Penalized versions of the Newman-Girvan modularity and their relation to normalized cuts and *k*-means clustering

Marianna Bolla
Institute of Mathematics, Budapest University of Technology and Economics, Egry József u. 1, Budapest 1111, Hungary
(Received 19 August 2010; revised manuscript received 24 June 2011; published 25 July 2011)

Two penalized—balanced and normalized—versions of the Newman-Girvan modularity are introduced and estimated by the non-negative eigenvalues of the modularity and normalized modularity matrix, respectively. In this way, the partition of the vertices that maximizes the modularity can be obtained by applying the k-means algorithm for the representatives of the vertices based on the eigenvectors belonging to the largest positive eigenvalues of the modularity or normalized modularity matrix. The proper dimension depends on the number of the structural eigenvalues of positive sign, while dominating negative eigenvalues indicate an anticommunity structure; the balance between the negative and the positive eigenvalues determines whether the underlying graph has a community, anticommunity, or randomlike structure.

DOI: 10.1103/PhysRevE.84.016108 PACS number(s): 89.75.Fb, 02.10.Ox

I. INTRODUCTION

Spectral clustering has evolved in the last decades. Usually, the underlying object is a graph and we want to classify its vertices by maximizing or minimizing some objective function favoring high or low edge densities within or between the clusters. For example, minimum multiway cut problems aim at minimizing the intercommunity edge densities. Though it is not always stated explicitly, an equal balance between the groups is preferable in order to form more realistic clusters and avoid trivial solutions. Thus we are looking for an optimum balanced partition of the vertices such that the objective function penalizes significantly different cluster sizes or volumes, where the volume of a vertex cluster—defined on an edge-weighted graph—is the sum of the weights of edges with at least one endpoint in the cluster. Even if the number of clusters (k) is given, it is nondeterministic polynomial (NP) complete to find the k partition optimizing the objective function. To reduce computational demand, spectral clustering methods were developed for minimizing multiway cuts, ratio cuts, and normalized cuts, cf. [1–7]. In the case of a convenient choice of k a good approximation of the optimal multiway cut can be found in polynomial time in the number of vertices (N). This significantly reduces computational costs, especially if N is very large; for example, genomic data [8] with tens of thousands of genes, and weighted graphs of social or communication networks, where the edge weights are pairwise similarities between a large number of sites. In [1] and [4] it is proved that the more dense the clusters themselves are, the larger the spectral gap between the k smallest normalized Laplacian eigenvalues and the others is. In his survey paper [9], Fortunato gives a nice overview of community detection in graphs.

The Newman-Girvan modularity introduced in [10] directly focuses on modules of higher intracommunity connections than expected based on the model of independent attachment of the vertices with probabilities proportional to their degrees. To maximize this modularity, hierarchical clustering methods based on the edge betweenness measure [10–13] and vector partitioning algorithms based on the spectral decomposition of the modularity matrix [14] are introduced. In [15] an extremal optimization algorithm is presented.

Our purpose is to extend the linear algebraic machinery developed for Laplacian based spectral clustering to the modularity based community detection. To this end, two penalized versions of the Newman-Girvan modularity are introduced in the general framework of an edge-weighted graph [16], and their relation to projections onto the subspace of partition vectors and to k variance of the clusters formed by the vertex representatives is investigated. These considerations give useful information on the choice of k and on the nature of the community structure. With an appropriate k (that is fairly small and corresponds to a spectral gap) a local maximum of the modularity can be guaranteed by processing an advanced version of the traditional k-means algorithm in $\mathcal{O}(kN)$ time. It will be proved by linear algebraic methods that the k-means algorithm used in spectral clustering problems is an efficient tool for community detection provided the "best" eigenvectors are selected for vertex representation. It is also important that we use the spectral decomposition of the matrix that is most adequate to our problem in the sense that the objective function can be sharply estimated by the sum of the largest or smallest eigenvalues of it. For example, the ratio/normalized cut (to be minimized over k partitions of vertices) can be sharply estimated from below by the sum of the k smallest (including the zero) eigenvalues of the positive semidefinite Laplacian/normalized Laplacian matrix; see [1–7]. Therefore the corresponding eigenvectors (apart from the trivial one belonging to the 0 eigenvalue) will be used for the representation of vertices in $\mathbb{R}^{k-1}$ that gives the input of the k-means algorithm with k clusters. Analogously, the balanced/normalized modularity (to be maximized over k partitions of vertices) can be estimated from above by the sum of the k largest eigenvalues of the modularity/normalized-modularity matrix. Similarly, the corresponding eigenvectors of this matrix will be used for the representation of vertices in $\mathbb{R}^k$ that gives the input of the k-means algorithm with $k+1$ clusters (the increase is due the orthogonality to the trivial eigenvector belonging to the 0 eigenvalue that is separated from the k largest ones). As the unpenalized Newman-Girvan modularity is concerned, in [14] no exact estimation for it is given by the leading modularity eigenvalues. To get sharp estimations, balanced or normalized partition vectors

are advisable to use so that they form a suborthogonal matrix and can be relaxed to orthonormal eigenvectors.

We also prove that maximizing the normalized modularity is equivalent to minimizing the normalized cut with the same k, therefore the normalized Laplacian can as well be used. However, the normalized modularity spectrum being in $[-1,1]$, the normalized modularity matrix is more capable to treat large graph problems. E.g., for convergent graph sequences (see [17]), the spectrum is accumulated around zero, and outstanding negative and positive eigenvalues indicate the block structure. This correlationlike matrix is also the kernel of the operator taking conditional expectation. Further, the zero is a watershed in its spectrum, and because of this symmetry, it is well suited for maximizing/minimizing the normalized modularity and may be the most adequate to the spectral characterization of randomlike graphs on several clusters (see [18]), where the probability of two vertices being connected depends merely on their cluster memberships.

Robustness of a community structure is also an important problem discussed in [19]. We also touch upon this problem, rather from a statistical point of view. For large N, if there are no outstanding positive eigenvalues in the modularity spectrum, there is no use in looking for modules of high intracommunity connections. There are other possibilities too: an anticommunity structure with lower than expected intracluster connections or a completely randomlike structure as discussed above. A shift toward the positive eigenvalues indicates community, while that toward the negative ones indicates an anticommunity structure; an equal balance between them may be an indication of a randomlike structure. It is shown through theoretical examples how the signs of the structural (large absolute value) eigenvalues decide the situation. By relaxing the notion of communities we may look for groups of vertices such that their intra- and intercluster connections mainly depend on their cluster memberships, and on this basis, noisy models are investigated [20]. These structures can as well be recovered by means of eigenvectors belonging to the structural eigenvalues of the modularity matrix. For example, equally functioning genes, people, or web sites in genomic, social, or communication networks may form clusters in this wider sense.

The paper is organized as follows. In Sec. II notation and two penalized versions of the Newman-Girvan modularity are introduced in the framework of an edge-weighted graph; further, efficiency of the k-means algorithm is discussed. Exact mathematical formulation of the balanced version is derived in Sec. III, while that of the normalized version is derived in Sec. IV, together with their relation to the k-variance minimization problem; hence the k-means algorithm is applicable. In Sec. V more general community structures are introduced and identified by spectra; further, real-life examples are presented. In Sec. VI some future directions, concerning consistency, are discussed.

II. PRELIMINARIES

A. Edge-weighted graphs

We shall use the general framework of an edge-weighted graph, cf. [1,9,16]. Let $G = (V,\mathbf{W})$ be a graph on N vertices, where the $N \times N$ symmetric matrix $\mathbf{W}$ has non-negative real entries and zero diagonal. Here w_{ij} is the similarity between vertices i and j, where 0 similarity means no connection or edge at all. A simple graph is a special case of it with 0 or 1 weights. In [16] the author first investigates multigraphs that correspond to a $\mathbf{W}$ of non-negative integer entries. Without loss of generality

$$\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} = 1 \tag{1}$$

will be supposed. Hence $\mathbf{W}$ is a joint distribution, with marginal entries

$$d_i = \sum_{j=1}^{N} w_{ij}, \quad i = 1,\ldots,N$$

that are called *generalized vertex degrees*. Let $\mathbf{d} := (d_1,\ldots,d_N)^T$ be the *degree vector* comprising the main diagonal of the diagonal *degree matrix* $\mathbf{D}$ (vectors are columns and ^{T} stands for the transposition). In [1] we estimated a variety of penalized minimum cuts by means of the spectrum of the Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$ or that of the normalized Laplacian $\mathbf{L}_D = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, where $\mathbf{I}$ denotes the identity matrix of appropriate size.

