Beyond the Expanders

Marianna Bolla

Institute of Mathematics, Budapest University of Technology and Economics,
Egry Jozsef u. 1, 1111 Budapest, Hungary; E-mail: `marib@math.bme.hu`

Abstract

Expander graphs are widely used in communication problems and construction of error correcting codes. In such graphs, information gets through very quickly. Typically, it is not true for social or biological networks, though we may find a partition of the vertices such that the induced subgraphs on them and the bipartite subgraphs between any pair of them exhibit regular behavior of information flow within or between the vertex subsets. Implications between spectral and regularity properties are discussed.

Keywords: Spectral gap; Spectral clustering; Volume regularity.

1 Introduction

We want to go beyond the expander graphs that – for four decades – have played an important role in communication networks; for a summary, see e.g., Chung [8] and Hoory et al. [14]. Roughly speaking, the expansion property means that each subset of the graph's vertices has "many" neighbors (combinatorial view), and hence, information gets through such a graph very "quickly" (probabilistic view). We will not give exact definitions of expanders here as those contain many parameters which are not used later. We rather refer to the spectral and random walk characterization of such graphs, as discussed, among others by Alon [1], and Meila and Shi [17].

The general framework of an edge-weighted graph will be used. Expanders have a spectral gap bounded away from zero, where – for a connected graph – this gap is defined as the minimum distance between the normalized Laplacian spectrum (apart from the trivial zero eigenvalue) and the endpoints of the [0,2] interval, the possible range of the spectrum. The larger the spectral gap, the more our graph resembles a random graph and exhibits quasi-random properties, e.g., the edge densities within any subset and between any two subsets of its vertices do not differ too much of what is expected, see the Expander Mixing Lemma 1 of Section 2. Quasi-random properties and spectral gap of random graphs with given expected degrees are discussed in Chung and Graham [9], and Coja-Oghlan and Lanka [11].

However, the spectral gap appears not at the ends of the normalized Laplacian spectrum in case of generalized random or generalized quasi-random graphs that, in the presence of $k \geq 2$ underlying clusters, have k eigenvalues (including the zero) separated from 1, while the bulk of the spectrum is located around 1, see e.g., [6]. These structures are usual in social or biological networks having k clusters of vertices (that belong to social groups or similarly functioning enzymes) such that the edge density within the clusters and between any pair of the clusters is homogeneous.

Our conjecture is that k so-called structural eigenvalues (separated from 1) in the normalized Laplacian spectrum are indications of such a structure, while the near 1 eigenvalues are responsible for the pairwise regularities. The clusters themselves can be recovered by applying the k-means algorithm for the vertex representatives obtained by the eigenvectors corresponding to the structural eigenvalues (apart from the zero). For the $k = 2$ case we will give an

exact relation between the eigenvalue separation (of the non-trivial structural eigenvalue from the bulk of the spectrum) and the volume regularity of the cluster pair that is obtained by the k-means algorithm applied for the coordinates of the transformed eigenvector belonging to the non-trivial structural eigenvalue, see Theorem 1 of Section 3. To eliminate the trivial eigenvalue-eigenvector pair, we shall rather use the normalized modularity spectrum of [7] that plays an important role in finding the extrema of some penalized versions of the Newman-Girvan modularity introduced in [18]. Theorem 2 of Section 4 gives an estimation for the extent of volume-regularity of the different cluster pairs in the $k > 2$ case based on the spectral gap and the k-variance of the vertex representatives.

In [10, 16], the authors give algorithms – based on low rank approximation – to find a regular partition if k is known and our graph comes from a generalized random graph model with k clusters. Without knowing k, there are constructions – like [13] – based on refinement of partitions and leading to a very fine partition with number of clusters depending merely on the constant ruling the regularity of the cluster pairs. On the contrary, our purpose is to estimate the extent of the regularity of the cluster pairs by means of spectral gaps and eigenvectors. The estimations given are relevant only in the presence of a large spectral gap (between some structural and the other eigenvalues) and special classification properties of the eigenvectors corresponding to the structural eigenvalues, see Theorem 2 of Section 4. In this case, the algorithm is straightforward via k-means clustering.

2   Preliminaries and statement of purpose

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/edge at all. A simple graph is a special case of it with 0-1 weights. 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

$$d_i = \sum_{j=1}^{n} w_{ij}, \quad i = 1, \dots, n$$

which are the *generalized vertex degrees* collected in the main diagonal of the diagonal *degree matrix* $\mathbf{D} = \operatorname{diag}(\mathbf{d})$, $\mathbf{d} = (d_1, \dots, d_n)^T$. In [4, 5] we investigated the spectral gap 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.

Suppose that our graph is connected ($\mathbf{W}$ is irreducible). Let $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ denote the eigenvalues of the symmetric normalized Laplacian $\mathbf{L}_D$ with corresponding unit-norm, pairwise orthogonal eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$. Namely, $\mathbf{u}_1 = (\sqrt{d_1}, \dots, \sqrt{d_n})^T = \sqrt{\mathbf{d}}$. In the random walk setup $\mathbf{D}^{-1} \mathbf{W}$ is the transition matrix (its entry in the (i, j)-th position is the conditional probability of moving from vertex i to vertex j in one step, given that we are in i) which is a stochastic matrix with eigenvalues $1 - \lambda_i$ and corresponding eigenvectors $\mathbf{D}^{-1/2} \mathbf{u}_i$ $(i = 1, \dots, n)$. "Good" expanders have a λ_2 bounded away from zero, that also implies the

separation of the isoperimetric number

$$h(G) = \min_{U \subset V : \, \mathrm{Vol}\,(U) \leq \frac{1}{2}} \frac{w(U, \overline{U})}{\mathrm{Vol}\,(U)}, \tag{2}$$

where for $X, Y \subset V$: $w(X, Y) = \sum_{i \in X} \sum_{j \in Y} w_{ij}$ is the weighted cut between X and Y, while $\mathrm{Vol}\,(U) = \sum_{i \in U} d_i$ is the volume of $U \subset V$. In view of (1), $\mathrm{Vol}\,(V) = 1$, this is why the minimum is taken on vertex sets having volume at most $\frac{1}{2}$. In [5], we proved that

$$\frac{1}{2}\lambda_2 \leq h(G) \leq \min\{1, \sqrt{2\lambda_2}\}, \tag{3}$$

while in the $\lambda_2 \leq 1$ case the stronger upper estimation

$$h(G) \leq \sqrt{\lambda_2 (2 - \lambda_2)}$$

holds. (We remark that $\lambda_2 \leq \frac{n}{n-1}$ always holds.)

