and by reasons of symmetry $q_{D+1} \geq 1$. If we use equation (23), we have

$$S_m = \frac{q_D (c - q_D) + 4q_{D-1} \cdot q_{D+1}}{4cL^2} \geq \frac{q_D (c - q_D) + 4}{4cL^2}. \tag{29}$$

It is possible to construct the case, where $q_{D-1} = q_{D+1} = 1$ and $q_D = c - 2$, as illustrated in Figure 5. The partitioning in Figure 5 consists of a very left *leaf* community having k internal links (set Q_{D+1}), which implies that the cumulative degree is $2k+1$; the very right community is a *leaf* and has $(k-1)$ internal links, which implies that the cumulative degree is $2k-1$ (set Q_{D-1}) and the $(c-2)$ central communities have 2 *inter-connecting* links and $(k-1)$ internal links, which implies that the cumulative degree is $2k$ (set Q_D).

Thus, equation (23) is reduced to

$$S_m = \frac{2^2 + 2^2 + 2(c-2) \cdot 1 + 2(c-2) \cdot 1}{8cL^2} = \frac{1}{2L^2}$$

and equality in inequality (29) is reached. □

In the previous lemmas we have examined all possible cases of the residue r in $L \equiv r \,(\mathrm{mod}\ c)$. The general, combined result is given in the next theorem.

Theorem 1. *The maximum modular connected graph with L links and partitioned into c communities, where $L \equiv r \,(\mathrm{mod}\ c)$ has modularity*

$$m = 1 - \frac{1}{c} - \frac{c-1}{L} - \begin{cases} \frac{1}{2L^2}, & r = 0 \\ \frac{r(c-2r)}{2cL^2}, & 1 \leq r \leq \left\lfloor \frac{c}{2} \right\rfloor \\ \frac{(c-r)(2r-c)}{2cL^2}, & \left\lfloor \frac{c}{2} \right\rfloor < r \leq c-1. \end{cases}$$

Proof. Lemma 1 shows that in the maximum modular connected graph always $L_{\mathrm{inter}} = c - 1$. The values for S_m for maximum modular graphs are determined in Lemmas 5 and 6. □

For all connected graphs, we can state the following corollary.

Corollary 3. *The maximum modularity of any connected graph with L links and partitioned into c communities, where $L \equiv r \,(\mathrm{mod}\ c)$ is bounded by*

$$m \leq 1 - \frac{1}{c} - \frac{c-1}{L} - \begin{cases} \frac{1}{2L^2}, & r = 0 \\ \frac{r(c-2r)}{2cL^2}, & 1 \leq r \leq \left\lfloor \frac{c}{2} \right\rfloor \\ \frac{(c-r)(2r-c)}{2cL^2}, & \left\lfloor \frac{c}{2} \right\rfloor < r \leq c-1 \end{cases}$$

5 Maximum modular graph for a given L, N and c

In this section, we consider the additional constraints that number of nodes N in the graph is also fixed. This is a subcase of the problem considered in Section 4, which implies that the maximum modular graph has modularity smaller than or equal to the modularity of the maximum modular graph obtained in Theorem 1. As demonstrated in this section, it is not always possible to reach the maximum modularity obtained in previous section when the number of links L and the number of communities c is fixed.

If we have a connected graph with N nodes and L links, we have the inequality

$$N - 1 \leq L \leq \binom{N}{2} \tag{30}$$

where the left-hand side equality is achieved when the graph is a tree and the right-hand side equality is achieved when the graph is a *complete graph*. Otherwise the original graph is not connected. If we write L and N as

$$\begin{cases} L = c \cdot k_L + r_L \\ N = c \cdot k_N + r_N \end{cases} \tag{31}$$

where the residues $r_L, r_N \in \{0, 1, \ldots, c-1\}$. Based on the left inequality in inequality (30), we have

$$r_N - r_L - 1 \leq c\,(k_L - k_N).$$

Because $r_N - r_L - 1 \geq 0 - (c-1) - 1 = -c$ we obtain $k_L - k_N \geq -1$.