Let $P_k = (V_1,\ldots,V_k)$ be a k partition of the vertices, where the disjoint, nonempty vertex subsets $V_1,\ldots,V_k$ will be referred to as modules, communities, or clusters; $\mathcal{P}_k$ denotes the set of all k partitions. In the edge-weighted case, with condition (1), the Newman-Girvan modularity of P_k given $\mathbf{W}$ is defined by

$$\begin{aligned} Q(P_k,\mathbf{W}) &= \sum_{a=1}^{k}\sum_{i,j\in V_a}(w_{ij} - d_i d_j) \\ &= \sum_{a=1}^{k}[e(V_a,V_a) - \mathrm{Vol}^2(V_a)], \end{aligned}$$

where $e(V_a,V_a) = \sum_{i,j\in V_a} w_{ij}$ is twice the sum of the weights of edges in V_a, and $\mathrm{Vol}(V_a) = \sum_{i\in V_a} d_i$ is the volume of V_a (the sum of the weights of edges with at least one endpoint in V_a); while the entries $d_i d_j$ of the null-model matrix $\mathbf{d}\mathbf{d}^T$ belong to the hypothesis of independence. In other words, under the null hypothesis, vertices i and j are connected to each other independently, with probabilities proportional [actually, because of condition (1), equal] to their generalized degrees. Hence for given k, maximizing $Q(P_k,\mathbf{W})$ is equivalent to looking for k modules of the vertices with intracommunity connections higher than expected under the null hypothesis. As $\sum_{a=1}^{k}\sum_{b=1}^{k}\sum_{i\in V_a}\sum_{j\in V_b}(w_{ij} - d_i d_j) = 0$, the above task is equivalent to minimizing

$$\sum_{a\neq b}\sum_{i\in V_a,\ j\in V_b}(w_{ij} - d_i d_j), \tag{2}$$

that is, to looking for k clusters of the vertices with intercluster connections lower than expected under the hypothesis of independence. In the minimum cut problem the cumulated intercluster connections themselves are minimized. In the edge-weighted case the *modularity matrix* is defined as $\mathbf{B} = \mathbf{W} - \mathbf{d}\mathbf{d}^T$ that is the negative of the so-called Q-Laplacian introduced in [21]. For large N, the authors use the structural

eigenvalues of the normalized matrix $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ that is equal to $\mathbf{I}-\mathbf{L}_D$, therefore its structural eigenvalues are 1 minus those of the normalized Laplacian, and they also coincide with the eigenvalues of the transition probability matrix $\mathbf{D}^{-1}\mathbf{W}$. Then they use specially normalized eigenvectors corresponding to the structural eigenvalues for vertex representation and process the k-means algorithm with the representatives. Therefore the spectral method introduced in [21] for maximizing the Newman-Girvan modularity is closely related to that of [3–5] for minimizing the normalized cut. We will show how these spectra are related to the maxima of the penalized modularities. Though it is not explicitly stated in the existing spectral algorithms, the normalization of their matrices and the eigenvectors implicitly favor balanced partitions defined precisely as follows.

We want to penalize partitions with clusters of extremely different sizes. To measure the size of cluster V_a either the number of its vertices $|V_a|$ or its volume $\mathrm{Vol}(V_a)$ is used. In [9] the author remarks that the Newman-Girvan modularity seems to attain its maximum for clusters of near equal sizes, though there is no explanation for it. In fact, it is true only for completely random networks; see Sec. V. In [11] and [22] the authors also define a good modularity structure as one having near equal sizes of modules. However, they do not make use of this idea in their objective function. As in the $k>2$ case there are more intercluster sums than intracluster ones; it is in Eq. (2) where we penalize clusters of too different size or volume by introducing a factor $\frac{1}{|V_a|}+\frac{1}{|V_b|}$ or $\frac{1}{\mathrm{Vol}(V_a)}+\frac{1}{\mathrm{Vol}(V_b)}$ for the $a\neq b$ pair that shifts the argmin toward balanced pairs.

On the one hand, communities of real-life networks have practical relevance if they do not differ too much in size. On the other hand, in their paper [22] Reichardt and Bornholdt prove that the Newman-Girvan modularity is a special ground-state energy, and in Bolla *et al.* [17] we use the convergence of ground-state energies to prove the testability of some balanced multiway cut densities (roughly speaking, testability means that they can be concluded by sampling from a large graph). Therefore to be testable, nonparametric statistics that can be interpreted as ground-state energies are to be maximized/minimized on conditions of balancing. However, these conditional extrema cannot be immediately related to spectra. As a compromise, we modify the modularity itself so that it would penalize clusters of significantly different sizes. Of course, real-life communities are sometimes very different in size. Our method is capable to find fundamental clusters, and further analysis is needed to separate small communities from the large ones. Another possibility is to distinguish a core of the graph that is free of low-degree vertices for which, usually near zero, eigenvalues are responsible.

For the above reasons, analogously to the the weighted cut of Bolla and Tusnády [1] and the ratio cut of Alpert and Yao [6], we define the *balanced Newman-Girvan modularity* of P_k given $\mathbf{W}$ as

$$Q_B(P_k,\mathbf{W})=\sum_{a=1}^{k}\frac{1}{|V_a|}\sum_{i,j\in V_a}(w_{ij}-d_id_j)$$
$$=\sum_{a=1}^{k}\left[\frac{e(V_a,V_a)}{|V_a|}-\frac{\mathrm{Vol}^2(V_a)}{|V_a|}\right],$$

and analogously to the normalized cut of Meilă and Shi [3] for $k=2$, further that of Bolla and Molnár-Sáska [5], and Azran and Ghahramani [7] for a general k, we define the the *normalized Newman-Girvan modularity* of P_k given $\mathbf{W}$ as

$$Q_N(P_k,\mathbf{W})=\sum_{a=1}^{k}\frac{1}{\mathrm{Vol}\,(V_a)}\sum_{i,j\in V_a}(w_{ij}-d_id_j)$$
$$=\sum_{a=1}^{k}\frac{e(V_a,V_a)}{\mathrm{Vol}\,(V_a)}-1,$$

where we used the fact that $\sum_{a=1}^{k}\mathrm{Vol}\,(V_a)=1$. In [7] it is shown that minimizing the normalized cut of $G=(V,\mathbf{W})$ over k partitions of vertices is equivalent to maximizing $\sum_{a=1}^{k}\frac{e(V_a,V_a)}{\mathrm{Vol}\,(V_a)}$. Hence maximizing the normalized Newman-Girvan modularity can be solved with the same spectral method (using the normalized Laplacian) as the normalized cut problem. However, in Sec. IV we introduce another method based on the normalized modularity matrix. Of course, for increasing N the value of the penalized modularity will decrease akin to the Pearson correlation, relatively small values of which can be significant for large sample sizes. However, for fixed N, the penalized modularity is larger as the modules are nearer to those of an equitable partition.

We also want to show another insight into the problem of the choice of k from the point of view of computational demand and by using the linear algebraic structure of our objective function. In this way, we will prove that for the selected k, maximizing the above adjusted modularities is equivalent to minimizing the k variance of the vertex representatives by choosing an appropriate representation; hence the k-means algorithm is applicable.

B. The k-means algorithm

Let us consider the following clustering problem in a finite dimensional Euclidean space. Given the points $\mathbf{x}_1,\ldots,\mathbf{x}_N\in\mathbb{R}^d$ and an integer $1<k<N$, we are looking for the k partition of the index set $\{1,\ldots,N\}$ (or equivalently, the clustering of the points into k disjoint nonempty subsets), which minimizes the following k *variance* of the points over all possible k partitions $(V_1,\ldots,V_k)$:

$$S_k^2(\mathbf{x}_1,\ldots,\mathbf{x}_N)=\sum_{a=1}^{k}\sum_{j\in V_a}\|\mathbf{x}_j-\mathbf{c}_a\|^2,\quad \mathbf{c}_a=\frac{1}{|V_a|}\sum_{j\in C_a}\mathbf{x}_j. \tag{3}$$

In general, $d\leqslant k$, and they are much less than N. To find the global minimum is NP complete, but the iteration of the k-means algorithm, first described in [23], is capable to find a local minimum in polynomial time. If there exists a well-separated k clustering of the points (even the largest intracluster distance is smaller than the smallest intercluster one), the convergence of the algorithm to the global minimum is proved in [24,25], with a convenient starting. Under relaxed conditions, the speed of the algorithm is increased by a filtration in [26]. The algorithm runs faster if the separation between clusters increases and an overall running time of $\mathcal{O}(kN)$ can be guaranteed. When we apply the k-means algorithm, these separation conditions will be kept in mind.