If a network does not have a "large" λ_2 (compared to the natural lower bound), or equivalently – in view of the above inequalities – it has a relatively "small" isoperimetric number, then the 2-partition of the vertices giving the minimum in (2) indicates a bottleneck, or equivalently, a low conductivity edge-set between two disjoint vertex clusters such that the random walk gets through with small probability between them, but – as some equivalent notions will indicate – it is rapidly mixing within the clusters. To find the clusters, the coordinates of the transformed eigenvector $\mathbf{D}^{-1/2}\mathbf{u}_2$ will be used. In [4], we proved that for the weighted 2-variance of this vector's coordinates

$$S_2^2(\mathbf{D}^{-1/2}\mathbf{u}_2) \leq \frac{\lambda_2}{\lambda_3} \tag{4}$$

holds. For a general $2 \leq k \leq n$, the notion of k-variance – in the Analysis of Variance sense – is the following. The weighted k-variance of the k-dimensional vertex representatives $\mathbf{x}_1, \ldots, \mathbf{x}_n$ comprising the row vectors of the $n \times k$ matrix $\mathbf{X}$ is defined by

$$S_k^2(\mathbf{X}) = \min_{P_k \in \mathcal{P}_k} S_k^2(P_k, \mathbf{X}) = \min_{P_k = (V_1, \ldots, V_k)} \sum_{a=1}^{k} \sum_{j \in V_a} d_j \|\mathbf{x}_j - \mathbf{c}_a\|^2, \tag{5}$$

where $\mathbf{c}_a = \frac{1}{\mathrm{Vol}\,(V_a)} \sum_{j \in V_a} d_j \mathbf{x}_j$ is the weighted center of cluster V_a $(a = 1, \ldots, k)$ and $\mathcal{P}_k$ denotes the set of k-partitions of the vertices. We remark that $S_2^2(\mathbf{D}^{-1/2}\mathbf{u}_1, \mathbf{D}^{-1/2}\mathbf{u}_2) = S_2^2(\mathbf{D}^{-1/2}\mathbf{u}_2)$, since $\mathbf{D}^{-1/2}\mathbf{u}_1 = \mathbf{1}$ is the all 1's vector.

The above results were generalized for minimizing the normalized k-way cut

$$f_k(P_k, G) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left(\frac{1}{\mathrm{Vol}\,(V_a)} + \frac{1}{\mathrm{Vol}\,(V_b)} \right) w(V_a, V_b) = k - \sum_{a=1}^{k} \frac{w(V_a, V_a)}{\mathrm{Vol}\,(V_a)} \tag{6}$$

of the k-partition $P_k = (V_1, \ldots, V_k)$ over the set of all possible k-partitions. Let

$$f_k(G) = \min_{P_k \in \mathcal{P}_k} f_k(P_k, G)$$

be the *minimum normalized k-way cut* of the underlying weighted graph $G = (V, \mathbf{W})$. In fact, $f_2(G)$ is the symmetric version of the isoperimetric number and $f_2(G) \leq 2h(G)$. In [5] we proved that

$$\sum_{i=1}^{k} \lambda_i \leq f_k(G) \leq c^2 \sum_{i=1}^{k} \lambda_i, \tag{7}$$

where the upper estimation is relevant only in the case when $S_k^2(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is small enough and the constant c depends on this minimum k-variance of the vertex representatives.

The *normalized Newman-Girvan modularity* is defined in [7] as the penalized version of the Newman-Girvan modularity [18] in the following way. The normalized k-way modularity of $P_k = (V_1, \ldots, V_k)$ is

$$Q_k(P_k, G) = \sum_{a=1}^{k} \frac{1}{\text{Vol}(V_a)} \sum_{i,j \in V_a} (w_{ij} - d_i d_j) = \sum_{a=1}^{k} \frac{1}{\text{Vol}(V_a)} [w(V_a, V_a) - \text{Vol}^2(V_a)]$$
$$= \sum_{a=1}^{k} \frac{w(V_a, V_a)}{\text{Vol}(V_a)} - 1 = k - 1 - f_k(P_k), \tag{8}$$

and

$$Q_k(G) = \max_{P_k \in \mathcal{P}_k} Q_k(P_k, G)$$

is the *maximum normalized k-way Newman-Girvan modularity* of the underlying weighted graph $G = (V, \mathbf{W})$. For given k, maximizing this modularity is equivalent to minimizing the normalized cut and can be solved by the same spectral technique. In fact, it is more convenient to use the spectral decomposition of the normalized modularity matrix $\mathbf{B}_D = \mathbf{I} - \mathbf{L}_D - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T$ with eigenvalues $\beta_1 \geq \cdots \geq \beta_n$, that are the numbers $1 - \lambda_i$ with eigenvectors $\mathbf{u}_i$ $(i = 2, \ldots, n)$ and the zero with corresponding unit-norm eigenvector $\sqrt{\mathbf{d}}$. In [5, 7], we also show that a spectral gap between λ_k and λ_{k+1} is an indication of k clusters with low inter-cluster connections; further, the intra-cluster connections (w_{ij}) between vertices i and j of the same cluster are higher than expected under the hypothesis of independence (in view of which the vertices are connected with probability $d_i d_j$). In the random walk framework, the random walk stays within the clusters with high probability.

Conversely, minimizing the above modularity will result in clusters with high inter- and low intra-cluster connections. In [7], we proved that

$$\min_{P_k \in \mathcal{P}_k} Q_k(P_k, G) \geq \sum_{i=1}^{k} \beta_{n+1-i}. \tag{9}$$

The existence of k "large" (significantly larger than 1) eigenvalues in the normalized Laplacian spectrum, or equivalently, the existence of k negative eigenvalues (separated from 0) in the normalized modularity spectrum is an indication of k clusters with the above property. In the random walk setup: the walk stays within the clusters with low probability.

These two types of network structures are frequently called community or anti-community structure. These are the two extreme cases, when $f_k(P_k, G)$ is either minimized or maximized, and the optimization gives k clusters with either strong intra-cluster and weak inter-cluster connections, or vice versa. Some networks exhibit a more general, still regular behavior: the vertices can be classified into k clusters such that the information-flow within them and between any pair of them is homogeneous. In terms of random walks, the walk stays within clusters or switches between clusters with probabilities characteristic for the cluster pair. That is, if the random walk moves from a vertex of cluster V_a to a vertex of cluster V_b, then the probability of doing this does not depend on the actual vertices, it merely depends on their cluster memberships, $a, b = 1, \ldots, k$.