1. The equality $(k_L = k_N - 1)$ is possible if and only if $r_L = c - 1$ and $r_N = 0$. Now we can place k_L links and k_N nodes in each community, forming a *tree-configuration*. Because $k_L = k_N - 1$ each community internally is a tree graph. The maximum modularity for $r_L = c - 1$, mentioned in the Theorem 1, is

$$m = 1 - \frac{1}{c} - \frac{c-1}{L} - \frac{c-2}{2cL^2}.$$

2. If $k_L = k_N$, based on the left inequality in inequality (30), we obtain

$$r_L \geq r_N - 1. \tag{32}$$

According to Lemma 5 in the maximum modular configurations there are $(c - r_L - 1)$ communities with

$(k_L - 1)$ links and $(r_L + 1)$ communities with k_L links. Here, we can place exactly k_N nodes in each community with $(k_L - 1)$ links. It is possible because, based on inequality (32), $c - r_L - 1 \leq c - r_N$. In r_N communities with k_L links we can place $(k_N + 1)$ nodes and k_N nodes in the remaining $(r_L + 1 - r_N)$ communities. In the final partitioning, we have communities with

(a) $(k_L - 1)$ internal links and k_N nodes,
(b) k_L internal links and k_N nodes,
(c) k_L internal links and $(k_N + 1)$ nodes.

Hence, the modularity obtained in Theorem 1 is achieved again.

3. If $k_L \geq k_N + 1$, then it is also possible to achieve the modularity from Theorem 1. According to the relations in equation (31), there is at least one:
   - community with at least $(k_N + 1)$ nodes if $r_N > 0$ or at least k_N nodes if $r_N = 0$,
   - community with at least k_L internal links.

   We can place k_N nodes in $(c - r_N)$ communities and $(k_N + 1)$ nodes in r_N communities. In this case, we have possibilities (a), (b) and (c) from the above, and in addition communities with (d) $(k_L - 1)$ internal links and $(k_N + 1)$ nodes are also possible.
   The communities from (a), (b), (c) and (d) should be connected graphs. The left inequality in inequality (30) is always satisfied. Indeed, for the right-hand side, the conditions for the connected graphs are

$$\text{(a) } k_L - 1 \leq \binom{k_N}{2}, \text{ (b) } k_L \leq \binom{k_N}{2},$$
$$\text{(c) } k_L \leq \binom{k_N + 1}{2}, \text{ (d) } k_L - 1 \leq \binom{k_N + 1}{2}. \quad (33)$$

   Based on inequality (33) we have two possibilities

   (a) If $r_L + 1 > r_N$ and $k_L \leq \binom{k_N}{2}$, we can construct: r_N communities each with k_L internal links could be placed $(k_N + 1)$ nodes; $(r_L - r_N + 1)$ communities each with k_L internal links could be placed k_N nodes; and $(c - r_L - 1)$ communities each with $(k_L - 1)$ internal links could be placed k_N nodes. In this case we have communities as in (a), (b) and (c). The existence condition for (b) given in inequality (33) is sufficient.
   (b) If $r_L + 1 \leq r_N$ and $k_L - 1 \leq \binom{k_N}{2}$, we can construct: $(r_L + 1)$ communities each with k_L internal links could be placed $(k_N + 1)$ nodes; $(r_N - r_L - 1)$ communities each with $(k_L - 1)$ internal links could be placed $(k_N + 1)$ nodes; and $(c - r_N)$ communities each with $(k_L - 1)$ internal links could be placed k_N nodes. In this case we have communities as in (a), (c) and (d). The existence condition for (a) given in inequality (33) is sufficient.
   In the cases 1 and 2 the modularity of the maximum modular graph will be the value in Theorem 1 and in all the other cases, the modularity of the maximum modular graph will be smaller. The main obstacle is that there is no space for additional links. There are two possible strategies for placing a new link:
   - re-placing a node from one community to another;
   - adding an *inter-connecting* link.

   According to Lemma 3 there is no general rule that the modularity increases/decreases by both steps. The increase of the modularity in each case is conditional. The general direction is that replacing a node from one community to another holds until the community, where the node will be placed, has a small enough cumulative degree. Therefore, only adding *inter-connecting* links will be possible.

5.1 The two communities case ($c = 2$)

In the case of dividing the graph into two communities, there is no *general rewiring*. When $c = 2$, we calculate the modularity of the maximum modular graph for a fixed number L of links and number N of nodes. We consider different cases for the residues of L and N, divided by $c = 2$. We already calculated in the general case for c the modularity of the maximum modular graph for smaller $L \leq 2\binom{\lfloor N/2 \rfloor}{2}$.

(a) if $k_L \leq \binom{k_N}{2} \iff L \leq 2\binom{\lfloor N/2 \rfloor}{2}$, placing k_N nodes in both communities and $(k_L - 1)$ links in one community and k_L in another (1 *inter-connecting* link) the modularity of the maximum modular graph is

$$m = \frac{1}{2} - \frac{1}{L} - \frac{1}{2L^2};$$

(b) if $L \geq 2\binom{N/2}{2} + 1$, we consider several cases for

$$L \in \left[2\binom{\lfloor N/2 \rfloor}{2} + (k-1)^2 + 2, 2\binom{\lfloor N/2 \rfloor}{2} + k^2 + 1\right] \quad (34)$$

where $k = 1, \ldots, \lfloor N/2 \rfloor - 1$. Lemma 1 states that the maximum modular graph has 1 *inter-connecting* link.

