Modularity spectra, eigen-subspaces, and structure of weighted graphs

Marianna Bolla *

Institute of Mathematics, Budapest University of Technology and Economics and Intra-University Centre for Telecommunications and Informatics, Debrecen

Abstract

The role of the normalized modularity matrix in finding homogeneous cuts will be presented. We also discuss the testability of the structural eigenvalues and that of the subspace spanned by the corresponding eigenvectors of this matrix. In the presence of a spectral gap between the $k-1$ largest absolute value eigenvalues and the remainder of the spectrum, this in turn implies the testability of the sum of the inner variances of the k clusters that are obtained by applying the k-means algorithm for the appropriately chosen vertex representatives.

Key words: Normalized modularity, Volume regularity, Spectral clustering, Testable weighted graph parameters, Structural eigenvalues, Spectral subspaces

1 Introduction

The purpose of this paper is to summarize the spectral properties and testability of the spectrum and spectral subspaces of the normalized modularity matrix introduced in [9] to find regular vertex partitions. We will generalize the Laplacian based spectral clustering methods to recover so-called volume regular cluster pairs such that the information flow between the pairs and within the clusters is as homogeneous as possible. For this purpose, we take into consideration both ends of the normalized Laplacian spectrum, i.e., large absolute value, so-called structural eigenvalues of our normalized modularity matrix introduced just for this convenience.

* Research supported by the Hungarian National Research Grant OTKA-KTIA 77778 and by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project, supported by the European Union, co-financed by the European Social Fund.

Email address: marib@math.bme.hu (Marianna Bolla).

In Theorem 3, we estimate the constant of volume regularity in terms of the gap between the structural and other eigenvalues, and the k-variance of the optimal vertex representatives constructed by the eigenvectors corresponding to the structural eigenvalues. Here we give a more detailed proof of this statement than in [10]. This theorem implies that for a general edge-weighted graph, the existence of $k-1$ structural eigenvalues of the normalized modularity matrix, separated from 0, is indication of a k-cluster structure such that the cluster-pairs are volume regular with constant depending on the spectral gap and the above k-variance. The clusters themselves can be recovered by applying the k-means algorithm for the vertex representatives. Hence, Theorem 3 implies that spectral clustering of the vertices into k parts gives satisfactory partition in the sense of volume regularity.

Furthermore, in Theorems 8 and 10, we prove the testability of the structural eigenvalues and the corresponding eigen-subspace of the normalized modularity matrix in the sense of [12]. In view of this, spectral clustering methods can be performed on a smaller part of the underlying graph and give good approximation for the cluster structure.

2 Preliminaries

Throughout the paper, we use the general framework of an edge-weighted graph. Let $G = G_n = (V, \mathbf{W})$ be an edge-weighted graph on vertex-set V ($|V| = n$) and $n \times n$ symmetric weight-matrix $\mathbf{W}$ of non-negative real entries and zero diagonal. We will call the numbers $d_i = \sum_{j=1}^{n} w_{ij}$ $(i = 1, \ldots, n)$ *generalized degrees*, and the diagonal matrix $\mathbf{D} = \mathtt{diag}\,(d_1, \ldots, d_n)$ *degree matrix*. In this and the next section, without loss of generality, $\mathtt{Vol}(V) = 1$ will be assumed, where the *volume* of the vertex-subset $U \subseteq V$ is $\mathtt{Vol}(U) = \sum_{i \in U} d_i$. In the sequel, we only consider connected graphs, which means that $\mathbf{W}$ is irreducible.

In [9], we defined the normalized version of the modularity matrix (introduced in [21]) as $\mathbf{M}_D = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T$, where $\sqrt{\mathbf{d}} = (\sqrt{d_1}, \ldots, \sqrt{d_n})^T$, and we called it *normalized modularity matrix*. The spectrum of this matrix is in the [-1,1] interval, and 0 is always an eigenvalue with unit-norm eigenvector $\sqrt{\mathbf{d}}$. Indeed, in [5] we proved that 1 is a single eigenvalue of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ with corresponding unit-norm eigenvector $\sqrt{\mathbf{d}}$, provided our graph is connected. This becomes a zero eigenvalue of $\mathbf{M}_D$ with the same eigenvector, whence 1 cannot be an eigenvalue of $\mathbf{M}_D$ if G is connected. In fact, the introduction of this matrix is rather technical, the spectral gap, further, Lemma 1 and Theorem 3 can better be formulated with it. It can also be obtained from the normalized Laplacian by subtracting it from the identity and depriving of its trivial factor. Normalized Laplacian was used for spectral clustering in several

papers (e.g., [3,5,6,14,20]), the idea of which can be summarized by means of the spectral decomposition of the normalized modularity matrix. We introduce the following notation: the *weighted cut* between the vertex-subsets $X, Y \subseteq V$ is $w(X,Y) = \sum_{i\in X}\sum_{j\in Y} w_{ij}$. We will frequently refer to the following facts.