Sometimes the points $\mathbf{x}_1,\ldots,\mathbf{x}_N$ are endowed with the positive weights $d_1,\ldots,d_N$, where without loss of generality $\sum_{i=1}^N d_i = 1$ can be supposed. In such cases the *weighted k variance* of the points

$$\tilde{S}_k^2(\mathbf{x}_1,\ldots,\mathbf{x}_N) = \sum_{a=1}^{k}\sum_{j\in V_a} d_j \|\mathbf{x}_j - \mathbf{c}_a\|^2, \qquad \mathbf{c}_a = \frac{1}{\sum_{j\in V_a} d_j}\sum_{j\in V_a} d_j \mathbf{x}_j \tag{4}$$

is minimized over all possible k partitions $(V_1,\ldots,V_k)$. The above algorithm can be easily adapted to this situation.

III. MAXIMIZING THE BALANCED NEWMAN-GIRVAN MODULARITY

The k partition P_k is uniquely defined by the $N\times k$ balanced partition matrix $\mathbf{Z}_k = (\mathbf{z}_1,\ldots,\mathbf{z}_k)$, where the ath balanced k-partition vector $\mathbf{z}_a = (z_{1a},\ldots,z_{Na})^T$ is the following: $z_{ia} = \frac{1}{\sqrt{|V_a|}}$ if $i\in V_a$ and 0, otherwise. This yields

$$Q_B(P_k,\mathbf{W}) = Q_B(\mathbf{Z}_k,\mathbf{B}) = \sum_{a=1}^{k}\mathbf{z}_a^T\mathbf{B}\mathbf{z}_a = \operatorname{tr}\mathbf{Z}_k^T\mathbf{B}\mathbf{Z}_k.$$

We want to maximize $\operatorname{tr}\mathbf{Z}_k^T\mathbf{B}\mathbf{Z}_k$ over balanced k-partition matrices $\mathbf{Z}_k\in\mathcal{Z}_k^B$. Observe that the k columns of $\mathbf{Z}_k$ form an orthonormal system in $\mathbb{R}^n$. Therefore $\mathbf{Z}_k$ is a suborthogonal matrix, and hence $\mathbf{Z}_k^T\mathbf{Z}_k = \mathbf{I}_k$ ($\mathbf{I}_k$ being the $k\times k$ identity matrix).

By the notation of Sec. II, let $\beta_1\geqslant\cdots\geqslant\beta_N$ denote the eigenvalues of the modularity matrix $\mathbf{B}$ with corresponding unit-norm, pairwise orthogonal eigenvectors $\mathbf{u}_1,\ldots,\mathbf{u}_N$. Let p denote the number of its positive eigenvalues; thus $\beta_{p+1}=0$ and $\mathbf{u}_{p+1} = \mathbf{1}/\sqrt{N}$. Now let $\mathbf{Y} = (\mathbf{y}_1,\ldots,\mathbf{y}_k)$ be an arbitrary $N\times k$ suborthogonal matrix ($k\leqslant N$). Then by a simple linear algebra (see, e.g., Bathia [27]),

$$\max_{\mathbf{Y}^T\mathbf{Y}=\mathbf{I}_k}\operatorname{tr}(\mathbf{Y}^T\mathbf{B}\mathbf{Y}) = \max_{\mathbf{y}_a^T\mathbf{y}_b=\delta_{ab}}\sum_{a=1}^{k}\mathbf{y}_a^T\mathbf{B}\mathbf{y}_a = \sum_{a=1}^{k}\beta_a$$

and equality is attained when $\mathbf{y}_1,\ldots,\mathbf{y}_k$ are eigenvectors of $\mathbf{B}$ corresponding to $\beta_1,\ldots,\beta_k$. Though the vectors themselves are not necessarily unique (e.g., in the case of multiple eigenvalues), the subspace Span $\{\mathbf{y}_1,\ldots,\mathbf{y}_k\}$ is unique if $\beta_k > \beta_{k+1}$.

Therefore

$$\max_{\mathbf{Z}_k\in\mathcal{Z}_k^B} Q_B(\mathbf{Z}_k,\mathbf{B}) \leqslant \sum_{a=1}^{k}\beta_a \leqslant \sum_{a=1}^{p+1}\beta_a. \tag{5}$$

Both inequalities can be attained by equality only in the $k=1$, $p=0$ case, when our underlying graph is the complete graph (all the weights are equal). This corresponds to perfectly assortative mixing. In this case there is only one cluster with partition vector of equal coordinates (balanced eigenvector belonging to the single 0 eigenvalue). For $k>1$, partition vectors for no graph can coincide with eigenvectors belonging to positive eigenvalues, since their coordinates do not sum to zero, which would be necessary to be orthogonal to the vector belonging to the 0 eigenvalue.

It is also obvious that the maximum with respect to k of the maximum in Eq. (5) is attained with the choice of $k=p+1$. In [14], for the nonpenalized case, the author shows how $p+1$ clusters can be constructed by applying a vector partitioning algorithm for $\mathbf{u}_1,\ldots,\mathbf{u}_p$. However, in the case of large networks, p can also be large, and computation of the positive eigenvalues together with eigenvectors is time consuming. As a compromise, it will be shown that choosing a $k<p$ such that there is a remarkable gap between β_{k-1} and β_k will also suffice. Further, even for a fixed "small" $k<p$, to find the true maxim over k partitions cannot be solved in polynomial time in N, but due to our estimations, spectral partitioning algorithms can be constructed like spectral clustering based on Laplacian eigenvectors; see [1,4]. Now, we are going to discuss this issue in detail.

We expand $Q_B(\mathbf{Z}_k,\mathbf{B})$ with respect to the eigenvalues and eigenvectors of the modularity matrix:

$$Q_B(\mathbf{Z}_k,\mathbf{B}) = \operatorname{tr}\mathbf{Z}_k^T\mathbf{B}\mathbf{Z}_k = \sum_{a=1}^{k}\mathbf{z}_a^T\left(\sum_{i=1}^{N}\beta_i\mathbf{u}_i\mathbf{u}_i^T\right)\mathbf{z}_a = \sum_{i=1}^{N}\beta_i\sum_{a=1}^{k}\left(\mathbf{u}_i^T\mathbf{z}_a\right)^2.$$

We can increase the last sum if we neglect the terms belonging to the negative eigenvalues, hence the outer summation stops at p, or equivalently, at $p+1$. In this case the inner sum is the largest in the $k=p+1$ case, when the partition vectors $\mathbf{z}_1,\ldots,\mathbf{z}_{p+1}$ are "close" to the eigenvectors $\mathbf{u}_1,\ldots,\mathbf{u}_{p+1}$, respectively. As both systems consist of orthonormal sets of vectors, the two subspaces spanned by them should be close to each other. The subspace $\mathcal{F}_{p+1} = \text{Span}\{\mathbf{z}_1,\ldots,\mathbf{z}_{p+1}\}$ consists of stepwise constant vectors on $p+1$ steps, therefore $\mathbf{u}_{p+1}\in\mathcal{F}_{p+1}$, and it suffices to process only the first p eigenvectors. The notation $Q'_{p+1,p}$ will be used for the increased objective function based on the first p eigenvalue-eigenvector pairs and looking for $p+1$ clusters:

$$Q_B(\mathbf{Z}_{p+1},\mathbf{B}) \leqslant Q'_{p+1,p}(\mathbf{Z}_{p+1},\mathbf{B}) := \sum_{i=1}^{p}\beta_i\sum_{a=1}^{p+1}\left(\mathbf{u}_i^T\mathbf{z}_a\right)^2,$$

and in the sequel, for given $\mathbf{B}$, we want to maximize $Q'_{p+1,p}(\mathbf{Z}_{p+1},\mathbf{B})$ over $\mathcal{Z}_{p+1}^B$.