In this context, we examined the following generalized random graph model, that corresponds to the ideal case: given the number of clusters k, the vertices of the graph independently belong

to the clusters; further, conditioned on the cluster memberships, vertices $i \in V_a$ and $j \in V_b$ are connected with probability p_{ab}, independently of each other, $1 \leq a, b \leq k$. Applying the results [6] for the spectral characterization of some noisy random graphs, we are able to prove that the normalized modularity spectrum of a generalized random graph is the following: there exists a positive number $\theta < 1$, independent of n, such that there are exactly $k - 1$ so-called structural eigenvalues of $\mathbf{B}_D$ that are greater than $\theta - o(1)$, while all the others are o(1) in absolute value. It is equivalent that $\mathbf{L}_D$ has k eigenvalues (including the zero) separated from 1.

The $k = 1$ case corresponds to quasi-random graphs and the above characterization corresponds to the eigenvalue separation of such graphs, discussed in [9]. The authors also prove some implications between the so-called quasi-random properties. For example, for dense graphs, "good" eigenvalue separation is equivalent to "low" discrepancy (of the induced subgraphs' densities from the overall edge density).

For the $k \geq 2$ case, generalized quasi-random graphs were introduced by Lovász and T. Sós [15]. These graphs are deterministic counterparts of generalized random graphs with the same spectral properties. In fact, the authors define so-called generalized quasi-random graph sequences by means of graph convergence that also implies the convergence of spectra. Though, the spectrum itself does not carry enough information for the cluster structure of the graph, together with some classification properties of the structural eigenvectors it does. We want to prove some implication between the spectral gap and the volume-regularity of the cluster pairs, also using the structural eigenvectors.

The notion of volume regularity was introduced by Alon et al. [2]. We shall use a slightly modified version of this notion.

Definition 1 *Let $G = (V, \mathbf{W})$ be weighted graph with $Vol(V) = 1$. The disjoint pair (A, B) is α-volume regular if for all $X \subset A$, $Y \subset B$ we have*

$$|w(X, Y) - \rho(A, B) \, Vol(X) \, Vol(Y)| \leq \alpha \sqrt{Vol(A) \, Vol(B)}, \tag{10}$$

where $\rho(A, B) = \frac{e(A,B)}{Vol(A) \, Vol(B)}$ is the relative inter-cluster density of (A, B).

Our definition was inspired by the Expander Mixing Lemma stated e.g., in [14] for regular graphs and in [8] for simple graphs in the context of quasi-random properties. Now we formulate it for edge-weighted graphs on a general degree sequence. We also include the proof as a preparation for the proof of Theorem 1 of Section 3.

Lemma 1 *(**Expander Mixing Lemma for Weighted Graphs**) Let $G = (V, W)$ be a weighted graph and suppose that $Vol(V) = 1$. Then for all $X, Y \subset V$:*

$$|w(X, Y) - Vol(X) \, Vol(Y)| \leq \|\mathbf{B}_D\| \cdot \sqrt{Vol(X)(1 - Vol(X)) \, Vol(Y)(1 - Vol(Y))}$$
$$\leq \|\mathbf{B}_D\| \cdot \sqrt{Vol(X) \, Vol(Y)},$$

where $\|\mathbf{B}_D\|$ is the spectral norm of the normalized modularity matrix of G.

Proof   Let $X \subset A$, $Y \subset B$ and $\mathbf{1}_U \in \mathbb{R}^n$ denote the indicator vector of $U \subset V$. Further, $\mathbf{x} := \mathbf{D}^{1/2} \mathbf{1}_X$ and $\mathbf{y} := \mathbf{D}^{1/2} \mathbf{1}_Y$.

We use the spectral decomposition $\mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} = \sum_{i=1}^{n} \rho_i \mathbf{u}_i \mathbf{u}_i^T$, where $\rho_i = 1 - \lambda_i$ ($i = 2, \ldots, n$) are eigenvalues of $\mathbf{B}_D$ and $\rho_1 = 1$ with corresponding unit-norm eigenvector $\mathbf{u}_1 =$

$\sqrt{\mathbf{d}} = \mathbf{D}^{1/2}\mathbf{1}$. We remark that $\mathbf{u}_1$ is also an eigenvector of $\mathbf{B}_D$ corresponding to the eigenvalue zero, hence $\|\mathbf{B}_D\| = \max_{i\geq 2}|\rho_i|$. Let $\mathbf{x} = \sum_{i=1}^n x_i\mathbf{u}_i$ and $\mathbf{y} = \sum_{i=1}^n y_i\mathbf{u}_i$ be the expansions of $\mathbf{x}$ and $\mathbf{y}$ in the orthonormal basis $\mathbf{u}_1,\ldots,\mathbf{u}_n$ with coordinates $x_i = \mathbf{x}^T\mathbf{u}_i$ and $y_i = \mathbf{y}^T\mathbf{u}_i$, respectively. Observe that $x_1 = \text{Vol}(X)$, $y_1 = \text{Vol}(Y)$ and $\sum_{i=1}^n x_i^2 = \|\mathbf{x}\|^2 = \text{Vol}(X)$, $\sum_{i=1}^n y_i^2 = \|\mathbf{y}\|^2 = \text{Vol}(Y)$. Based on these,

$$|w(X,Y) - \text{Vol}(X)\text{Vol}(Y)| = |\sum_{i=2}^n \rho_i x_i y_i| \leq \|\mathbf{B}_D\| \cdot |\sum_{i=2}^n x_i y_i|$$

$$\leq \|\mathbf{B}_D\| \cdot \sqrt{\sum_{i=2}^n x_i^2 \sum_{i=2}^n y_i^2}$$

$$\leq \|\mathbf{B}_D\| \cdot \sqrt{\text{Vol}(X)(1-\text{Vol}(X))\text{Vol}(Y)(1-\text{Vol}(Y))}$$

$$\leq \|\mathbf{B}_D\| \cdot \sqrt{\text{Vol}(X)\text{Vol}(Y)},$$

where we also used the triangle and the Cauchy-Schwarz inequalities. ■

We remark that the spectral gap of G is $1 - \|\mathbf{B}_D\|$, hence – in view of Lemma 1 – the density between any two subsets of "good" expanders is near to what is expected. On the contrary, in the above definition of volume regularity, the X, Y pairs are disjoint, and a "small" α indicates that the (A, B) pair is like a bipartite expander, see e.g., [8].

