Spectral inequalities for the modularity of a graph

Francesco Tudisco
Rome-Moscow school 2014

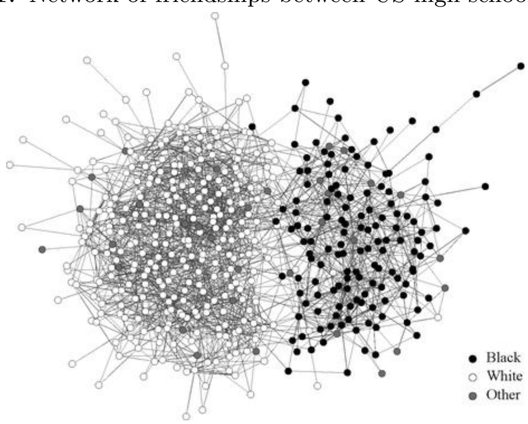
In this short lecture we would like to propose some inequalities that use the eigenproperties of modularity matrices to analyze the community structure of a network. Complex networks appear in a large variety of applications, important examples are the World Wide Web, biological networks like food webs or protein protein interaction networks, social networks, communication networks, and many other. Loosely speaking a complex network is a graph which occur in real life, therefore we need some standard graph theory concepts. Let me fix some preliminary notation and definition:

A complex network is modelled by a graph G and is assumed for the sake of simplicity to be **finite, simple, undirected, unoriented and loop-free**. It turns out that G is the pair (V, E) where V is the set of n nodes of the graph and $E \subseteq V \times V$ is the set of edges. A natural bijection there exists between graphs and the cone of entrywise nonnegative matrices, indeed to any such G we can associate a unique nonnegative adjacency matrix A. For our discussion, in particular, A will be a symmetric matrix with boolean entries, i.e. a_{ij} and a_{ji} are either 1, if nodes i and j are joined by an edge, or 0, otherwise.

1 Community detection

The discover and description of communities in a network is a central problem in modern graph analysis. Observations on real life graphs reveal that complex networks are intrinsically divided into groups. Think for instance at a network describing the scientific collaborations between a set of researchers, everybody would expect it to be divided into groups depending on the scientific interests and the geographical distance among the individuals, for instance. Similarly for a social network describing the friendship relations between a group of people, see for example Figure 1 where the friendship relations between students of an high-school in the United States are represented. So given a generic graph describing some real life interaction, the community detection problem consists in discovering and revealing the groups (if any) in which it is subdivided.

Figure 1: Network of friendships between US high-school students



Given two sets $S, T \subseteq V$ let

$$e_{in}(S) = 2|\{ij \in E \mid i, j \in S\}| \qquad e(S,T) = |\{ij \in E \mid i \in S, j \in T\}|$$

An informal statement of the community detection problem is the following:

find the *natural* partition $S_1, \dots, S_m$ of V such that $e_{in}(S_i)$ contains *many* edges and $e(S_i, S_j)$ contains *few* edges, when $i \neq j$. $(*)$

Here I use the word "natural" to stress the fact that such partition is a property of the network itself. The size of each S_i and the number m are not specified and may vary significantly from one network to another.

Of course the problem in this form is not well posed, at least we need to formalize the concepts of "many" and "few". Let me devote the remaining part of the section to briefly recall how the modularity function is used to such aim.

Let $G = (V, E)$ be a given connected graph. Let A be its adjacency matrix and $\mathbb{1}$ be the vector of all ones. The degree vector of G is given by

$$d = A\mathbb{1} \qquad d_i = |\{ij \in E \mid j \in V\}|$$

Consider any subset $S \subseteq V$. The characteristic vector of S is denoted by $\mathbb{1}_S$ and is defined by

$$(\mathbb{1}_S)_i = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases},$$

whereas the volume of S is the sum of the degrees of the vertices in S, i.e.

$$\mathrm{vol}\, S = \sum_{i \in S} d_i = \mathbb{1}_S^T d = \mathbb{1}_S^T A \mathbb{1} = 2e_{in}(S) + e_{out}(S)$$

where $e_{out}(S)$ is a shorthand for $e(S, \overline{S})$. Correspondingly, $\mathrm{vol}\, G$ is commonly used to denote the volume of the whole set V.