For this purpose, let us project the vectors $\sqrt{\beta_i}\mathbf{u}_i$ onto the subspace $\mathcal{F}_{p+1}$:

$$\sqrt{\beta_i}\mathbf{u}_i = \sum_{a=1}^{p+1}[(\sqrt{\beta_i}\mathbf{u}_i)^T\mathbf{z}_a]\mathbf{z}_a + \operatorname{ort}_{\mathcal{F}_{p+1}}(\sqrt{\beta_i}\mathbf{u}_i), \qquad i=1,\ldots,p. \tag{6}$$

The first term is the component in the subspace, and the second is orthogonal to it. In fact, the projected copies will be in a p-dimensional subspace of $\mathcal{F}_{p+1}$ orthogonal to the $\mathbf{1}$ vector (scalar multiple of $\mathbf{u}_{p+1}$). They will be stepwise constant vectors on $p+1$ steps, and their coordinates sum to 0. This is why one less eigenvector is used than the number of clusters looked for.

By the Pythagorean theorem, for the squared lengths of the vectors in the decomposition (6) we get that

$$\beta_i = \|\sqrt{\beta_i}\mathbf{u}_i\|^2 = \sum_{a=1}^{p+1}[(\sqrt{\beta_i}\mathbf{u}_i)^T\mathbf{z}_a]^2 + \mathrm{dist}^2(\sqrt{\beta_i}\mathbf{u}_i,\mathcal{F}_{p+1}),$$
$$i = 1,\ldots,p.$$

By summing for $i = 1,\ldots,p$, the cumulated second term will turn out to be the sum of inner variances of the vertex representatives in an appropriate representation, defined as follows. For a given positive integer $d \leqslant p$, let the d-dimensional representatives $\mathbf{x}_1,\ldots,\mathbf{x}_N$ of the vertices be row vectors of the $N \times d$ matrix $\mathbf{X}_d = (\sqrt{\beta_1}\mathbf{u}_1,\ldots,\sqrt{\beta_d}\mathbf{u}_d)$. For brevity's sake, the k variance $S_k^2(\mathbf{x}_1,\ldots,\mathbf{x}_N)$ is denoted by $S_k^2(\mathbf{X}_d)$, cf. the notation of Sec. II. Since $\mathcal{F}_k$ consists of stepwise constant vectors on the partition $(V_1,\ldots,V_k)$, by an analysis of variance argument (see [1]) it follows that

$$S_k^2(\mathbf{X}_d) = \sum_{i=1}^{d}\mathrm{dist}^2(\sqrt{\beta_i}\mathbf{u}_i,\mathcal{F}_k).$$

Indeed, $\mathrm{dist}^2(\sqrt{\beta_i}\mathbf{u}_i,\mathcal{F}_k)$ is the minimum squared distance between $\sqrt{\beta_i}\mathbf{u}_i$ and the subspace of stepwise constant vectors on $(V_1,\ldots,V_k)$. In view of Steiner's theorem, the minimum is attained by the stepwise constant vector with coordinates having at most k different values $c_{1i},\ldots,c_{ki}$. Namely, if $j \in V_a$, the jth coordinate of the distance minimizing stepwise constant vector is $c_{ai} = \frac{1}{|V_a|}\sum_{\ell\in V_a}\sqrt{\beta_i}u_i(\ell)$, yielding

$$\mathrm{dist}^2(\sqrt{\beta_i}\mathbf{u}_i,\mathcal{F}_k) = \sum_{a=1}^{k}\sum_{j\in V_a}[\sqrt{\beta_i}u_i(j) - c_{\mathrm{ai}}]^2.$$

By summing for $i = 1,\ldots,d$ and rearranging the summation, $\sum_{i=1}^{d}\mathrm{dist}^2(\sqrt{\beta_i}\mathbf{u}_i,\mathcal{F}_k)$ equals $S_k^2(\mathbf{X}_d)$ with cluster centers $\mathbf{c}_a = (c_{a1},\ldots,c_{ad})$, $a = 1,\ldots,k$, cf. (3). Hence

$$\sum_{i=1}^{p}\beta_i = \sum_{i=1}^{p}\sum_{a=1}^{p+1}[(\sqrt{\beta_i}\mathbf{u}_i)^T\mathbf{z}_a]^2 + \sum_{i=1}^{p}\mathrm{dist}^2(\sqrt{\beta_i}\mathbf{u}_i,\mathcal{F}_{p+1})$$
$$= Q'_{p+1,p}(\mathbf{Z}_{p+1},\mathbf{B}) + S_{p+1}^2(\mathbf{X}_p),$$

where the rows of $\mathbf{X}_p = (\sqrt{\beta_1}\mathbf{u}_1,\ldots,\sqrt{\beta_p}\mathbf{u}_p)$ are regarded as p-dimensional representatives of the vertices. We could as well take $(p+1)$-dimensional representatives as the last coordinates are zeros, and hence $S_{p+1}^2(\mathbf{X}_p) = S_{p+1}^2(\mathbf{X}_{p+1})$. Thus maximizing $Q'_{p+1,p}$ is equivalent to minimizing $S_{p+1}^2(\mathbf{X}_p)$, which can be obtained by applying the k-means algorithm for the p-dimensional representatives with $p+1$ clusters.

More generally, if there is a gap between β_d and $\beta_{d+1} > 0$, then we may look for k clusters based on d-dimensional representatives of the vertices. Analogously to the above calculations, for $d < k \leqslant p+1$ we have that

$$\sum_{i=1}^{d}\beta_i = \sum_{i=1}^{d}\sum_{a=1}^{k}[(\sqrt{\beta_i}\mathbf{u}_i)^T\mathbf{z}_a]^2 + \sum_{i=1}^{d}\mathrm{dist}^2(\sqrt{\beta_i}\mathbf{u}_i,\mathcal{F}_k)$$
$$=: Q'_{k,d}(\mathbf{Z}_k,\mathbf{B}) + S_k^2(\mathbf{X}_d). \quad (7)$$

If β_d is much greater than β_{d+1}, the k variance $S_k^2(\mathbf{X}_{d+1})$ is not significantly greater than $S_k^2(\mathbf{X}_d)$, since $\mathbf{X}_{d+1}$'s last column, $\sqrt{\beta_{d+1}}\mathbf{u}_{d+1}$, will not increase too much the k variance of the d-dimensional representatives, its norm being much less than that of the first d columns. As the left hand side of Eq. (7) is not increased significantly by adding β_{d+1}, the quantity $Q'_{k,d+1}(\mathbf{Z}_k,\mathbf{B})$ is not much greater than $Q'_{k,d}(\mathbf{Z}_k,\mathbf{B})$. Neither the classification nor the value of the modularity is changed much compared to the cost of taking one more eigenvector into consideration. After d has been selected, we can process the k-means algorithm with $k = d+1,\ldots,p+1$ clusters. By an easy argument, $S_{k+1}^2(\mathbf{X}_d) \leqslant S_k^2(\mathbf{X}_d)$, but we can stop if it is much less. These considerations would minimize computational demand and proved good for randomly generated graphs from different block structures; see Sec. V.

Calculating eigenvectors is costly; the Lánczos method performs well if we calculate only eigenvectors belonging to some leading eigenvalues followed by a spectral gap. In [6] the authors suggest to use as many eigenvectors as possible. In fact, using more eigenvectors (up to p) is better from the point of view of accuracy, but using fewer eigenvectors (up to a gap in the positive part of the spectrum) is better from the computational point of view, cf. [4,11]. We have to compromise. By these arguments, a local maximum of the modularity can be expected at $k = d+1$.

The advantage of the modularity matrix versus the Laplacian is that here 0 is a watershed, and for small graphs, the $d = p$, $k = p+1$ choice is feasible; for large graphs we look for gaps (like in the case of the Laplacian) in the positive part of the spectrum, and the number of clusters is one more than the number of the largest positive eigenvalues with corresponding eigenvectors entered into the classification.

IV. MAXIMIZING THE NORMALIZED NEWMAN-GIRVAN MODULARITY

The k partition P_k is also uniquely defined by the $N \times k$ normalized partition matrix $\mathbf{Z}_k = (\mathbf{z}_1,\ldots,\mathbf{z}_k)$, where the ath normalized k-partition vector $\mathbf{z}_a = (z_{1a},\ldots,z_{Na})^T$ is the following: $z_{ia} = \frac{1}{\sqrt{\mathrm{Vol}(V_a)}}$, if $i \in V_a$ and 0, otherwise. By these,

$$Q_N(P_k,\mathbf{W}) = Q_N(\mathbf{Z}_k,\mathbf{B}) = \sum_{a=1}^{k}\mathbf{z}_a^T\mathbf{B}\mathbf{z}_a$$
$$= \mathrm{tr}\,(\mathbf{D}^{1/2}\mathbf{Z}_k)^T\mathbf{B}_D(\mathbf{D}^{1/2}\mathbf{Z}_k),$$

where $\mathbf{B}_D = \mathbf{D}^{-1/2}\mathbf{B}\mathbf{D}^{-1/2}$ will be called *normalized modularity matrix*. Since the matrix $\mathbf{D}^{1/2}\mathbf{Z}_k$ is suborthogonal, the maximization here happens with respect to $\mathbf{Z}_k^T\mathbf{D}\mathbf{Z}_k = \mathbf{I}_k$, that is, over normalized k-partition matrices $\mathbf{Z}_k \in \mathcal{Z}_k^N$.