In the next section we shall prove the following statement for the $k = 2$ case: if one eigenvalue jumps out of the bulk of the normalized modularity spectrum, then clustering the coordinates of the corresponding transformed eigenvector into 2 parts (by minimizing the 2-variance of its coordinates) will result in an α-volume regular partition of the vertices, where α depends on the spectral gap.

We may go further: if $k - 1$ (so-called structural) eigenvalues jump out of the normalized modularity spectrum, then clustering the representatives of the vertices – obtained by the corresponding eigenvectors in the usual way – into k clusters will result in α-volume regular pairs, where α depends on the spectral gap (between the structural eigenvalues and the bulk of the spectrum) and the k-variance of the vertex representatives based on the eigenvectors corresponding to the structural eigenvalues. In Section 4, we give an estimation for α in the $k \geq 2$ case; further, we extend the estimation to the clusters themselves.

3   Eigenvalue separation and volume regularity (k=2 case)

Theorem 1 *Let $G = (V, \mathbf{W})$ is an edge-weighted graph on n vertices, with generalized degrees $d_1, \ldots, d_n$ and $\mathbf{D} = diag(d_1, \ldots, d_n)$. Suppose that $Vol(V) = 1$. Let the eigenvalues of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, enumerated in decreasing absolute values, be*

$$1 = \rho_1 > |\rho_2| = \theta > \varepsilon \geq |\rho_i|, \quad i \geq 3.$$

The partition (A, B) of V is defined so that it minimizes the weighted 2-variance of the coordinates of $\mathbf{D}^{-1/2}\mathbf{u}_2$, where $\mathbf{u}_2$ is the unit-norm eigenvector belonging to ρ_2. Then the (A, B) pair is $\mathcal{O}(\sqrt{\frac{1-\theta}{1-\varepsilon}})$-volume regular.

Proof   We use the notations of Lemma 1's proof. Let $X \subset A$, $Y \subset B$. Fort short, $\mathbf{x} := \mathbf{D}^{1/2}\mathbf{1}_X$, $\mathbf{y} := \mathbf{D}^{1/2}\mathbf{1}_Y$, $\mathbf{a} := \mathbf{D}^{1/2}\mathbf{1}_A$, $\mathbf{b} := \mathbf{D}^{1/2}\mathbf{1}_B$. With $\rho := \rho(A, B)$ and $\mathbf{M} := \mathbf{W} - \rho\mathbf{d}\mathbf{d}^T$,

$$|w(X, Y) - \rho\text{Vol}(X)\text{Vol}(Y)| = |\mathbf{1}_X^T\mathbf{M}\mathbf{1}_Y| = |\mathbf{x}^T(\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} - \rho\sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T)\mathbf{y}|. \tag{11}$$

Using the spectral decomposition $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} = \sum_{i=1}^n \rho_i\mathbf{u}_i\mathbf{u}_i^T$ and the fact that $\mathbf{u}_1 = \sqrt{\mathbf{d}} = \mathbf{D}^{1/2}\mathbf{1}$, we can write (11) as

$$|(1 - \rho)x_1y_1 + \rho_2x_2y_2 + \sum_{i=3}^n \rho_ix_iy_i|, \tag{12}$$

where $\mathbf{x} = \sum_{i=1}^n x_i\mathbf{u}_i$ and $\mathbf{y} = \sum_{i=1}^n y_i\mathbf{u}_i$ is the expansion of $\mathbf{x}$ and $\mathbf{y}$ in the orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ with coordinates $x_i = \mathbf{x}^T\mathbf{u}_i$ and $y_i = \mathbf{y}^T\mathbf{u}_i$, respectively.

First we will prove that $1 - \rho$ is governed by ρ_2; more precisely, $|1 - \rho| \le |\rho_2| + \varepsilon$. Applying the arguments of Lemma 1 and the above formulas for the special $A, B \subset V$ yields

$$\text{Vol}(A)\text{Vol}(B) \cdot (\rho - 1) = w(A, B) - \text{Vol}(A)\text{Vol}(B) =$$
$$= \mathbf{a}^T(\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T)\mathbf{b} = \rho_2a_2b_2 + \sum_{i=3}^n \rho_ia_ib_i, \tag{13}$$

where $\mathbf{a} = \sum_{i=1}^n a_i\mathbf{u}_i$ and $\mathbf{b} = \sum_{i=1}^n b_i\mathbf{u}_i$ is the expansion of $\mathbf{a}$ and $\mathbf{b}$ in the orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$, respectively. The separation of A and B is based on the vector $\mathbf{D}^{-1/2}\mathbf{u}_2$ which has both negative and positive coordinates, since $\mathbf{u}_2$ is orthogonal to $\mathbf{u}_1$ of all positive coordinates. With formulas, $\mathbf{a} + \mathbf{b} = \mathbf{u}_1$, and hence, $a_2 + b_2 = \mathbf{u}_1^T\mathbf{u}_2 = 0$. (If it is the eigenvalue λ_2 of the normalized Laplacian that is the farthest from 1, then the corresponding eigenvector, our $\mathbf{u}_2$, is also called "Fiedler-vector" as the two-partition of the vertices into two loosely connected parts was based on the signs of its coordinates in the early paper of Fiedler [12]). If θ is much larger than ε, the first term in the last formula of (13) – apart from a term of $\mathcal{O}(|\varepsilon|)$ – will dominate the sign of $\rho - 1$ which is therefore opposite to the sign of ρ_2.

Therefore, we will distinguish between two cases.

- If $\lambda_2 < 1 - \varepsilon$, then $\rho_2 = 1 - \lambda_2 > \varepsilon > 0$, and in view of the inequalities between the minimum normalized cut and the smallest positive normalized Laplacian eigenvalue (apply (7) for the $k = 2$ case):

$$\rho \ge f_2(G) = \min_{U \subset V} \frac{w(U, \bar{U})}{\text{Vol}(U)\text{Vol}(\bar{U})} \ge \lambda_2 = 1 - \rho_2, \tag{14}$$

  therefore $1 - \rho \le \rho_2$, as $1 - \rho$ is also positive due to the considerations before. Further, the estimation, due to (4),

$$S_2^2(\mathbf{D}^{-1/2}\mathbf{u}_2) \le \frac{\lambda_2}{\lambda_3} \le \frac{1 - \rho_2}{1 - \varepsilon} = \frac{1 - \theta}{1 - \varepsilon} \tag{15}$$

  also follows.