Definition 1.1 *Let $S \subseteq V$, and let $G(S)$ be the induced subgraph. The principal submatrix of A corresponding to the vertices in S is the adjacecncy matrix of $G(S)$ and is denoted by $A(S)$.*
The modularity of S is

$$Q(S) = e_{in}(S) - \frac{(\mathrm{vol}\, S)^2}{\mathrm{vol}\, G}\,,$$

*and if $Q(S) > 0$, we say that $G(S)$ is a **module** in G.*

Simple algebraic manipulations can be used to show that $Q(S) = Q(\overline{S})$, indeed:

$$\begin{aligned} Q(S) &= \underbrace{\mathrm{vol}\, S - e_{out}(S)}_{e_{in}(S)} - \frac{(\mathrm{vol}\, S)^2}{\mathrm{vol}\, G} \\ &= \mathrm{vol}\, S \left(1 - \frac{\mathrm{vol}\, S}{\mathrm{vol}\, G}\right) - e_{out}(S) \\ &= \frac{\mathrm{vol}\, S \cdot \mathrm{vol}\, \overline{S}}{\mathrm{vol}\, G} - e_{out}(S) \end{aligned}$$

which is symmetric in S and $\overline{S}$.

The *modularity matrix* of G has been introduced in [New06a] as the following rank one perturbation of A:

$$M = A - \frac{1}{\mathrm{vol}\, G} dd^T.$$

With the help of this matrix we can express $Q(S)$ as the following quadratic form:

$$Q(S) = \mathbb{1}_S^T M \mathbb{1}_S. \tag{1}$$

Undoubtely, the modularity of a vertex set is one of the most efficient indicators of its consistency as a community in G. Indeed the usefulness of Definition 1.1 lies in the fact that, in practice, if $G(S)$ is a connected module whose *size is significant*, then it can be actually recognized as a community in G. Thus, for the sake of simplicity, we shall say that *a connected graph $G(S)$ such that $Q(S) > 0$ is a community in G.*

Definition 1.1 leads naturally to an efficiency measure of a partitioning of G into modules. Indeed, let $S_1, \dots, S_m$ be a partitioning of V into pairwise disjoint subsets. The *(normalized) modularity of* $S_1, \dots, S_m$ is defined as

$$q(S_1, \dots, S_k) = \sum_{i=1}^{k} q(S_i) = \sum_{i=1}^{k} \frac{Q(S_i)}{|S_i|}.$$

The normalization factor $1/\mathrm{vol}\, G$ is purely conventional and has been introduced in [New06a, NG04] for compatibility with previous works, to settle the value of $q(S)$ in a range independent on G and m.

The problem of **partitioning a graph into** an arbitrary number of **subrgaphs whose overall modularity is maximized** is known as the ***modularity-based community detection problem*** and is a possible (and very popular) rigorous formulation of the informal statement we have given in $(*)$. Of course such optimization problem is not easy and deserve a deep analysis.

First of all we should stress the fact that in this form the problem has not a unique solution. Second, we see that the quantities $Q(S)$ and $q(S_1, \dots, S_m)$ appears to be related with the spectrum of M, due to the variational characterization of the eigenvalues of a symmetric matrix. The first theorem that I want to discuss shows indeed that, if $\{S_1, \dots, S_m\}$ is a partition that maximizes the modularity function on G, then m is not larger than the number of positive eigenvalues of M, plus one. I shall discuss such result in the next section

Upper bound on the number of modules

Given two subsets $S, T \subseteq V$, let us define their joint modularity as

$$Q(S,T) = e(S,T) - \frac{\mathrm{vol}\, S \,\mathrm{vol}\, T}{\mathrm{vol}\, G}$$

and let us observe the following facts:

- Clearly, $Q(S,T) = Q(T,S)$ and $Q(S) = Q(S, \overline{S})$. Furthermore, we can express the joint modularity of S and T equivalently as

$$Q(S,T) = \mathbb{1}_S^T M \mathbb{1}_T.$$

- From the equation $(\mathbb{1}_S + \mathbb{1}_T)^T M(\mathbb{1}_S + \mathbb{1}_T) = \mathbb{1}_S^T M \mathbb{1}_S + \mathbb{1}_T^T M \mathbb{1}_T + 2\mathbb{1}_S^T M \mathbb{1}_T$ we have

$$Q(S \cup T) = Q(S) + Q(T) + 2Q(S,T).$$

  In particular, $Q(S,T) > 0$ if and only if $Q(S \cup T) > Q(S) + Q(T)$. It follows that if $\{S_1, \dots, S_m\}$ is a partition that maximizes the modularity it is necessary that the joint modularity of any two subset is non-positive, i.e. $Q(S_i, S_j) \leq 0$ for any $i \neq j$, otherwise we can increase the overall modularity by merging two subgraphs into one. Moreover, if $\{S_1, \dots, S_m\}$ is a partition that maximizes the modularity and it has minimal cardinality among all those partition that maximize q (recall that the optimal partition is not unique in general), then $Q(S_i, S_j) < 0$ for all $i \neq j$. As a consequence:

- If $\{S_1, \dots, S_m\}$ is a partition that maximizes the modularity, it has minimal cardinality, and it is made up entirely by modules, then the matrix

$$(L_Q)_{ij} = \begin{cases} Q(S_i) > 0 & i = j \\ Q(S_i, S_j) < 0 & i \neq j \end{cases}$$

  is symmetric and irreducible.

We are ready for the announced theorem:

Theorem 1.2 Let $\{S_1, \dots, S_k\}$ be a partition that maximizes the modularity, which has minimal cardinality, and which is made up entirely by modules. Then $k \leq \pi(M) + 1$, where $\pi(X)$ denotes the number of positive eigenvalues of a matrix X.

Proof. Consider the matrix $X = [\mathbb{1}_{S_1} \cdots \mathbb{1}_{S_k}]$. If L_Q is defined as above then

$$L_Q = X^T M X$$

Furthermore, L_Q is weakly diagonally dominant. Indeed,

$$\sum_{j=1}^{k} (L_Q)_{ij} = \mathbb{1}_{S_i}^T M \sum_{j=1}^{k} \mathbb{1}_{S_j} = \mathbb{1}_{S_i}^T M \mathbb{1} = 0$$

thus $(L_Q)_{ii} = -\sum_{j \neq i} (L_Q)_{ij}$ and $\mathbb{1}$ is an eigenvector with zero eigenvalue of L_Q. Now let

$$A_Q = (a_{ij})_{ij} = \begin{cases} -Q(S_i, S_j) = & i \neq j \\ 0 & \textit{otherwise} \end{cases}$$

It follows that L_Q is the Laplacian matrix of the weighted unoriented graph G_Q, whose adjacency matrix is A_Q. We deduce that L_Q is a symmetric positive semidefinite matrix, with a zero eigenvalue which is associated to the eigenvector $\mathbb{1}$.

Consider now the matrix

$$B = \alpha I - L_Q$$

and observe that for a sufficient large $\alpha > 0$ (say $\alpha > \max_i Q(S_i)$) it is entrywise nonnegative and irreducible. Hence, by Perron-Frobenius theory, its largest eigenvalue is simple. Since the eigenspaces of B and L_Q coincide, the zero eigenvalue of L_Q must be simple.

We deduce that $\pi(L_Q) = k - 1$, i.e. L_Q has exactly $k - 1$ positive eigenvalues. Let us finally observe that this implies that M has at least $k - 1$ positive eigenvalues.

Given any vector v let me introduce the notation $\overline{v}$ to denote its normalized counterpart, i.e. $\overline{v} = v/\|v\|$. So that $\overline{\mathbb{1}}_{S_i} = \mathbb{1}_{S_i}/\|\mathbb{1}_{S_i}\| = \mathbb{1}_{S_i}|S_i|^{-1}$. Consider now any set of $n - k$ real unitary vectors $y_{k+1}, \dots, y_n$ such that $\{\overline{\mathbb{1}}_{S_1}, \dots, \overline{\mathbb{1}}_{S_k}, y_{k+1}, \dots, y_n\}$ is a set of orthonormal vectors (it would be sufficient to require linear independence of course). Now let Z be the following unitary matrix

$$Z = \begin{pmatrix} \vdots & & \vdots & \vdots & & \vdots \\ \overline{\mathbb{1}}_{S_1} & \cdots & \overline{\mathbb{1}}_{S_k} & y_{k+1} & \cdots & y_n \\ \vdots & & \vdots & \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} \overline{X} & Y \end{pmatrix}$$