Let $\beta'_1 \geqslant \cdots \geqslant \beta'_N$ denote the eigenvalues of the symmetric normalized modularity matrix $\mathbf{B}_D$ with corresponding unit-norm, pairwise orthogonal eigenvectors $\mathbf{u}'_1,\ldots,\mathbf{u}'_N$. First we establish the range of the eigenvalues of $\mathbf{B}_D = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} - \mathbf{D}^{-1/2}\mathbf{d}\mathbf{d}^T\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T$, where $\sqrt{\mathbf{d}} := (\sqrt{d_1},\ldots,\sqrt{d_N})^T$. The correlationlike eigenvalues of the first term are in the $[-1,1]$ interval; the largest eigenvalue is always 1 with corresponding unit-norm eigenvector $\sqrt{\mathbf{d}}$. The only nonzero eigenvalue of the rank 1 second term is also 1 with the same eigenvector. Therefore the spectrum of the matrix $\mathbf{B}_D$ is the same as the spectrum of the first term, with the only exception that—due to the subtraction of the second term—the

eigenvalue 1 of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ becomes an eigenvalue 0 of $\mathbf{B}_D$ with eigenvector $\sqrt{\mathbf{d}}$. Hence the spectrum of $\mathbf{B}_D$ is in $[-1,1]$ and includes the 0. These considerations also give an exact relation between the normalized Laplacian and modularity matrix: $\mathbf{B}_D = \mathbf{I} - \mathbf{L}_D - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T$. If the eigenvalues of $\mathbf{L}_D$ are $0 = \lambda_1 \leqslant \cdots \leqslant \lambda_N \leqslant 2$, then the spectrum of $\mathbf{B}_D$ consists of the numbers $1 - \lambda_i$ $(i = 2, \ldots, N)$ and the zero with corresponding eigenvector $\sqrt{\mathbf{d}}$. Further, the multiplicity of 0 is one more than the multiplicity of the eigenvalue 1 of $\mathbf{L}_D$. The multiplicity of 1 is one less than multiplicity of the eigenvalue 0 of $\mathbf{L}_D$; hence 1 cannot be an eigenvalue of $\mathbf{B}_D$ if G is connected ($\mathbf{W}$ is irreducible).

Let p denote the number of positive eigenvalues of $\mathbf{B}_D$ (this p does not necessarily coincide with that of Sec. III). Now let $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_k)$ be an arbitrary $N \times k$ matrix $(k \leqslant N)$ such that $\mathbf{Y}^T\mathbf{D}\mathbf{Y} = \mathbf{I}_k$. With the same linear algebra as used in Sec. III,

$$\max_{\mathbf{Y}^T\mathbf{D}\mathbf{Y}=\mathbf{I}_k} \mathrm{tr}\,(\mathbf{Y}^T\mathbf{B}\mathbf{Y}) = \sum_{a=1}^{k} \beta'_a \tag{8}$$

and equality is attained with $\mathbf{y}_1 = \mathbf{D}^{-1/2}\mathbf{u}'_1, \ldots, \mathbf{y}_k = \mathbf{D}^{-1/2}\mathbf{u}'_k$. Therefore

$$\max_{\mathbf{Z}_k \in \mathcal{Z}_k^N} Q_N(\mathbf{Z}_k, \mathbf{B}) \leqslant \sum_{a=1}^{k} \beta'_a \leqslant \sum_{a=1}^{p+1} \beta'_a.$$

For further investigation, we expand our objective function with respect to the eigenvectors:

$$Q_N(\mathbf{Z}_k, \mathbf{B}) = \sum_{i=1}^{N} \beta'_i \sum_{a=1}^{k} [(\mathbf{u}'_i)^T(\mathbf{D}^{1/2}\mathbf{z}_a)]^2.$$

We can increase this sum if we neglect the terms belonging to the negative eigenvalues, hence the outer summation stops at p, or equivalently, at $p+1$. The inner sum is the largest in the $k = p+1$ case, when the unit-norm, pairwise orthogonal vectors $\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_{p+1}$ are close to the eigenvectors $\mathbf{u}'_1, \ldots, \mathbf{u}'_{p+1}$, respectively. In fact, the two subspaces spanned by them should be close to each other. Now the subspace $\mathcal{F}_{p+1} = \mathrm{Span}\,\{\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_{p+1}\}$ does not consist of stepwise constant vectors, but the following argument is valid. By the notation $Q''_{p+1,p}(\mathbf{Z}_{p+1}, \mathbf{B})$ for the increased objective function based on the first p eigenvalue–eigenvector pairs and looking for $p+1$ clusters we get that

$$\begin{aligned} Q_N(\mathbf{Z}_{p+1}, \mathbf{B}) &\leqslant Q''_{p+1,p}(\mathbf{Z}_{p+1}, \mathbf{B}) \\ &:= \sum_{i=1}^{p} \beta'_i \sum_{a=1}^{p+1} [(\mathbf{u}'_i)^T(\mathbf{D}^{1/2}\mathbf{z}_a)]^2. \end{aligned}$$

In the sequel, for given $\mathbf{B}$, we want to maximize $Q''_{p+1,p}(\mathbf{Z}_{p+1}, \mathbf{B})$ over $\mathcal{Z}_{p+1}^N$.

With the argument of Sec. III, now the vectors $\sqrt{\beta'_i}\,\mathbf{u}'_i$ are projected onto the subspace $\mathcal{F}_{p+1}$:

$$\sqrt{\beta'_i}\,\mathbf{u}'_i = \sum_{a=1}^{p+1} [(\sqrt{\beta'_i}\,\mathbf{u}'_i)^T\mathbf{D}^{1/2}\mathbf{z}_a]\,\mathbf{D}^{1/2}\mathbf{z}_a + \mathrm{ort}_{\mathcal{F}_{p+1}}(\sqrt{\beta'_i}\,\mathbf{u}'_i), \quad i = 1, \ldots, p.$$

As $\sqrt{\beta'_{p+1}}\,\mathbf{u}'_{p+1} = \mathbf{0}$, there is no use in projecting it.

By the Pythagorean theorem, for the squared lengths of the vectors in the above orthogonal decomposition we get that

$$\beta'_i = \|\sqrt{\beta'_i}\,\mathbf{u}'_i\|^2 = \sum_{a=1}^{p+1} [(\sqrt{\beta'_i}\,\mathbf{u}'_i)^T\mathbf{D}^{1/2}\mathbf{z}_a]^2 + \mathrm{dist}^2(\sqrt{\beta'_i}\,\mathbf{u}'_i, \mathcal{F}_{p+1}), \quad i = 1, \ldots, p.$$

Let the vertex representatives $\mathbf{x}'_1, \ldots, \mathbf{x}'_N \in \mathbb{R}^p$ be the row vectors of the $N \times p$ matrix $\mathbf{X}'_p = (\sqrt{\beta'_1}\,\mathbf{D}^{-1/2}\mathbf{u}'_1, \ldots, \sqrt{\beta'_p}\,\mathbf{D}^{-1/2}\mathbf{u}'_p)$. Then

$$\mathrm{dist}^2(\sqrt{\beta'_i}\,\mathbf{u}'_i, \mathcal{F}_{p+1}) = \sum_{j=1}^{N} d_j(x'_{ji} - c_{ji})^2, \quad i = 1, \ldots, p$$

where x'_{ji} is the ith coordinate of the vector $\mathbf{x}'_j$ and c_{ji} is the same for vector $\mathbf{c}_j \in \mathbb{R}^p$, where there are at most $p+1$ different ones among the centers $\mathbf{c}_1, \ldots, \mathbf{c}_N$ assigned to the vertex representatives, namely,

$$c_{ji} = \frac{1}{\sum_{\ell \in V_a} d_\ell} \sum_{\ell \in V_a} d_\ell x'_{\ell i}, \quad j \in V_a, \quad i = 1, \ldots, p.$$

In other words, the column vectors of the $N \times p$ matrix of rows $\mathbf{c}_1, \ldots, \mathbf{c}_N$ are stepwise constant vectors on the same $p+1$ steps belonging to the $(p+1)$ partition of the vertices encoded into the partition matrix $\mathbf{Z}_{p+1}$.