- If $1 - \varepsilon \le \lambda_2 \le \frac{n}{n-1}$, then – provided $\frac{1}{n-1} \le \varepsilon$ – it is the eigenvalue λ_n that is the farthest from 1, and hence, greater than $1 + \varepsilon$. Consequently, $-\varepsilon < \rho_2 = 1 - \lambda_n < 0$, and hence, by (8) and (9):

$$\rho_2 + \rho_{neg} \le Q_2((A, B), \mathbf{W}) = (2 - 1) - f_2((A, B), \mathbf{W}) = 1 - \rho,$$

where $\rho_{neg} = \min\{1 - \lambda_{n-1}, 0\}$, and $|\rho_{neg}| < \varepsilon$. Note, that in this case $1 - \rho$ is negative that yields $|1 - \rho| \leq |\rho_2| + \varepsilon$. Now the optimum A, B is obtained by minimizing the 2-variance of the coordinates of the transformed eigenvector $\mathbf{D}^{-1/2}\mathbf{u}_2$ (now $\mathbf{u}_2$ belongs to λ_n and ρ_2 at the same time) for which the following relation – like (4) – can be proved:

$$S_2^2(\mathbf{D}^{-1/2}\mathbf{u}_2) = \mathcal{O}(\frac{2 - \lambda_n}{2 - \lambda_{n-1}}) = \mathcal{O}(\frac{\beta_n + 1}{\beta_{n-1} + 1}) = \mathcal{O}(\frac{1 - \theta}{1 - \varepsilon}), \tag{16}$$

where β's are eigenvalues of the normalized modularity matrix. Indeed, in lack of dominant vertices, there is a relation between the largest and smallest normalized Laplacian eigenvalues of G and $\overline{G}$, respectively, where the complement graph $\overline{G} = (V, \overline{\mathbf{W}})$ is defined such that $\overline{w}_{ij} = 1 - w_{ij}$ $(i \neq j)$ and $\overline{w}_{ii} = 0$ $(i = 1, \ldots, n)$.

If the two largest absolute value eigenvalues of the normalized modularity matrix are of different sign, then we are able to find a gap at least $\theta - \varepsilon$ between eigenvalues of the same sign.

Therefore, (12) can be estimated from above with

$$|\rho_2| \cdot |x_1 y_1 + x_2 y_2| + \varepsilon x_1 y_1 + \max_{i \geq 3} |\rho_i| \cdot |\sum_{i=3}^{n} x_i y_i|. \tag{17}$$

As for the second term, $\varepsilon x_1 y_1 = \varepsilon \mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)$, so it does not need further treatment. Using the Cauchy-Schwarz inequality, the last term can be estimated from above with

$$\varepsilon\sqrt{\sum_{i=3}^{n} x_i^2 \sum_{i=3}^{n} y_i^2}\varepsilon\sqrt{\sum_{i=2}^{n} x_i^2 \sum_{i=2}^{n} y_i^2} \leq \varepsilon\sqrt{\mathrm{Vol}\,(X)(1 - \mathrm{Vol}\,(X))\mathrm{Vol}\,(Y)(1 - \mathrm{Vol}\,(Y))} \leq \varepsilon\sqrt{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)},$$

since $x_1 = \mathrm{Vol}\,(X)$, $y_1 = \mathrm{Vol}\,(Y)$ and $\sum_{i=1}^{n} x_i^2 = \|\mathbf{x}\|^2 = \mathrm{Vol}\,(X)$, $\sum_{i=1}^{n} y_i^2 = \|\mathbf{y}\|^2 = \mathrm{Vol}\,(Y)$. The first term is reminiscent of an equation for the coordinates of orthogonal vectors. Therefore, we project the vectors $\mathbf{u}_1$, $\mathbf{u}_2$ onto the subspace $F = \mathrm{Span}\,\{\mathbf{a}, \mathbf{b}\}$. In fact, $\mathbf{u}_1 = \mathbf{a} + \mathbf{b}$, and hence, $\mathbf{u}_1 \in F$. The vector $\mathbf{u}_2$ can be decomposed as

$$\mathbf{u}_2 = \frac{\mathbf{u}_2^T \mathbf{a}}{\mathrm{Vol}\,(A)}\mathbf{a} + \frac{\mathbf{u}_2^T \mathbf{b}}{\mathrm{Vol}\,(B)}\mathbf{b} + \mathbf{q}, \tag{18}$$

where $\mathbf{q}$ is the component orthogonal to F. For the squared distance $\|\mathbf{q}\|^2$ between $\mathbf{u}_2$ and F, in [4], we proved that it is equal to the weighted 2-variance $S_2^2(\mathbf{D}^{-1/2}\mathbf{u}_2)$ and in (15) we estimated it from above with $\frac{1-\theta}{1-\varepsilon}$. (In the $\rho_2 = 1 - \lambda_n$ case similar upper estimation works using (16)). Let s^2 denote this minimum 2-variance of the coordinates of $\mathbf{D}^{-1/2}\mathbf{u}_2$ (in both cases).

To estimate $a_1 b_1 + a_2 b_2 = (\mathbf{u}_1^T \mathbf{a})(\mathbf{u}_1^T \mathbf{b}) + (\mathbf{u}_2^T \mathbf{a})(\mathbf{u}_2^T \mathbf{b})$, the problem is that the pairwise orthogonal vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{a}$, $\mathbf{b}$ are not in the same subspace of $\mathbb{R}^n$ as, in general, $\mathbf{u}_2 \notin F$. However, by an argument proved in [4], we can find orthogonal, unit-norm vectors $\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2 \in F$ such that

$$\|\mathbf{u}_1 - \tilde{\mathbf{u}}_1\|^2 + \|\mathbf{u}_2 - \tilde{\mathbf{u}}_2\|^2 \leq 2s^2, \tag{19}$$

where, in view of $\mathbf{u}_1 \in F$, $\tilde{\mathbf{u}}_1 = \mathbf{u}_1$. Let $\mathbf{r} := \mathbf{u}_2 - \tilde{\mathbf{u}}_2$. Since $\tilde{\mathbf{u}}_1^T \mathbf{a}, \tilde{\mathbf{u}}_2^T \mathbf{a}$ and $\tilde{\mathbf{u}}_1^T \mathbf{b}, \tilde{\mathbf{u}}_2^T \mathbf{b}$ are coordinates of the orthogonal vectors $\mathbf{a}, \mathbf{b}$ in the basis $\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2$,

$$(\tilde{\mathbf{u}}_1^T \mathbf{a})(\tilde{\mathbf{u}}_1^T \mathbf{b}) + (\tilde{\mathbf{u}}_2^T \mathbf{a})(\tilde{\mathbf{u}}_2^T \mathbf{b}) = 0,$$

and because of $\tilde{\mathbf{u}}_2^T \mathbf{a} + \tilde{\mathbf{u}}_2^T \mathbf{b} = \tilde{\mathbf{u}}_2^T \mathbf{u}_1 = 0$,

$$\tilde{\mathbf{u}}_2^T \mathbf{a} = -\tilde{\mathbf{u}}_2^T \mathbf{b} = \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)} =: c.$$