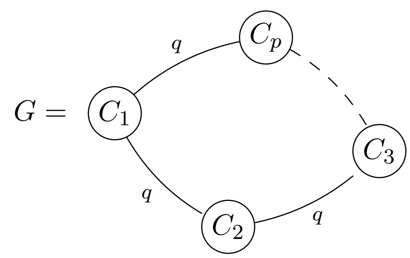
We immediately observe that there exists a nonsingular diagonal matrix D such that $\overline{X} = XD$.

Now consider a set $\{x_1, \dots, x_k\}$ of orthonormal eigenvectors of L_Q corresponding to the eigenvalues $\gamma_1, \dots, \gamma_k$ (one and only one of them is zero), then

$$\begin{aligned} &\begin{pmatrix} x_i^T D^{-1} & 0 \end{pmatrix} (Z^T M Z) \begin{pmatrix} D^{-1} x_i \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} x_i^T D^{-1} & 0 \end{pmatrix} \begin{pmatrix} \overline{X}^T \\ Y^T \end{pmatrix} M \begin{pmatrix} \overline{X} & Y \end{pmatrix} \begin{pmatrix} D^{-1} x_i \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} x_i^T D^{-1} & 0 \end{pmatrix} \begin{pmatrix} \overline{X}^T M \overline{X} & \overline{X}^T M Y \\ Y^T M \overline{X} & Y^T M Y \end{pmatrix} \begin{pmatrix} D^{-1} x_i \\ 0 \end{pmatrix} \\ &= x_i^T D^{-1} (XD)^T M (XD) D^{-1} x_i = x_i^T L_Q x_i = \gamma_i \end{aligned}$$

It follows that there exists a linear vector space $\mathcal{L}$ of dimension $k-1$ such that $x^{\mathsf{T}}(Z^{\mathsf{T}}MZ)x > 0$, for all $x \in \mathcal{L}$. Therefore $\pi(Z^{\mathsf{T}}MZ) \geq k-1$ and since $Z^{\mathsf{T}}MZ$ and M are similar, we conclude that $k \leq \pi(M)+1$. $\square$

It is important to note that the theorem we have just stated actually gives a sharp bound. Indeed there are several graphs for which the maximal partition is made up by exactly $\pi(M)+1$ modules. An example is the *circulant ring of cliques* where G is made by $n = pq$ nodes, divided as in the picture below



each G_i is a complete graph on q nodes and each pair (C_i, C_{i+1}) is joined by exactly q edges each having a weight of $1/q$. The adjacency matrix is

$$A = \begin{pmatrix} E_q & q^{-1}I & \cdots & q^{-1}I \\ q^{-1}I & E_q & \ddots & \\ & \ddots & \ddots & q^{-1}I \\ q^{-1}I & & q^{-1}I & E_q \end{pmatrix}$$

where $E_q = \mathbb{1}\mathbb{1}^{\mathsf{T}} - I$, of order $q \times q$. It can be shown that M has exactly $p-1$ positive eigenvalues, i.e. exactly the number of "communities" minus one.

In the forthcoming discussion we shall consider a simpler version of the community detection problem, known as the *cut-version* of the community detection problem.

2 The cut-modularity of a graph

In what follows, we consider the *cut version* of the community detection problem, that is the problem of finding a subset $S \subseteq V$ having maximal modularity (uniqueness is not ensured in the general case). More precisely, we consider estimating the maximum

$$q_G = \max_{S \subseteq V} q(S, \overline{S}) = \max_{S \subseteq V} Q(S) \frac{n}{|S||\overline{S}|} \tag{2}$$

by means of spectral techniques. The quantity q_G is referred henceforth as the *cut-modularity of the graph* G. The probably best known methods for detecting a subset whose modularity well approximates q_G are based on the idea of spectral partitioning. The idea of such methods is quite simple and reads as follows: we have already seen that, if