By summing for $i = 1, \ldots, p$, in view of the analysis of variance argument of Sec. III, the cumulated second term will turn out to be the weighted $(p+1)$ variance (4) of the vertex representatives in the $(p+1)$ partition designated by the partition matrix $\mathbf{Z}_{p+1}$:

$$\tilde{S}^2_{p+1}(\mathbf{X}'_p) = \sum_{i=1}^{p} \mathrm{dist}^2(\sqrt{\beta'_i}\,\mathbf{u}'_i, \mathcal{F}_{p+1}) = \sum_{j=1}^{n} d_j \|\mathbf{x}_j - \mathbf{c}_j\|^2.$$

Therefore

$$\sum_{i=1}^{p} \beta'_i = Q''_{p+1,p}(\mathbf{Z}_{p+1}, \mathbf{B}) + \tilde{S}^2_{p+1}(\mathbf{X}'_p).$$

This applies to a given $(p+1)$ partition of the vertices. We are looking for the $(p+1)$ partition maximizing the first term. In view of the above formula, increasing $Q''_{p+1,p}$ can be achieved by decreasing $\tilde{S}^2_{p+1}(\mathbf{X}'_p)$; the latter one is obtained by applying the k-means algorithm with $p+1$ clusters for the p-dimensional representatives $\mathbf{x}'_1, \ldots, \mathbf{x}'_N$ with respective weights $d_1, \ldots, d_N$.

Analogously, for $d < k \leqslant p+1$:

$$\begin{aligned} \sum_{i=1}^{d} \beta'_i &= \sum_{i=1}^{d} \sum_{a=1}^{k} [(\sqrt{\beta'_i}\,\mathbf{u}'_i)^T\mathbf{D}^{1/2}\mathbf{z}_a]^2 + \sum_{i=1}^{d} \mathrm{dist}^2(\sqrt{\beta'_i}\,\mathbf{u}'_i, \mathcal{F}_k) \\ &= Q''_{k,d}(\mathbf{Z}_k, \mathbf{B}) + \tilde{S}^2_k(\mathbf{X}'_d), \end{aligned}$$

where the row vectors of the $N \times d$ matrix $\mathbf{X}'_d = (\sqrt{\beta'_1}\,\mathbf{D}^{-1/2}\mathbf{u}'_1, \ldots, \sqrt{\beta'_d}\,\mathbf{D}^{-1/2}\mathbf{u}'_d)$ are d-dimensional representatives of the vertices. Hence in the presence of a spectral gap between β'_d and $\beta'_{d+1} > 0$—in the miniature world of the [0,1] interval—neither $\sum_{i=1}^{d} \beta'_i$ nor $\tilde{S}^2_k(\mathbf{X}'_d)$ can be increased significantly by introducing one more eigenvalue-eigenvector

pair [by using $(d+1)$-dimensional representatives instead of d-dimensional ones]. Consequently, $Q''_{k,d}(\mathbf{Z}_k,\mathbf{B})$ would not change much, and by the argument of Sec. III, $k = d+1$ clusters based on d-dimensional representatives will suffice.

In their new paper [28], Karrer and Newman introduce a model that takes into consideration the heterogeneity in the degrees of vertices. While the usual block model is biased toward placing vertices of similar degrees in the same cluster, the new model is capable to find clusters of vertices of heterogeneous degrees. I had the same motivation when introduced the normalized modularity matrix. To get this matrix, the edge weights in the modularity matrix are divided by the square roots of the degrees of their end vertices; therefore the normalized modularity matrix supports the effort for eliminating degree differences.

V. COMMUNITIES IN GENERAL

A. Anticommunity structure

Given the weighted graph $G = (V,\mathbf{W})$ instead of taking the maximum, we take the minimum of $Q_B(P_k,\mathbf{W}) = Q_B(\mathbf{Z}_k,\mathbf{B})$ over balanced k-partition matrices $\mathbf{Z}_k$. As for fixed k, analogously to Eq. (5),

$$\min_{\mathbf{Y}^T\mathbf{Y}=\mathbf{I}_k} \operatorname{tr}(\mathbf{Y}^T\mathbf{B}\mathbf{Y}) = \min_{\mathbf{y}_a^T\mathbf{y}_b=\delta_{ab}} \sum_{a=1}^{k} \mathbf{y}_a^T\mathbf{B}\mathbf{y}_a = \sum_{a=1}^{k} \beta_{N+1-a},$$

and similarly to the inference of Sec. III,

$$\min_{P_k\in\mathcal{P}_k} Q_B(\mathbf{Z}_k,\mathbf{B}) = \min_{\mathbf{Z}_k^T\mathbf{Z}_k=\mathbf{I}_k} \operatorname{tr}\mathbf{Z}_k^T\mathbf{B}\mathbf{Z}_k \geqslant \sum_{a=1}^{k} \beta_{N+1-a} \geqslant \sum_{a=1}^{n+1} \beta_{N+1-a},$$

where n is the number of negative eigenvalues of $\mathbf{B}$ ($n+p<N$). For the classification, here we use the scaled (by the square root of the absolute value of the corresponding eigenvalue) eigenvectors belonging to the negative eigenvalues for the representation to find $n+1$ clusters. For large N, it suffices to choose $d<n$ structural negative eigenvalues such that there is a remarkable spectral gap between β_{N+1-d} and β_{N-d}. Then with $\mathbf{X}_d = (\sqrt{|\beta_N|}\cdot\mathbf{u}_N,\ldots,\sqrt{|\beta_{N+1-d}|}\cdot\mathbf{u}_{N+1-d})$, we find the minimum of $S^2_{d+1}(\mathbf{X}_d)$ by the k-means algorithm with $d+1$ clusters.

The same can be done by minimizing the normalized modularity $Q_N(\mathbf{Z}_k,\mathbf{B})$ based on the largest absolute value negative eigenvalues and the corresponding eigenvectors of the normalized modularity matrix.

B. Examples

The following theoretical examples illustrate that large positive eigenvalues of the modularity matrix reflect a community, while large absolute value negative ones reflect an anticommunity structure.

Pure community structure. G is the disjoint union of k complete graphs on $N_1,\ldots,N_k$ vertices, respectively (there are no intercommunity edges, but all possible intracommunity edges are present). This belongs to perfectly assortative mixing. G's modularity matrix has $k-1$ positive eigenvalues, $\beta_k = 0$ with corresponding eigenvector $\mathbf{1}/\sqrt{N}$, and there is only one negative eigenvalue with multiplicity $N-k$. (In the $N_1=\cdots=N_k$ special case $\beta_1=\cdots=\beta_{k-1}$ is a multiple positive eigenvalue.) Here k communities are detected by the k-means algorithm applied for the $(k-1)$-dimensional representatives based on the eigenvectors corresponding to the positive eigenvalues. As these eigenvectors themselves have piecewise constant structures on the steps belonging to the vertex clusters, the k variance of the representatives is 0, and the maximum $Q_B(\mathbf{Z}_k,\mathbf{B})$ is a slightly smaller positive number than the maximum $Q'_{k,k-1}(\mathbf{Z}_k,\mathbf{B})$, the latter one being the sum of the positive eigenvalues. In the $k=1$ case the modularity matrix is negative semidefinite, and both the maximum $Q_B(\mathbf{Z}_k,\mathbf{B})$ and $Q'_{k,k-1}(\mathbf{Z}_k,\mathbf{B})$ are zeros. The normalized modularity matrix $\mathbf{B}_D$ has the eigenvalue 1 with multiplicity $k-1$, one 0 eigenvalue, and all the other eigenvalues are in the $(-1,0)$ interval taking on at most $k-1$ different values. (In the $N_1=\cdots=N_k$ case there is only one negative eigenvalue with multiplicity $N-k$.)