The exact result is determined in the remaining part of this section.

Lemma 7. *If N_{C_1} and N_{C_2} ($N_{C_1} \leq N_{C_2}$) are the number of nodes and L_{C_1} and L_{C_2} the number of internal links in the communities C_1 and C_2, respectively then*

(a) $L_{C_1} \leq L_{C_2}$
(b) k_L the modularity $m(L_{C_1})$ is increasing function
(c) $N_{C_1} \leq N - k$, if L belongs to the interval (34)

Proof. Details are given in Appendix A.4 □

Corollary 4. *If L lies in the interval (34), then*

$$N_{C_1} = \left\lfloor \frac{N}{2} \right\rfloor - k$$
$$L_{C_1} = \binom{\lfloor \frac{N}{2} \rfloor - k}{2}.$$

We use the term *bi-modular* for partitioning into 2 communities ($c = 2$).

Lemma 8. *The maximum bi-modular graph with L links and N nodes, when L is in the interval (34) is*

$$m = \frac{1}{2} - \frac{1}{L} - \frac{\left(L-1-2\binom{\lfloor \frac{N}{2}\rfloor - k}{2}\right)^2}{2L^2}.$$

Proof. Corollary 4 shows that $L_{C_1} = \binom{\lfloor \frac{N}{2}\rfloor - k}{2}$ and $L_{\text{inter}} = 1$, hence, $L_{C_2} = L - 1 - \binom{\lfloor \frac{N}{2}\rfloor - k}{2}$. The definition in equation (4) leads to

$$m = \frac{1}{2} - \frac{1}{L} - \frac{\left(L-1-2\binom{\lfloor \frac{N}{2}\rfloor - k}{2}\right)^2}{2L^2}.$$

It remains to prove the existence of a graph with $N_{C_2} = \left[\frac{N}{2}\right] + k$ nodes and L_{C_2} links. We have

$$\begin{aligned} L_{C_2} &= L - 1 - L_{C_1} \\ &\le 2\binom{\lfloor N/2\rfloor}{2} + k^2 + 1 - 1 - \binom{\lfloor \frac{N}{2}\rfloor - k}{2} \\ &= \binom{\lfloor \frac{N}{2}\rfloor + k}{2}. \end{aligned} \tag{35}$$

On the other hand

$$\begin{aligned} L_{C_2} &= L - 1 - L_{C_1} \\ &\ge 2\binom{\lfloor N/2\rfloor}{2} + (k-1)^2 + 2 - 1 - \binom{\lfloor \frac{N}{2}\rfloor - k}{2} \\ &= 1 + \binom{\lfloor N/2\rfloor - k + 1}{2} - \binom{\lfloor \frac{N}{2}\rfloor - k}{2} \\ &\quad + \binom{\lfloor N/2\rfloor + k - 1}{2} \\ &= \binom{\lfloor \frac{N}{2}\rfloor + k - 1}{2} + \left\lfloor \frac{N}{2}\right\rfloor - k + 1 \\ &= \left\lfloor \frac{N}{2}\right\rfloor + k - 1 + \frac{\left(\lfloor \frac{N}{2}\rfloor + k - 2\right)\left(\lfloor \frac{N}{2}\rfloor + k - 3\right)}{2} \\ &\ge \binom{\left\lfloor \frac{N}{2}\right\rfloor + k}{} - 1. \end{aligned} \tag{36}$$

Based on inequalities (36) and (35) and the condition in inequality (30), it is possible to construct a graph with $\left\lfloor\frac{N}{2}\right\rfloor + k$ nodes and L_{C_2} nodes and the proof is complete. □

The next Lemma 9 resolves the problem of finding the modularity of a maximum modular connected graph, when L is significantly larger than N.

Lemma 9. *If* $L \in \left[2 + \binom{N-1}{2}, \binom{N}{2}\right]$ *the modularity of a maximum bi-modular graph* ($c = 2$) *is*

$$m = -\frac{1}{2}\left(1 - \frac{\binom{N-1}{2}}{L}\right)^2.$$

Proof. Details are given in Appendix A.5. □

An immediate corollary of the last Lemma 9 is

Corollary 5. *In the complete and bi-modular graph, the maximum modularity is* 0.

We conclude the discussion by the following Theorem 2.