$$v_S = \frac{1}{n}(|\overline{S}|\mathbb{1}_S - |S|\mathbb{1}_{\overline{S}})$$

then $v_S^T v_S = \frac{|S||\overline{S}|}{n^2}$ and $v_S^T M v_S = Q(S)/n$, so that

$$\frac{v_S^T M v_S}{v_S^T v_S} = q(S) + q(\overline{S})\,.$$

Observe now that v_S can be wrote as $v_S = n\mathbb{1}_S - |S|\mathbb{1}$, thus it belongs to $\mathcal{L}$, the following subset of $\mathbb{R}^n$: let $\{0,1\}^n$ be the set of n-dimensional vectors whose components are only 0 or 1, then

$$\mathcal{L} = \{n\,x - \|x\|\mathbb{1} \mid x \in \{0,1\}^n\}$$

Clearly

$$q_G = \max_{v\in\mathcal{L}} \frac{v^T M v}{v^T v}\,.$$

Now let $m_1 \geq \cdots \geq m_n$ be the eigenvalues of M and $u_1, \dots, u_n$ the corresponding orthogonal eigenvectors. Then

$$M = \sum_{i=1}^n m_i u_i u_i^T \quad \text{and} \quad q_G = \max_{v\in\mathcal{L}} \sum_{i=1}^n m_i \frac{(u_i^T v)^2}{\|v\|^2}\,.$$

Therefore it is clear that, if v could be chosen to be proportional to u_1, then the former sum would be maximized and it would equal a positive multiple of m_1. However the constrain $v \in \mathcal{L}$ prevents us to such a simple choice and makes the optimization problem much more difficult. In fact it has been pointed out in several works, see e.g., [New06b, New06a], that it is extremely unlikely that a simple procedure exists for finding the optimal $v \in \mathcal{L}$. As a consequence **the spectral partitioning based methods** essentially select v accordingly with the sign of the elements in u_1, by setting

$$v = \text{sing}(u_1) \text{ i.e. } v_i = \begin{cases} 1 & \text{if } (u_1)_i \text{ is positive (or nonnegative)} \\ 0 & \text{otherwise} \end{cases}$$

Then the vertex set V is partitioned into $S_+ = \{i \in V \mid v_i = 1\}$ and $\overline{S_+} = V \setminus S_+$, $G(S_+)$ is proposed as an approximation of the module having maximal modularity in G and, analogously, m_1 is proposed as an approximation of the modularity q_G.

Although the described procedure proposes the subgraph $G(S_+)$ as a leading module, next theorem will show that if $q(S_+) > 0$ then $G(S_+)$ is indeed a community.

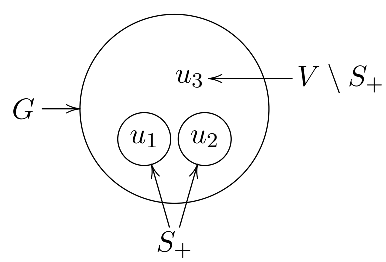
Given a real symmetric matrix X, let me write $\lambda_i(X)$ to denote the i^{th} eigenvalue of X, according to the non-increasing order $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$. We need the following lemmas

Lemma 2.1 Let y be a nonzero real vector, and let X be a real symmetric matrix. Then

$$\lambda_1(X) \geq \lambda_1(X - yy^T) \geq \lambda_2(X)$$

Lemma 2.2 Let $A \geq O$ be irreducible. Let $\lambda_1(A) \geq \mu \geq \lambda_2(A)$ and let u be such that $Au \geq \mu u$. If $S_+ = \{i \mid u_i \geq 0\}$, then $G(S_+)$ is connected.

Proof. Assume by contradiction that $G(S_+)$ has two distinct connected components, and let $u^T = (u_1^T, u_2^T, u_3^T)$. Each u_i corresponds to a distinct component of $G(S_+)$, as in the picture below



Note that by the hypothesis we have done the following entrywise inequality holds

$$u_3 < 0$$

Now, up to a permutation, we have the following block structure

$$\mu \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \leq \begin{pmatrix} A_1 & & B_1 \\ & A_2 & B_2 \\ B_1^T & B_2^T & B_3 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$$

Since A_1 and A_2 are the adjacency matrices of the two distinct connected components of $G(S_+)$, then both of them are irreducible. By the Perron-Frobenius theorem there exist vectors $x_i > 0$ such that $x_i^T A_i = \rho(A_i) x_i^T$.