Pure anticommunity structure. G is the complete k-partite graph on $N_1,\ldots,N_k$ vertices, respectively (there are no intracommunity edges, but all possible intercommunity edges are present). These modules may model hub authorities and belong to perfectly disassortative mixing. G's modularity matrix has $k-1$ negative eigenvalues, all the other eigenvalues are zeros. (In the $N_1=\cdots=N_k$ special case there is one negative eigenvalue with multiplicity $k-1$.) Here k communities are detected by the k-means algorithm applied for the $(k-1)$-dimensional representatives based on the eigenvectors corresponding to the negative eigenvalues. As these eigenvectors themselves have piecewise constant structures on the steps belonging to the vertex clusters, the k variance of the representatives is 0, the minimum $Q_B(\mathbf{Z}_k,\mathbf{B})$ is negative, but slightly larger than the minimum $Q'_{k,k-1}(\mathbf{Z}_k,\mathbf{B})$, the latter one being the sum of the negative eigenvalues. The normalized modularity matrix $\mathbf{B}_D$ has $k-1$ negative eigenvalues in the $[-1,0)$ interval; all the other eigenvalues are zeros. (In the $N_1=\cdots=N_k$ case the negative eigenvalue has multiplicity $k-1$.)

Noisy community structure. Now we investigate a case close to the community structure. Let $\mathbf{W}$ be a noisy matrix obtained by burdening a blown up $k\times k$ symmetric pattern matrix with a so-called Wigner noise (its entries in and above the main diagonal are independent, uniformly bounded random variables; otherwise it is symmetric).

The blown up matrix is a symmetric block matrix with k^2 blocks of $N_i\times N_j$ sizes and it has the same entries within the blocks. If $N=\sum_{i=1}^{k} N_i$ is large and the block sizes are of the same magnitude, the spectral decomposition of the edge-weighted graph $G=(V,\mathbf{W})$ is very "close" to that of the following random graph model on the vertex set V: edges come into existence within/between the blocks with probabilities given in the pattern matrix; see [20] for details. If the pattern matrix has "large" diagonal and "small" off-diagonal entries and, further, the blow-up sizes are of the same magnitude, then the modularity matrix has $k-1$ outstanding positive eigenvalues (larger than the absolute value of the smallest negative eigenvalue), and the representatives based on the corresponding eigenvectors can be well classified into k clusters: $S^2_k(\mathbf{X}_{k-1})$ is much smaller than $S^2_{k-1}(\mathbf{X}_{k-1})$, but not much larger than $S^2_{k+1}(\mathbf{X}_{k-1})$.

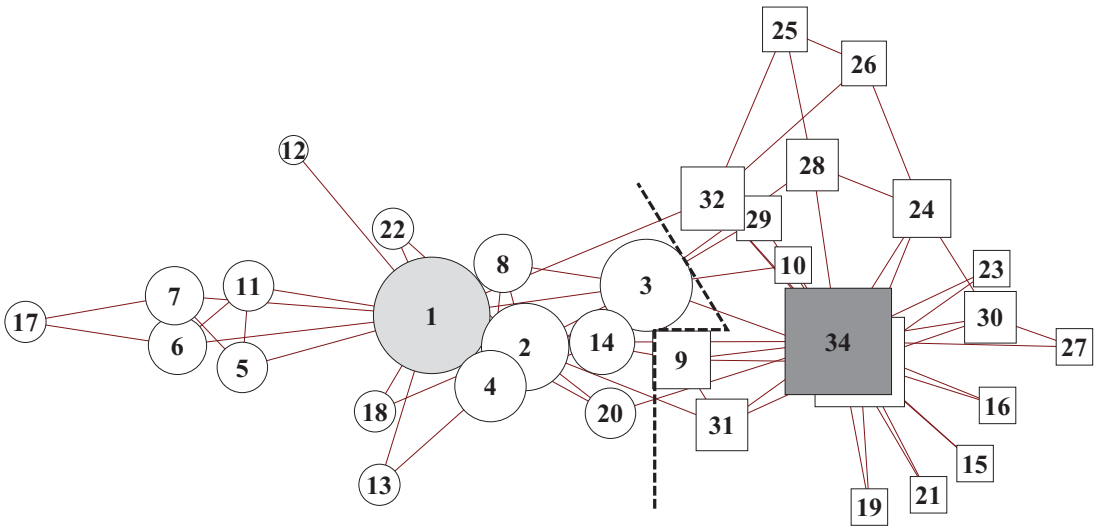

FIG. 1. (Color online) The network of social connections in the karate club network of Zachary [29]. Circles and squares represent nodes of the two clusters with sizes proportional to their degrees. The shaded nodes are the administrator (1) and the instructor (34, covering 33). The separation of the two clusters found by our spectral algorithm maximizing the balanced modularity coincides with the real-life separation of the club members found in the original paper and denoted by the dashed line.

Noisy anticommunity structure. Now we investigate a case close to the anticommunity structure. Let $\mathbf{W}$ be a noisy matrix obtained from a pattern matrix with small diagonal and large off-diagonal entries. The modularity matrix has $k-1$ protruding negative eigenvalues (in absolute value larger than the positive ones), and the representatives based on the corresponding eigenvectors can be well classified into k clusters: $S_k^2(\mathbf{X}_{k-1})$ is much smaller than $S_{k-1}^2(\mathbf{X}_{k-1})$, but not much larger than $S_{k+1}^2(\mathbf{X}_{k-1})$.

We also applied our spectral algorithm for real-life networks, as follows:

Zachary's karate club data. Maximizing the balanced modularity by means of applying the k-means algorithm (with $k=2$ clusters) for the coordinates of the eigenvector belonging to the leading positive eigenvalue of the modularity matrix, our algorithm gave exactly the same partition of the club members as found in the original paper [29]; see Fig. 1.

The bottlenose dolphin community of Doubtful Sound. We investigated the graph of social connections of 40 bottlenose dolphins retained for association analysis by Lusseau *et al.* [30]. They found three groups with individuals most frequently seen together (see Lusseau's Fig. 5), though the groups were not separated clearly by their hierarchical clustering algorithm. Based on one and two leading positive eigenvalues and the corresponding eigenvectors of the modularity matrix, by k-means algorithm, we found two and three clusters, respectively. Though we processed the algorithm for the two- and three-cluster cases separately, one cluster of the three

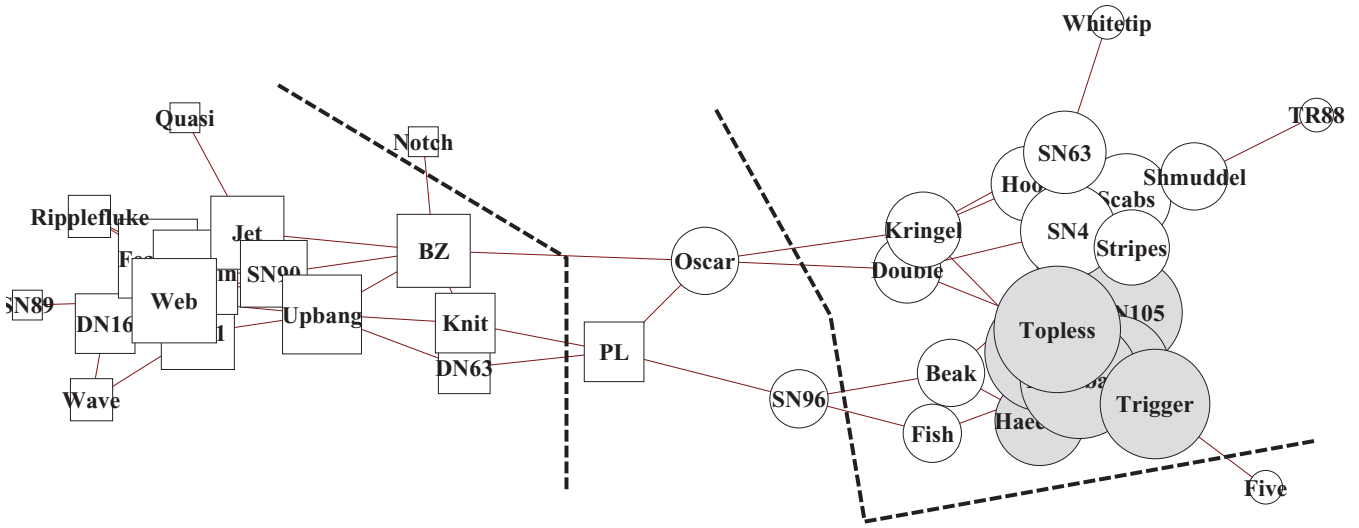

FIG. 2. (Color online) The network of social connections between 40 bottlenose dolphins retained for association analysis by Lusseau *et al.* [30]. Squares and circles represent individuals of the two main clusters obtained by our spectral clustering algorithm maximizing the balanced modularity with $k=2$. In the $k=3$ case, the squares remained in the same cluster, while circles separated into the shaded and open ones. The dense parts of these clusters coincide with the three communities described in [30]. The two main communities observed in the original paper are separated by dashed lines, while the intermediate low degree nodes are not classified uniquely by the original paper.