(a) The spectral decomposition of $\mathbf{M}_D$ solves the following quadratic placement problem. For a given positive integer k ($1 < k < n$), we want to minimize $Q_k = \sum_{i<j} w_{ij}\|\mathbf{r}_i - \mathbf{r}_j\|^2$ on the conditions

$$\sum_{i=1}^{n} d_i \mathbf{r}_i \mathbf{r}_i^T = \mathbf{I}_{k-1} \quad \text{and} \quad \sum_{i=1}^{n} d_i \mathbf{r}_i = \mathbf{0} \tag{1}$$

where the vectors $\mathbf{r}_1, \ldots, \mathbf{r}_n$ are $(k-1)$-dimensional *representatives* of the vertices, which form the row vectors of the $n \times (k-1)$ matrix $\mathbf{X}$. Denote the eigenvalues of $\mathbf{M}_D$, in decreasing order, by $1 > \lambda_1 \geq \cdots \geq \lambda_n \geq -1$ with corresponding unit-norm, pairwise orthogonal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. In [5], we proved that the minimum of Q_k subject to (1) is $k-1-\sum_{i=1}^{k-1}\lambda_i$ and is attained by the representation such that the optimum vertex representatives $\mathbf{r}_1^*, \ldots, \mathbf{r}_n^*$ are row vectors of the matrix $\mathbf{X}^* = (\mathbf{D}^{-1/2}\mathbf{u}_1, \ldots, \mathbf{D}^{-1/2}\mathbf{u}_{k-1})$. Instead of $\mathbf{X}$, the augmented $n \times k$ matrix $\tilde{\mathbf{X}}$ can as well be used, which is obtained from $\mathbf{X}$ by inserting the column $\mathbf{x}_0 = \mathbf{1}$ of all 1's. In fact, $\mathbf{x}_0 = \mathbf{D}^{-1/2}\mathbf{u}_0$, where $\mathbf{u}_0 = \sqrt{\mathbf{d}}$ is the eigenvector corresponding to the eigenvalue 1 of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$. Then

$$Q_k = \mathtt{tr}\,(\mathbf{D}^{1/2}\tilde{\mathbf{X}})^T(\mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})(\mathbf{D}^{1/2}\tilde{\mathbf{X}}),$$

and minimizing Q_k on the constraint (1) is equivalent to minimizing the above expression subject to $\tilde{\mathbf{X}}^T\mathbf{D}\tilde{\mathbf{X}} = \mathbf{I}_k$. This problem is the *continuous relaxation* of minimizing

$$Q_k(P_k) = \mathtt{tr}\,(\mathbf{D}^{1/2}\tilde{\mathbf{X}}(P_k))^T(\mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})(\mathbf{D}^{1/2}\tilde{\mathbf{X}}(P_k))$$

over the set of k-partitions $P_k = (V_1, \ldots, V_k)$ of the vertices such that P_k is planted into $\tilde{\mathbf{X}}$ in the way that the columns of $\tilde{\mathbf{X}}(P_k)$ are so-called *normalized partition-vectors* belonging to P_k. Namely, the coordinates of the ith column are zeros, except those indexing vertices of V_i, which are equal to $\frac{1}{\sqrt{\mathtt{Vol}(V_i)}}$ ($i = 1, \ldots, k$). In fact, this is the *normalized cut* problem, which is discussed in [20] for $k = 2$, further, in [3] and [6] for a general k, and the solution is based on the above continuous relaxation.

(b) Now, let us maximize the *normalized Newman–Girvan modularity* of G induced by P_k, defined in [9] as

$$M_k(P_k) = \sum_{a=1}^{k} \frac{1}{\mathtt{Vol}(V_a)} \sum_{i,j\in V_a} (w_{ij} - d_i d_j) = \sum_{a=1}^{k} \frac{w(V_a, V_a)}{\mathtt{Vol}(V_a)} - 1$$

over the set $\mathcal{P}_k$ of the k-partitions of V. It is easy to see that $M_k(P_k) =$

$k-1-Q_k(P_k)$, and hence, the above task has the same spectral relaxation as the normalized cut problem. Let $M_k = \max_{P_k \in \mathcal{P}_k} M_k(P_k)$ denote the maximum k-way normalized Newman-Girvan modularity of the weighted graph G.