Theorem 2. *The modularity of the maximum modular and connected graph with N nodes and L links divided into* $c = 2$ *communities is*

$$m = \begin{cases} \frac{1}{2} - \frac{1}{L} - \frac{1}{2L^2}, \text{if} N-1 \le L \le 2\binom{\lfloor \frac{N}{2}\rfloor}{2} + 1 \\ \frac{1}{2} - \frac{1}{L} - \frac{W^2}{2L^2}, \text{if } \ 2\binom{\lfloor \frac{N}{2}\rfloor}{2} + 2 \le L \le 1 + \binom{N-1}{2} \\ -\frac{1}{2}\left(1 - \frac{\binom{N-1}{2}}{L}\right)^2, \text{ if } \ 2 + \binom{N-1}{2} \le L \le \binom{N}{2} \end{cases}$$

where $W = L - 1 - 2\left(\begin{matrix}\lfloor \frac{N}{2}\rfloor - 1 - \left\lfloor\sqrt{L - 2 - 2\binom{[N/2]}{2}}\right\rfloor \\ 2\end{matrix}\right)$

Proof. The results for $N - 1 \le L \le 2\binom{\lfloor \frac{N}{2}\rfloor}{2} + 1$ and $2 + \binom{N-1}{2} \le L \le \binom{N}{2}$, we already obtained. For $L \le N-2$, there is no connected graph. For $L \ge \binom{N}{2} + 1$ there is no graph without self or multiple links between nodes. Lemma 8 shows that

$$m = \frac{1}{2} - \frac{1}{L} - \frac{\left(L-1-2\binom{\lfloor \frac{N}{2}\rfloor - k}{2}\right)^2}{2L^2} \tag{37}$$

for

$$2\binom{\lfloor N/2\rfloor}{2} + (k-1)^2 + 2 \le L \le 2\binom{\lfloor N/2\rfloor}{2} + k^2 + 1$$

which implies

$$k \le 1 + \sqrt{L - 2\binom{\lfloor N/2\rfloor}{2} - 2} \le 1 + \sqrt{k^2 - 1} < k + 1.$$

Hence $k = 1 + \left\lfloor\sqrt{L - 2 - 2\binom{\lfloor N/2\rfloor}{2}}\right\rfloor$. If we use this result into Lemma 8 we will obtain the result stated in the theorem. □

6 Applications

The theoretical upper bounds determined in the paper can be used in transportation planning. Suppose that the organization responsible for world transport connections needs to make a plan and the condition is that it should spend

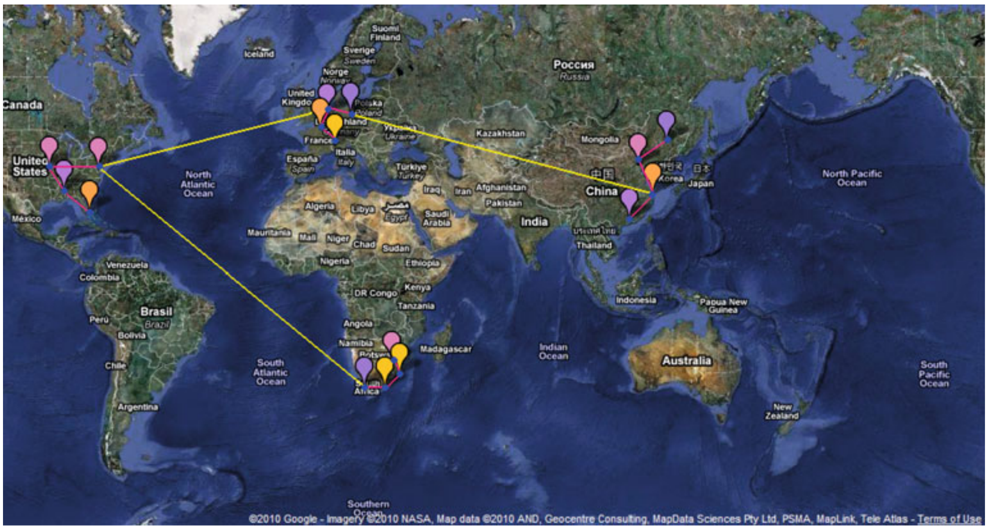


Fig. 6. (Color online) Constructing maximum modular world connections (©Google Inc.).

money for establishing exactly L lines in exactly c countries. It should be possible from each place to reach every other place, but also the cost of inter-continental lines should be minimized. Moreover, for political and diplomatic reasons, the distribution of used money per country should be fair and approximately equal. In practice, we suppose that the organization should connect Europe, the United States, China and South Africa, and should spend an equal amount of money for connecting exactly 4 cities in those countries or continent. This is a problem of constructing a maximum modular network of cities as nodes and countries/continent as communities. Figure 6 gives a construction of a maximum modular network, given $c = 4$ number of countries/continents and $L = 19$, number of connections that should be established.