turned out to be the same as one of the clusters of the two-cluster case. In our Fig. 2, squares and circles represent individuals of the two main clusters obtained by our spectral clustering algorithm maximizing the balanced modularity with $k = 2$, and the shaded and open circles denote the separation of the second cluster when we processed our algorithm with $k = 3$. These three communities are practically the same as discussed on page 401 of [30]. Squares correspond to a male group with an unknown sex individual at the bottom, shaded circles correspond to the group of six males and one female (Trigger) at the left upper corner, finally, open squares represent the female band at the top right of Lusseau's Fig. 5; there were loose connections between these two kinds of circles in accord with the fact that they belonged to one cluster in the two-cluster situation. The two main communities observed in the original paper are separated by dashed lines, while the intermediate low degree nodes—corresponding to the middle part of Lusseau's Fig. 5— are not classified uniquely in the original paper.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

In [31] the total modularity is normalized by a factor so that the perfectly assortative network's modularity attains the maximum value 1, and that of the perfectly disassortative network is in the $[-1,0)$ interval. The author also remarks that the optimum modularity of a perfectly disassortative network is closer to that of a randomly mixing network. Our idea is that a real-life network is generally the superposition of these types; however, for large N, we are able to identify the dominating structure. If N is large, in the noisy cases there are a lot of positive/negative eigenvalues, but for detecting the community/anticommunity structure it suffices to take only the structural ones. We can spare memory and computational time in this way.

Summarizing, a shift toward the positive/negative eigenvalues indicates community/anticommunity structure. The number of structural eigenvalues plus one can be taken for the number of clusters, while the cluster memberships can be concluded by applying the k-means algorithm for the representatives based on the corresponding normalized eigenvectors. Equal balance between the positive and the negative eigenvalues (not only in their number, but also in their magnitudes) indicates a randomlike structure of [18]. Normalized modularity spectrum may play an important role in the spectral characterization of these random-looking graphs. It seems that the bulk of the normalized modularity spectrum is responsible for the pairwise regularities, while the structural (large absolute value) eigenvalues together with eigenvectors indicate the blocks. Though the number of clusters may be large, at the cost of the accuracy it can be decreased by applying spectral methods using eigenvectors belonging to the largest absolute value eigenvalues; cf. [32].

A drawback of the spectral methods is that the structural eigenvalues with corresponding eigenvectors are only capable to reveal fundamental clusters (this is why they are related to balanced modularities), while small communities are hidden behind the near zero eigenvalues. It seems straightforward to enter eigenvectors belonging to small eigenvalues into the representation based classification, but it would cause complications: partly because in the case of large graphs there are too many small eigenvalues (normalized modularity spectrum has tendency accumulate around zero) and partly because small eigenvalues can be indications of small clusters and low degree vertices at the same time. These considerations are valid for large and dense enough graphs. For sparse ones, a so-called core of the graph can be separated which is used to decide whether the graph has a community or anticommunity structure. In [33] the separation is done in terms of the normalized Laplacian eigenvalues. In view of Sec. IV, the normalized modularity spectrum can as well be used.

In [34] Bickel and Chen state the asymptotic consistency of the Newman-Girvan modularity in a submodel of their block model. As these modularities are nonparametric statistics, and the conditions apply to the unknown model parameters, it is possible to substitute their estimates for the parameters, and if these satisfy the conditions, we may expect consistency. It is a future direction to check the consistency conditions for the penalized modularities. We conjecture that for large N, the balanced modularity is a consistent estimator of the true modularity structure of the underlying weighted graph if the optimum k and optimum k partition $(V_1, \ldots, V_k)$ of its vertices satisfy the following requirement: for every $a = 1, \ldots, k$,

$$\sqrt{|V_a|}e(V_a, V_a) > \sum_{b \neq a} \sqrt{|V_b|}e(V_a, V_b)$$

holds. For the normalized modularity this condition seems to be

$$\sqrt{\mathrm{Vol}\,(V_a)}e(V_a, V_a) > \sum_{b \neq a} \sqrt{\mathrm{Vol}\,(V_b)}e(V_a, V_b).$$

The conditions formulated in the above conjectures are more likely to be satisfied by balanced clusters. If these requirements are violated, one should treat carefully the result of the classification and suspect other possibilities: an anticommunity or randomlike structure.

ACKNOWLEDGMENTS

The author thanks János Tóth and Erik Bodzsár for processing computer programs using the spectral clustering algorithms proposed in the present paper. Research was supported by the Hungarian National Research Fund and by the National Office of Research and Technology (Grants No. OTKA 76481 and No. OTKA-NKTH 77778).

[1] M. Bolla and G. Tusnády, Discrete Math. **128**, 1 (1994).

[2] J. Shi and J. Malik, IEEE Trans. Pattern Anal. Machine Intell. **22**, 888 (2000).

[3] M. Meilă and J. Shi, in *Proceedings of the NIPS (Neural Information Processing Systems) 13 Conference*, edited by T. K. Leen, T. G. Dietterich, and V. Tresp (MIT Press, Cambridge, MA, 2001), pp. 873–879.

[4] A. Y. Ng, M. I. Jordan, and Y. Weiss, in *Proceedings of the NIPS (Neural Information Processing Systems) 14 Conference*, edited by T. G. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, Cambridge, MA, 2001), pp. 849–856.
[5] M. Bolla and G. Molnár-Sáska, Discrete Math. **282**, 23 (2004).
[6] C. J. Alpert and S.-Z. Yao, in *Proceedings of the 32nd ACM/IEEE International Conference on Design Automation*, edited by B. T. Preas, P. G. Karger, B. S. Nobandegani, and M. Pedram (Association for Computer Machinery, New York, NY, 1995), pp. 195–200.
[7] A. Azran and Z. Ghahramani, in *Proceedings of the CVPR Conference* (IEEE Computer Society, New York, 2006), pp. 190–197.
[8] D. M. Wilkinson and B. A. Huberman, Proc. Natl. Acad. Sci. USA **101**, 5241 (2004).
[9] S. Fortunato, Phys. Rep. **486**, 75 (2010).
[10] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[11] M. E. J. Newman, Eur. Phys. J. B **38**, 321 (2004).
[12] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).
[13] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E **70**, 066111 (2004).
[14] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).
[15] J. Duch and A. Arenas, Phys. Rev. E **72**, 027104 (2005).
[16] M. E. J. Newman, Phys. Rev. E **70**, 056131 (2004).
[17] M. Bolla, T. Kói, and A. Krámli, e-print arXiv:1001.1623v1 [math.PR] (2010).
[18] L. Lovász and V. T. Sós, J. Comb. Theory, Ser. B **98**, 146 (2008).
[19] B. Karrer, E. Levina, and M. E. J. Newman, Phys. Rev. E **77**, 046119 (2008).
[20] M. Bolla, Lin. Alg. Appl. **402**, 228 (2005).
[21] S. White and P. Smyth, in *Proceedings of SIAM International Conference on Data Mining* (SIAM, Newport Beach, CA, 2005), pp. 76–84.
[22] J. Reichardt and S. Bornholdt, Phys. Rev. E **76**, 015102(R) (2007).
[23] J. B. MacQueen, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. Cam and J. Neyman (University of California Press, Berkeley, CA, 1967), Vol. 1, pp. 281–297.
[24] J. C. Dunn, J. Cybernetics **3**, 32 (1973).
[25] J. C. Dunn, J. Cybernetics **4**, 95 (1974).
[26] T. Kanungo *et al.*, IEEE Trans. Pattern Anal. Machine Intell. **24**, 881 (2002).
[27] R. Bathia, *Matrix Analysis* (Springer, New York, 1997).
[28] B. Karrer and M. E. J. Newman, Phys. Rev. E **83**, 016107 (2011).
[29] W. W. Zachary, J. Antropol. Res. **33**, 452 (1977).
[30] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behav. Ecol. Sociobiol. **54**, 396 (2003).
[31] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[32] A. Coja-Oghlan, Combinatorics, Probab. Comput. **19**, 227 (2010).
[33] N. Alon *et al.*, Lect. Notes Comput. Sci. **4596**, 789 (2007).
[34] P. J. Bickel and A. Chen, Proc. Natl. Acad. Sci. USA **106**, 21068 (2009).