Therefore,

$$
\begin{aligned}
|(\mathbf{u}_1^T \mathbf{a})(\mathbf{u}_1^T \mathbf{b}) + (\mathbf{u}_2^T \mathbf{a})(\mathbf{u}_2^T \mathbf{b})| &= |\mathrm{Vol}\,(A)\mathrm{Vol}\,(B) + [(\tilde{\mathbf{u}}_2 + \mathbf{r})^T \mathbf{a}][(\tilde{\mathbf{u}}_2 + \mathbf{r})^T \mathbf{b}]| \\
&= |\mathrm{Vol}\,(A)\mathrm{Vol}\,(B) + [c + \mathbf{r}^T \mathbf{a}][-c + \mathbf{r}^T \mathbf{b}]| = |c(-\mathbf{r}^T \mathbf{a} + \mathbf{r}^T \mathbf{b}) + (\mathbf{r}^T \mathbf{a})(\mathbf{r}^T \mathbf{b})| \\
&\le |c|\sqrt{\|\mathbf{r}\|^2 \|\mathbf{b} - \mathbf{a}\|^2} + \sqrt{\|\mathbf{r}\|^2 \|\mathbf{a}\|^2}\sqrt{\|\mathbf{r}\|^2 \|\mathbf{b}\|^2} \\
&\le \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}(\|\mathbf{r}\| + \|\mathbf{r}\|^2) \le \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}(\sqrt{2}s + 2s^2),
\end{aligned}
$$

using (19) and the fact that $\|\mathbf{b} - \mathbf{a}\|^2 = 1$.
Now we estimate $x_1 y_1 + x_2 y_2 = (\mathbf{u}_1^T \mathbf{x})(\mathbf{u}_1^T \mathbf{y}) + (\mathbf{u}_2^T \mathbf{x})(\mathbf{u}_2^T \mathbf{y})$. Going back to (18) we have

$$\mathbf{u}_2^T \mathbf{x} = \frac{\mathbf{u}_2^T \mathbf{a}}{\mathrm{Vol}\,(A)} \mathbf{a}^T \mathbf{x} + \frac{\mathbf{u}_2^T \mathbf{b}}{\mathrm{Vol}\,(B)} \mathbf{b}^T \mathbf{x} + \mathbf{q}^T \mathbf{x} = \frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(A)} \mathbf{u}_2^T \mathbf{a} + \mathbf{q}^T \mathbf{x},$$

and similarly,

$$\mathbf{u}_2^T \mathbf{y} = \frac{\mathbf{u}_2^T \mathbf{a}}{\mathrm{Vol}\,(A)} \mathbf{a}^T \mathbf{y} + \frac{\mathbf{u}_2^T \mathbf{b}}{\mathrm{Vol}\,(B)} \mathbf{b}^T \mathbf{y} + \mathbf{q}^T \mathbf{y} = \frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(B)} \mathbf{u}_2^T \mathbf{b} + \mathbf{q}^T \mathbf{y},$$

that in view of $\|\mathbf{q}\|^2 = s^2$ yields

$$
\begin{aligned}
x_1 y_1 + x_2 y_2 &= |(\mathbf{u}_1^T \mathbf{x})(\mathbf{u}_1^T \mathbf{y}) + (\mathbf{u}_2^T \mathbf{x})(\mathbf{u}_2^T \mathbf{y})| = \\
&|\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y) + (\frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(A)} \mathbf{u}_2^T \mathbf{a} + \mathbf{q}^T \mathbf{x})(\frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(B)} \mathbf{u}_2^T \mathbf{b} + \mathbf{q}^T \mathbf{y})| \\
&\le |\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y) + (\frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(A)} \mathbf{u}_2^T \mathbf{a})(\frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(B)} \mathbf{u}_2^T \mathbf{b})| \\
&+ |(\mathbf{q}^T \mathbf{x})(\frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(B)} \mathbf{u}_2^T \mathbf{b}) + (\mathbf{q}^T \mathbf{y})(\frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(A)} \mathbf{u}_2^T \mathbf{a}) + (\mathbf{q}^T \mathbf{x})(\mathbf{q}^T \mathbf{y})| \\
&\le \frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(A)} \frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(B)} |\mathrm{Vol}\,(A)\mathrm{Vol}\,(B) + (\mathbf{u}_2^T \mathbf{a})(\mathbf{u}_2^T \mathbf{b})| \\
&+ \|\mathbf{q}\|\|\mathbf{x}\| \frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(B)} \|\mathbf{u}_2\|\|\mathbf{b}\| + \|\mathbf{q}\|\|\mathbf{y}\| \frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(A)} \|\mathbf{u}_2\|\|\mathbf{a}\| + \|\mathbf{q}\|^2 \|\mathbf{x}\|\|\mathbf{y}\| \\
&\le \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}(\sqrt{2}s + 2s^2) + \|\mathbf{q}\|\sqrt{\mathrm{Vol}\,(X)} \frac{\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(B)} \sqrt{\mathrm{Vol}\,(B)} \\
&+ \|\mathbf{q}\|\sqrt{\mathrm{Vol}\,(Y)} \frac{\mathrm{Vol}\,(X)}{\mathrm{Vol}\,(A)} \sqrt{\mathrm{Vol}\,(A)} + \|\mathbf{q}\|^2 \sqrt{\mathrm{Vol}\,(X)}\sqrt{\mathrm{Vol}\,(Y)} \\
&= \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}(\sqrt{2}s + 2s^2) + \|\mathbf{q}\|\sqrt{\mathrm{Vol}\,(X)}\sqrt{\mathrm{Vol}\,(Y)}(\frac{\sqrt{\mathrm{Vol}\,(Y)}}{\sqrt{\mathrm{Vol}\,(B)}} + \frac{\sqrt{\mathrm{Vol}\,(X)}}{\sqrt{\mathrm{Vol}\,(A)}} + \|\mathbf{q}\|) \\
&\le \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}[(\sqrt{2}s + 2s^2 + s(2 + s)] = \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}[(\sqrt{2} + 2)s + 3s^2] \\
&\le \sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}(\sqrt{2} + 5)s.
\end{aligned}
$$

Summarizing, the second and third terms in (17) are estimated from above with $\varepsilon\sqrt{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)} \le \varepsilon\sqrt{\mathrm{Vol}\,(A)\mathrm{Vol}\,(B)}$. Because of $\varepsilon < \theta$, by an easy calculation it follows

that it is less than $\sqrt{\frac{1-\theta}{1-\varepsilon}}$. Therefore, the constant α of the (A, B) pair's regularity is $\mathcal{O}(\sqrt{\frac{1-\theta}{1-\varepsilon}})$.