Theorem 1 and the fact that $19 \equiv 3 \pmod 4$ shows that the maximum modularity is ($r = c-1, c = 4, L = 19$)

$$m = 1 - \frac{1}{4} - \frac{c-1}{L} - \frac{1\,(2c-2-c)}{2cL^2} \approx 0.5914.$$

The communities are arranged according to Lemma 5, such that each country/continent has 4 internal connections and only the 2 *central* communities, Europe and US, have 2 external connections. Their cumulative degree is 10, and those of China and South Africa are 9.

Another possible application is designing a distributed data-center to keep redundant copies which will increase the robustness under malicious attacks. The additional constraint consists of preserving the connectivity and communication between the units. If we want to design 3 identical redundant parts, apart from the increase in security and reliability level, additionally the maximum modularity is achieved over all the networks with the same number of direct connections. For instance, a topological plan of a data-center with 3 redundant parts each with unique control unit and 8 functional units connected in a *star-configuration* structure is drawn in Figure 7. According to Theorem 1 and the fact that $26 \equiv 2 (\mathrm{mod}\ 3)$, the maximum modularity of $m = 0.5895$ is achieved

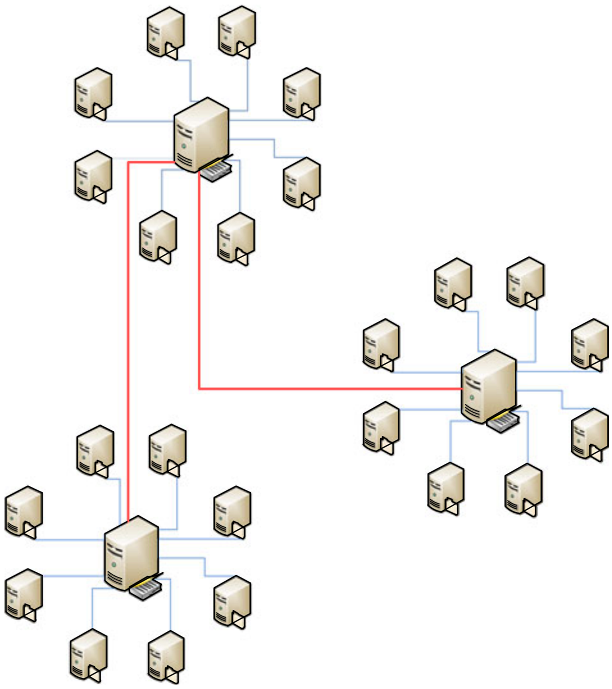


Fig. 7. (Color online) Reliable data-redundant and maximum modular data-center.

$$m = 1 - \frac{1}{3} - \frac{c-1}{L} - \frac{(c-2)\,(2\times 2-c)}{2cL^2} \approx 0.5895$$

7 Conclusion

We introduced and described the notation of *maximum modular graph*, which is a graph in the set of all graphs satisfying given constraint(s) that has a maximum modularity. Using the Cauchy identity [16] we found an alternative form of Newman's modularity [11] that is suitable to deduce upper bounds. Considering the divisibility of the number L of links over the number c of communities, exact upper bounds are given. For a given number L of links and for a given number L of links and N of nodes, we derive the *attainable* value for the upper bound, which means that there is at least one graph and partitioning where that maximum modularity is reached. We also propose two possible applications for maximum modular partitioning of a transportation network and maximum modular topology of reliable data-center.

An interesting problem is finding the maximum modularity for a fixed degree vector $d = \left[d_1\ d_2\ \dots\ d_N\right]$. Because of the basic law of the degrees, $\sum_{k=1}^{c} d_i = 2L$, we have a fixed number of links and because of the size of the degree vector d, the number of nodes is also fixed. This leads to the conclusion that the maximum modularity is smaller than or equal to the result obtained in Theorem 1. However, the problem is not equivalent. Having the fixed number of neighbors for each node this problem could be considered as a problem of grouping nodes in the communities to maximize a multi-variable function. The defined steps for *local/global re-wiring* and *shifting* do not work in this case, because they change the nodal degrees so that we need to consider a degree preserving rewiring process [17].

We would like to thank Wynand Winterbach for his useful suggestions.