Therefore $\mu u_i \leq A_i u_i + B_i u_3$ implies

$$\mu(x_i^T u_i) \leq \rho(A_i)(x_i^T u_i) + x_i^T B_i u_3 < \rho(A_i)(x_i^T u_i) \quad i = 1, 2$$

Now, since $x_i > 0$, we have $x_i^T u_i > 0$ so that the previous inequalities imply

$$\mu < \rho(A_i) \qquad i = 1, 2$$

By Cauchy's interlacing theorem A has at least two eigenvalues strictly larger than μ, that is $\lambda_2(A) > \mu$, but this is impossible since $\lambda_2(A) \leq \mu$ by definition. □

Theorem 2.3 Let u be the eigenvector corresponding to the largest nonzero eigenvalue of M. Assume $u^T d \geq 0$ and let $0 \leq \sigma \leq \frac{u^T d}{\operatorname{vol} G}$. If $S_\sigma = \{i \mid u_i \geq \sigma\}$ then $G(S_\sigma)$ is a connected subgraph of G.

Proof. Observe that

$$M(u - \sigma\mathbb{1}) = A(u - \sigma\mathbb{1}) - d\, d^T(u - \sigma\mathbb{1})/\operatorname{vol} G \leq A(u - \sigma\mathbb{1})$$

thus

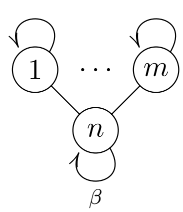
$$A(u-\sigma\mathbb{1}) \geq M(u-\sigma\mathbb{1}) = Mu = \lambda_1(M)(u-\sigma\mathbb{1}) + \lambda_1(M)\sigma\mathbb{1} \geq \lambda_1(M)(u-\sigma\mathbb{1})$$

Using Lemma 2.1 we get $\lambda_1(A) \geq \lambda_1(M) \geq \lambda_2(A)$, therefore by Lemma 2.2 we obtain the thesis. □

Observe that despite the case of the Laplacian nodal domain, we can not apply the above theorem to the set $S_- = \{i \mid u_i \leq 0\}$, indeed the hypothesis $u^T d \geq 0$ would not hold anymore. Of course one may guess that a different argument could be used to overcome such a request. However, unfortunately, if the sign of u is chosen so that $G(S_+)$ is connected, it is not possible to ensure that $G(S_-)$ is connected as well, at least in the general case. Indeed the following counterexample holds

Example 2.4 Let S_m be the star graph with m leaves (thus $m+1$ nodes). Assume in addition that each vertex has a non-weighted self-loop except for the root that has a self-loop with a positive weight $\beta > 0$. The adjacency matrix of S_m and its degree vector are as follows

$S_m =$



$$A = \begin{pmatrix} 1 & & & 1 \\ & \ddots & & \vdots \\ & & 1 & 1 \\ 1 & \cdots & 1 & \beta \end{pmatrix}, \qquad d = \begin{pmatrix} 2 \\ \vdots \\ 2 \\ \beta + m \end{pmatrix}$$

It is not difficult to see that the eigenvalues of the modularity matrix M are the following:

- 0, with associated eigenvector $\mathbb{1}$;
- 1, with multiplicity $m-1$ and associated eigenvectors $\mathbb{1}_{\{1\}} - \mathbb{1}_{\{j\}}$ for $j = 2, \dots, m$;
- $\lambda_\beta = (\beta - m)(m+1)/\mathrm{vol}\, G$, with associated eigenvector $u_\beta = (-1, \dots, -1, m)^T$.

Observe that, when β is large enough we have

$$u_\beta^T d \geq 0 \qquad \text{and} \qquad \lambda_\beta > 1$$

thus $\lambda_1(M) = \lambda_\beta$, and then $V \setminus S_0 = \{1, \dots, m\}$. However $G(S_-)$ is completely non-connected, it consists of m distinct *singletons*, given by the leaf nodes.

References

[New06a] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 69:321–330, 2006.

[New06b] M. E. J. Newman. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA*, 103:8577–8582, 2006.

[NG04] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69(026113), 2004.