∎

Remark 1 *The statement has relevance only if θ is much larger than ε. In this case the spectral gap between the largest absolute value eigenvalue and the others in the normalized modularity spectrum indicates a regular 2-partition of the graph that can be constructed based on the eigenvector belonging to the structural eigenvalue.*

4   Analysis of Variance setup (the $k > 2$ case)

Theorem 2 *Let $G = (V, \mathbf{W})$ is an edge-weighted graph on n vertices, with generalized degrees $d_1, \ldots, d_n$ and $\mathbf{D} = diag(d_1, \ldots, d_n)$. Suppose that $Vol(V) = 1$. Let the eigenvalues of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, enumerated in decreasing absolute values, be*

$$1 = \rho_1 > |\rho_2| \geq \cdots \geq |\rho_k| > \varepsilon \geq |\rho_i|, \quad i \geq k+1.$$

The partition $(V_1, \ldots, V_k)$ of V is defined so that it minimizes the weighted k-variance of the vertex representatives obtained as row vectors of the $n \times k$ matrix $\mathbf{X}$ of column vectors $\mathbf{D}^{-1/2}\mathbf{u}_i$, where $\mathbf{u}_i$ is the unit-norm eigenvector belonging to ρ_i $(i = 1, \ldots, k)$. With the notation $s^2 = S_k^2(\mathbf{X})$, the (V_i, V_j) pairs are $2(\sqrt{2}s + \varepsilon)$-volume regular $(i \neq j)$ and for the clusters V_i $(i = 1, \ldots, k)$ the following holds: for all $X, Y \subset V_i$ we have that

$$|w(X, Y) - \rho(V_i) Vol(X) Vol(Y)| \leq 2(\sqrt{2}s + \varepsilon) Vol(V_i), \tag{20}$$

where $\rho(V_i) = \frac{w(V_i, V_i)}{Vol^2(V_i)}$ is the relative intra-cluster density of V_i.

Proof   Denoting by $\mathbf{u}_1, \ldots, \mathbf{u}_k$ the eigenvectors belonging to the so-called structural eigenvalues $\rho_1, \ldots, \rho_k$, the representatives $\mathbf{r}_1, \ldots, \mathbf{r}_n$ of the vertices are row vectors of the matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$, where $\mathbf{x}_i = \mathbf{D}^{-1/2}\mathbf{u}_i$ $(i = 1, \ldots, k)$ and the trivial $\mathbf{x}_1 = \mathbf{1}$ (belonging to $\rho_1 = 1$) can be omitted, see (5). The minimum k-variance $S_k^2(\mathbf{X})$ of the k-dimensional (actually, $(k-1)$-dimensional) representatives is as small as s^2. Suppose that the minimum k-variance is attained by the k-partition $(V_1, \ldots, V_k)$ of the vertices.
By an easy analysis of variance argument of [5, 6] it follows that

$$s^2 = \sum_{i=1}^{k} \mathrm{dist}^2(\mathbf{u}_i, F),$$

where $F = \mathrm{Span}\{\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_k\}$ with the so-called normalized partition vectors $\mathbf{z}_1, \ldots, \mathbf{z}_k$ of coordinates $z_{ji} = \frac{1}{\sqrt{Vol(V_i)}}$ if $j \in V_i$ and 0, otherwise $(i = 1, \ldots, k)$. Note that the vectors $\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_k$ form an orthonormal system. By [4, 5] we can find another orthonormal system $\mathbf{v}_1, \ldots, \mathbf{v}_k \in F$ such that

$$\sum_{i=1}^{k} \|\mathbf{u}_i - \mathbf{v}_i\|^2 \leq 2s^2.$$

With these vectors, we construct the following k-rank approximation of the matrix $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} = \sum_{i=1}^{n} \rho_i \mathbf{u}_i \mathbf{u}_i^T$: it is approximated by $\sum_{i=1}^{k} \rho_i \mathbf{v}_i \mathbf{v}_i^T$ with the following accuracy (in spectral norm):

$$\|\sum_{i=1}^{n} \rho_i \mathbf{u}_i \mathbf{u}_i^T - \sum_{i=1}^{k} \rho_i \mathbf{v}_i \mathbf{v}_i^T\| \leq \sum_{i=1}^{k} |\rho_i| \cdot \|\mathbf{u}_i \mathbf{u}_i^T - \mathbf{v}_i \mathbf{v}_i^T\| + \|\sum_{i=k+1}^{n} \rho_i \mathbf{u}_i \mathbf{u}_i^T\| \leq \sqrt{\sum_{i=1}^{k} \sin^2 \sigma_i} + \varepsilon \leq \sqrt{2}s + \varepsilon,$$

(21)

where σ_i is the angle between $\mathbf{u}_i$ and $\mathbf{v}_i$, and for it, $\sin \frac{\sigma_i}{2} = \frac{1}{2}\|u_i - v_i\|$ holds, therefore

$$\sin^2 \sigma_i = (2 \sin \frac{\sigma_i}{2} \cos \frac{\sigma_i}{2})^2 = \frac{1}{4}\|u_i - v_i\|^2 (4 - \|u_i - v_i\|^2), \quad i = 1, \ldots, k.$$

Hence, the above difference can be estimated from above with $\sqrt{2}s + \varepsilon$ in spectral norm. Based on these considerations and the fact that the cut norm is less than or equal to the spectral norm, the densities to be estimated in the defining formula (10) of volume regularity can be written in terms of stepwise constant vectors in the following way. The vectors $\mathbf{y}_i := \mathbf{D}^{-1/2}\mathbf{v}_i$ are stepwise constants on the partition $(V_1, \ldots, V_k)$, $i = 1, \ldots, k$. The matrix $\sum_{i=1}^{k} \rho_i \mathbf{y}_i \mathbf{y}_i^T$ is therefore a symmetric block-matrix on $k \times k$ blocks belonging to the above partition of the vertices. Let $\tilde{w}_{ab}$ denote its entries in the (a,b) block $(a, b = 1, \ldots, k)$. Using (21), the following approximation of the matrix $\mathbf{W}$ is performed:

$$\|\mathbf{W} - \mathbf{D}(\sum_{i=1}^{k} \rho_i \mathbf{y}_i \mathbf{y}_i^T)\mathbf{D}\| = \|\mathbf{D}^{1/2}(\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} - \sum_{i=1}^{k} \rho_i \mathbf{v}_i \mathbf{v}_i^T)\mathbf{D}^{1/2}\| \leq \|\mathbf{D}\|^{1/2}(\sqrt{2}s + \varepsilon)\|\mathbf{D}\|^{1/2}.$$