Appendix: The derivations used for the proofs for m′ > m

According to the definition of modularity, the difference after and before the change we have

$$m' - m = \frac{L_{\text{inter}} - L'_{\text{inter}}}{L} - \frac{1}{8cL^2}\sum_{p=1}^{c}\sum_{k=1}^{c}\left[\left(D_{C_p} - D_{C_k}\right)^2 - \left(D'_{C_p} - D'_{C_k}\right)^2\right]. \quad \text{(A.1)}$$

In equation (A.1), the expression $\left(D_{C_p} - D_{C_k}\right)^2 - \left(D'_{C_p} - D'_{C_k}\right)^2$ is different from 0, only if $p \in \{i, j\}$ or $k \in \{i, j\}$. Using equation (A.1), we obtain

$$\begin{aligned} m' - m = {} & \frac{L_{\text{inter}} - L'_{\text{inter}}}{L} \\ & - \frac{2}{8cL^2}\sum_{k=1}^{c}\Big[\left(D'_{C_i} - D'_{C_k}\right)^2 - \left(D_{C_i} - D_{C_k}\right)^2 \\ & + \left(D'_{C_j} - D_{C_k}\right)^2 - \left(D_{C_j} - D_{C_k}\right)^2\Big] \\ = {} & \frac{L_{\text{inter}} - L'_{\text{inter}}}{L} - \frac{1}{4cL^2} \\ & \times \sum_{\substack{k=1\\k\neq i,j}}^{c}\Big[\left(D_{C_i} + D'_{C_i} - 2D_{C_k}\right)\left(D'_{C_i} - D_{C_i}\right) \\ & + \left(D_{C_j} + D'_{C_j} - 2D_{C_k}\right)\left(D'_{C_j} - D_{C_j}\right)\Big] \\ & - \frac{1}{4cL^2}\left(D'_{C_i} - D_{C_i} - \left(D'_{C_j} - D_{C_j}\right)\right) \\ & \times \left(D'_{C_i} + D_{C_i} - \left(D'_{C_j} + D_{C_j}\right)\right). \end{aligned} \quad \text{(A.2)}$$

A.1 Details of Lemma 1

In this case according to the condition in the Lemma 1 and the expression in equation (A.2)

$$\begin{aligned} m' - m = {} & \frac{1}{L} - \frac{1}{4cL^2}\sum_{\substack{k=1\\k\neq i,j}}^{c}\left[\left(2D_{C_i} - 2D_{C_k}\right) - \left(2D_{C_j} - 2D_{C_k}\right) + 2\right] \\ & - \frac{1}{4cL^2}\left(D_{C_i} - D_{C_i} - \left(D_{C_j} - D_{C_j}\right) + 2\right) \\ & \times \left(D_{C_i} + D_{C_i} - \left(D_{C_j} + D_{C_j}\right) + 2\right) \\ = {} & \frac{1}{L} - \frac{2}{4cL^2}\sum_{\substack{k=1\\k\neq i,j}}^{c}\left(D_{C_i} - D_{C_j} + 1\right) - \frac{2\left(2D_{C_i} - 2D_{C_j} + 2\right)}{4cL^2} \\ = {} & \frac{1}{L} - \frac{c-2}{2cL^2}\left(D_{C_i} - D_{C_j} + 1\right) - \frac{D_{C_i} - D_{C_j} + 1}{cL^2} \\ = {} & \frac{1}{L} - \frac{c-2+2}{2c \cdot L^2}\left(D_{C_i} - D_{C_j} + 1\right) \\ = {} & \frac{1}{L} - \frac{1}{2L^2}\left(D_{C_i} - D_{C_j} + 1\right) \\ = {} & \frac{2L - 1 - D_{C_i} + D_{C_j}}{2L^2}. \end{aligned}$$

A.2 Details of Lemma 2

Using equation (A.2) and the condition in the lemma one can get

$$\begin{aligned} m' - m = {} & -\frac{1}{4cL^2}\sum_{\substack{k=1\\k\neq i,j}}^{c}\Big[\left(D_{C_i} + D'_{C_i} - 2D_{C_k}\right)\left(D'_{C_i} - D_{C_i}\right) \\ & + \left(D_{C_j} + D'_{C_j} - 2D_{C_k}\right)\left(D'_{C_j} - D_{C_j}\right)\Big] \\ & - \frac{1}{4cL^2}\left(D'_{C_i} - D_{C_i} - \left(D'_{C_j} - D_{C_j}\right)\right) \\ & \times \left(D'_{C_i} + D_{C_i} - \left(D'_{C_j} + D_{C_j}\right)\right) \\ = {} & \frac{4}{4cL^2}\sum_{\substack{k=1\\k\neq i,j}}^{c}\left[\left(D_{C_i} - D_{C_k} - 1\right) - \left(D_{C_j} - D_{C_k} + 1\right)\right] \\ & + \frac{4}{4cL^2}\left(2D_{C_i} - 2 - 2D_{C_j} - 2\right) \\ = {} & \frac{1}{cL^2}\sum_{\substack{k=1\\k\neq i,j}}^{c}\left(D_{C_i} - D_{C_j} - 2\right) + \frac{2}{cL^2}\left(D_{C_i} - D_{C_j} - 2\right) \\ = {} & \frac{c-2+2}{cL^2}\left(D_{C_i} - D_{C_j} - 2\right) = \frac{1}{L^2}\left(D_{C_i} - D_{C_j} - 2\right). \end{aligned}$$