Therefore, the entries of $\mathbf{W}$ – for $i \in V_a$, $j \in V_b$ – can be decomposed as

$$w_{ij} = d_i d_j \tilde{w}_{ab} + \eta_{ij},$$

where the cut norm and spectral norm of the $n \times n$ symmetric error matrix $\mathbf{E} = (\eta_{ij})$ is at most $\|\mathbf{D}\|(\sqrt{2}s + \varepsilon)$. But we will restrict the error matrix to $V_a \times V_b$: its entries are η_{ij}'s for $i \in V_a, j \in V_b$, and zeros otherwise. Denoting the restricted matrix by $\mathbf{E}^{ab}$, and the restricted diagonal matrices by $\mathbf{D}^a$ and $\mathbf{D}^b$, respectively, the following finer estimation holds:

$$\|\mathbf{E}^{ab}\|_\square \leq |\mathbf{D}^a\|^{1/2} \cdot \|\mathbf{D}^b\|^{1/2} \cdot (\sqrt{2}s + \varepsilon) \leq \sqrt{\mathrm{Vol}\,(V_a)\mathrm{Vol}\,(V_b)}(\sqrt{2}s + \varepsilon).$$

Consequently, for $a, b = 1, \ldots, k$:

$$|w(X, Y) - \rho(V_a, V_b)\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)| = \left| \sum_{i \in X} \sum_{j \in Y} (d_i d_j \tilde{w}_{ab} + \eta_{ij}^{ab}) - \frac{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(V_a)\mathrm{Vol}\,(V_b)} \sum_{i \in V_a} \sum_{j \in V_b} (d_i d_j \tilde{w}_{ab} + \eta_{ij}^{ab}) \right|$$

$$= \left| \sum_{i \in X} \sum_{j \in Y} \eta_{ij}^{ab} - \frac{\mathrm{Vol}\,(X)\mathrm{Vol}\,(Y)}{\mathrm{Vol}\,(V_a)\mathrm{Vol}\,(V_b)} \sum_{i \in V_a} \sum_{j \in V_b} \eta_{ij}^{ab} \right| \leq 2(\sqrt{2}s + \varepsilon)\sqrt{\mathrm{Vol}\,(V_a)\mathrm{Vol}\,(V_b)},$$

that gives the required statement both in the $a \neq b$ and $a = b$ case.

Remark 2 *In the $k = 2$ case, the estimate of Theorem 1 has the same order of magnitude as that of Theorem 2, since $s^2 = \mathcal{O}(\sqrt{\frac{1-\theta}{1-\varepsilon}})$. The statement has only relevance for an integer*

$k \in [2, n)$ *such that there is a remarkable spectral gap between* $\theta := |\rho_k|$ *and* $|\rho_{k+1}|$ *in the normalized modularity spectrum, i.e., the so-called structural eigenvalues* $\rho_1, \ldots, \rho_k$ *are far apart from zero, while the others are in an* ε *distance from zero, in absolute value. This is a necessary condition for* s^2 *to be "small". As it is not sufficient, instead of* θ *and* ε, *the estimation of Theorem 2 is given in terms of* s *and* ε. *Indeed, by perturbation results of spectral subspaces for symmetric matrices [3],* s^2 *itself can be estimated from above by the spectral gap between the* k *structural and the other eigenvalues when* $\rho_2, \ldots, \rho_k$ *have the same sign (the situation of strong community or anti-community structure).*

Acknowledgement

References

[1] N. Alon, Eigenvalues and expanders, *Combinatorica* **6**, 86-96, 1986.

[2] N. Alon, A. Coja-Oghlan, H. Han, M. Kang, V. Rödl, and M. Schacht, Quasi-randomness and algorithmic regularity for graphs with general degree distributions, *Siam J. Comput.* **39** (6), 2336-2362, 2010.

[3] R. Bathia, *Matrix analysis*, Springer, New York, 1997.

[4] M. Bolla and G. Tusnády, Spectra and Optimal Partitions of Weighted Graphs, *Discrete Mathematics* **128**, 1-20, 1994.

[5] M. Bolla and G. Molnár–Sáska, Optimization problems for weighted graphs and related correlation estimates, *Discrete Mathematics* **282**, 23-33, 2004.

[6] M. Bolla, Noisy random graphs and their Laplacians, *Discrete Mathematics* **308**, 4221-4230, 2008.

[7] M. Bolla, Penalized versions of the Newman–Girvan modularity and their relation to multi-way cuts and k-means clustering, preprint.

[8] F. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics **92**, American Mathematical Society, 1997.

[9] F. Chung and R. Graham, Quasi-random graphs with given degree sequences, *Random Structures and Algorithms* **12**, 1-19, 2008.

[10] A. Coja-Oghlan, Graph Partitioning via Adaptive Spectral Techniques, *Combinatorics, Probability and Computing* **19**, 227-284, 2010.

[11] A. Coja-Oghlan and A. Lanka, The spectral gap of random graphs with given expected degrees, *The Electronic Journal of Combinatorics* **16**, R138, 2009.

[12] Fiedler, M., Algebraic connectivity of graphs, *Czechoslovak Math. J.* **23**, 298-305, 1973.

[13] Frieze, A., Kannan, R., Quick approximation to matrices and applications, *Combinatorica* **19**, 175-220, 1999.

[14] S. Hoory, N. Linial, and A. Widgerson, Expander graphs and their applications, *Bulletin (New series) of the American Mathematical Society* **43** (4), 439-561, 2006.

[15] L. Lovász and V. T. Sós, Generalized quasirandom graphs, *J. Comb. Theory B* **98**, 146-163, 2008.

[16] McSherry, F., Spectral partitioning of random graphs, *Proc. 42nd FOCS*, 529-537, 2001.

[17] M. Meilă and J. Shi, Learning segmentation by random walks, *Proceedings of the NIPS (Neural Information Processing Systems) 13 Conference*, ed. T. K. Leen, T. G. Dietterich, and V. Tresp, MIT Press, Cambridge, 873-879, 2001.

[18] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Physical Review E* **69**, 026113, 2004.