A.3 Details of Lemma 3

Using equation (A.2) and the condition in the lemma one can get

$$\begin{aligned} m' - m = &\frac{1}{L} - \frac{1}{4cL^2} \sum_{\substack{p=1\\ p\neq i,j,k}}^{c} \left[\left(D_{C_i} + D'_{C_i} - 2D_{C_p}\right)\left(D'_{C_i} - D_{C_i}\right) \right. \\ &+ \left(D_{C_j} + D'_{C_j} - 2D_{C_p}\right)\left(D'_{C_j} - D_{C_j}\right) \\ &\left. + \left(D_{C_k} + D'_{C_k} - 2D_{C_p}\right)\left(D'_{C_k} - D_{C_k}\right)\right] \\ &- \frac{1}{4cL^2}\left(D'_{C_i} - D_{C_i} - \left(D'_{C_j} - D_{C_j}\right)\right) \\ &\times \left(D'_{C_i} + D_{C_i} - \left(D'_{C_j} + D_{C_j}\right)\right) \\ &- \frac{1}{4cL^2}\left(D'_{C_i} - D_{C_i} - \left(D'_{C_k} - D_{C_k}\right)\right) \\ &\times \left(D'_{C_i} + D_{C_i} - \left(D'_{C_k} + D_{C_k}\right)\right) \\ &- \frac{1}{4cL^2}\left(D'_{C_j} - D_{C_j} - \left(D'_{C_k} - D_{C_k}\right)\right) \\ &\times \left(D'_{C_j} + D_{C_j} - \left(D'_{C_k} + D_{C_k}\right)\right) \\ = &\frac{1}{L} + \frac{1}{4cL^2} \sum_{\substack{p=1\\ p\neq i,j,k}}^{c} \left[\left(2D_{C_i} - 2D_{C_p} - 1\right) \right. \\ &+ \left(2D_{C_j} - 2D_{C_p} - 1\right) \\ &\left. - 2\left(2D_{C_k} - 2D_{C_p} + 2\right)\right] \\ &- \frac{1}{4cL^2}(-1-2)\left(2D_{C_i} - 1 - \left(2D_{C_k} + 2\right)\right) \\ &- \frac{1}{4cL^2}(-1-2)\left(2D_{C_j} - 1 - \left(2D_{C_k} + 2\right)\right) \\ = &\frac{1}{L} + \frac{1}{2cL^2} \sum_{\substack{p=1\\ p\neq i,j,k}}^{c} \left(D_{C_i} + D_{C_j} - 2D_{C_k} - 3\right) \\ &+ \frac{3}{4cL^2}\left(2D_{C_i} + 2D_{C_j} - 4D_{C_k} - 6\right) \\ = &\frac{1}{L} + \frac{2c-6+6}{4cL^2}\left[D_{C_i} + D_{C_j} - 2D_{C_k} - 3\right] \\ = &\frac{2L + D_{C_i} + D_{C_j} - 2D_{C_k} - 3}{2L^2}. \end{aligned}$$

A.4 Details of the Lemma 7

If $c = 2$ and the number of links is L for maximum modularity we have the expression

$$m(L_{C_1}) = \frac{1}{2} - \frac{1}{L} - \frac{\left(L - 1 - 2L_{C_1}\right)^2}{2L^2}$$

(a) if $L_{C_1} > L_{C_2}$ and using the fact that number of nodes in C_1 is $N_{C_1} < \lfloor \frac{N}{2} \rfloor$ we have

$$\begin{aligned} L_{C_1} &< \binom{\lfloor \frac{N}{2} \rfloor}{2} \\ L &= L_{C_1} + L_{C_2} + 1 < 2L_{C_1} + 1 < 2\binom{\lfloor \frac{N}{2} \rfloor}{2} + 1. \end{aligned} \tag{A.3}$$

The inequality in (A.3) is in contradiction with (34)

(b) if we derive the modularity as a function of L_{C_1} we have

$$\begin{aligned} \frac{dm(L_{C_1})}{dL_{C_1}} &= -\frac{2\left(L - 1 - 2L_{C_1}\right)(-2)}{2L^2} \\ &= \frac{2\left(L - 1 - 2L_{C_1}\right)}{L^2}. \end{aligned}$$

In (a) we have just proved that $L_{C_1} < L_{C_2} \Longrightarrow L = L_{C_1} + L_{C_2} + 1 > 2L_{C_1} + 1 \Longrightarrow (L - 1 - 2L_{C_1} > 0)$ and $m(L_{C_1})$ is increasing function;

(c) because of the Lemma 1 and (a) if the number of links in the community C_1 is smaller than $\binom{N_{C_1}}{2}$, the modularity will increase if we *local re-wire* one link from the community C_2 into C_1. Hence, $L_{C_1} = \binom{N_{C_1}}{2}$. For the number of links in the community C_2 we have $L_{C_2} \leq \binom{N - N_{C_1}}{2}$. If $\lfloor \frac{N}{2} \rfloor > N_{C_1} \geq \lfloor \frac{N}{2} \rfloor - k + 1$ then

$$\begin{aligned} L &= L_{C_1} + L_{C_2} + 1 \leq \binom{N_{C_1}}{2} + \binom{N - N_{C_1}}{2} + 1 \\ L &\leq 2\binom{\lfloor \frac{N}{2} \rfloor}{2} + \left(\left\lfloor \frac{N}{2} \right\rfloor - N_{C_1}\right)^2 + 1 \\ &\leq 2\binom{\lfloor \frac{N}{2} \rfloor}{2} + (k-1)^2 + 1. \end{aligned}$$

This is in contradiction with the interval (34).

A.5 Details of Lemma 9

Based on the Lemma 1 the maximum modularity is obtained in the case where the number of *inter-connecting* links is minimal. It is in the case $N_{C_1} = 1$ and $N_{C_2} = N - 1$ and C_2 is a full clique i.e. $L_{C_2} = \binom{N-1}{2}$. Clearly there are no links in the communities C_1 ($L_{C_1} = 0$), thus $L_{\text{inter}} = L - L_{C_2} = L - \binom{N-1}{2}$. In this case for modularity we obtain

$$\begin{aligned} m(L_{C_1}) &= \frac{1}{2} - \frac{L - \binom{N-1}{2}}{L} - \frac{\binom{N-1}{2}^2}{2L^2} \\ &= -\frac{1}{2}\left(1 - 2\frac{\binom{N-1}{2}}{L} + \left(\frac{\binom{N-1}{2}}{L}\right)^2\right) \\ &= -\frac{1}{2}\left(1 - \frac{\binom{N-1}{2}}{L}\right)^2. \end{aligned}$$

References

1. M.E.J. Newman, M. Girvan, Phys. Rev. E **69**, 026113 (2004)
2. U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, D. Wagner, in *Graph-Theoretic Concepts in Computer Science*, *Lecture Notes in Computer Science* (Springer, Berlin/Heidelberg, 2007), Vol. 4769, Chap. 12, pp. 121–132.
3. M.E.J. Newman, Phys. Rev. E **69**, 066133 (2004)
4. R. Guimerà, L.A.N. Amaral, Nature **433**, 895 (2005)
5. R. Guimerà, M. Sales-Pardo, L.A.N. Amaral, Phys. Rev. E **70**, 025101 (2004)
6. J. Duch, A. Arenas, Phys. Rev. E **72**, 027104 (2005)
7. S. Fortunato, Phys. Rep. **486**, 75 (2010)
8. P. Van Mieghem, X. Ge, P. Schumm, S. Trajanovski, H. Wang, Phys. Rev. E **82**, 056113 (2010)
9. S. Fortunato, M. Barthélemy, Proc. Natl. Acad. Sci. **104**, 36 (2007)
10. M. Fiedler, Czechoslovak Math. J. **23**, 298 (1973)
11. M.E.J. Newman, Proc. Natl. Acad. Sci. **103**, 8577 (2006)
12. P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, 2010)
13. A. Clauset, M.E.J. Newman, C. Moore, Phys. Rev. E **70**, 066111 (2004)
14. G. Agarwal, D. Kempe, Eur. Phys. J. B **66**, 409 (2008)
15. J.-C. Delvenne, S.N. Yaliraki, M. Barahona, Proc. Natl. Acad. Sci. **107**, 12755 (2010)
16. A. Cauchy, *Cours d'Analyse de l'Ecole Royale Polytechnique: Analyse Algébrique*. Debure (reissued by Cambridge University Press, Cambridge, 2009), p. 1821
17. P. Van Mieghem, H. Wang, X. Ge, S. Tang, F.A. Kuipers, Eur. Phys. J. B **76**, 643 (2010)