(c) Finally, from the above considerations it is straightforward that $M_k \leq \sum_{i=1}^{k-1} \lambda_i$, or equivalently, the minimum normalized k-way cut is at least the the sum of the $k-1$ smallest positive normalized Laplacian eigenvalues. As for the minimum normalized k-way cut, in [6] we also gave an upper estimate by constant times the sum of the $k-1$ smallest positive normalized Laplacian eigenvalues, which constant depends on the so-called *k-variance* of the vertex representatives defined in the following way.

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

where $\mathbf{c}_a = \frac{1}{\mathtt{Vol}(V_a)} \sum_{j \in V_a} d_j \mathbf{r}_j$ is the weighted center of cluster V_a and $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^{k-1}$ are rows of $\mathbf{X}$. (The augmented $\tilde{\mathbf{X}}$ would give the same k-variance.) The constant of our estimation depended on $S_k^2(\mathbf{X}^*)$, and it was close to 1 if this k-variance of the optimum $(k-1)$-dimensional vertex representatives was small enough. Note that $S_k^2(\mathbf{X}, P_k)$ is the objective function of the weighted k-means algorithm.

In this way, we showed that large positive eigenvalues of the normalized modularity matrix are responsible for clusters with high intra- and low inter-cluster densities. Likewise, maximizing $Q_k(P_k)$ instead of minimizing over $\mathcal{P}_k$, small negative eigenvalues of the normalized modularity matrix are responsible for clusters with low intra- and high inter-cluster densities (see [9]). Our idea is that taking into account eigenvalues from both ends of the normalized modularity spectrum, we can recover so-called regular cluster pairs. For this purpose, we use the notion of volume regularity to be introduced in the next section.

3 Normalized modularity and volume regularity

With the normalized modularity matrix, the well-known Expander Mixing Lemma (for simple graphs see, e.g., [17]) is formulated for edge-weighted graphs in the following way (see [8]).

Lemma 1 *Provided* $\mathtt{Vol}(V) = 1$*, for all* $X, Y \subseteq V$*,*

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

where $\|\mathbf{M}_D\|$ *denotes the spectral norm of the normalized modularity matrix of* $G = (V, \mathbf{W})$*.*

Since the spectral gap of G is $1 - \|\mathbf{M}_D\|$, a large spectral gap indicates small discrepancy as a quasi-random property discussed in [15]. If there is a gap not at the ends of the spectrum, we want to partition the vertices into clusters so that a relation similar to the above property for the edge-densities between the cluster pairs would hold. For this purpose, we use a slightly modified version of the volume regularity's notion introduced in [2].

Definition 2 *Let $G = (V, \mathbf{W})$ be an edge-weighted graph with $\mathtt{Vol}(V) = 1$. The disjoint pair $A, B \subseteq V$ is α-volume regular if for all $X \subseteq A$, $Y \subseteq B$ we have*

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

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

In the ideal k-cluster case, let us consider the following generalized random simple graph model: given the partition $(V_1, \ldots, V_k)$ of V ($|V| = n$), vertices $i \in V_a$ and $j \in V_b$ are connected with probability p_{ab}, independently of each other, $1 \leq a, b \leq k$. We can think of the probability p_{ab} as the inter-cluster density of the pair (V_a, V_b). Since generalized random graphs can be viewed as edge-weighted graphs with a special block-structure burdened with random noise, based on [7], we are able to give the following spectral characterization of them. Fixing k, and tending with n to infinity in such a way that the cluster sizes grow at the same rate, there exists a positive number $\theta < 1$, independent of n, such that for every $0 < \tau < 1/2$ there are exactly $k-1$ eigenvalues of $\mathbf{M}_D$ greater than $\theta - n^{-\tau}$, while all the others are at most $n^{-\tau}$ in absolute value. Further, the k-variance of the vertex representatives constructed by the $k-1$ transformed structural eigenvectors is $\mathcal{O}(n^{-2\tau})$, and the cluster pairs are α-volume regular with any small α, almost surely. Note that generalized quasirandom graphs defined in [18] are deterministic counterparts of generalized random graphs with the same spectral properties.

Theorem 3 *Let $G = (V, \mathbf{W})$ be a connected edge-weighted graph on n vertices, with generalized degrees $d_1, \ldots, d_n$ and degree matrix $\mathbf{D}$. Assume that $\mathtt{Vol}(V) = 1$, and there are no dominant vertices, i.e., $d_i = \Theta(1/n)$, $i = 1, \ldots, n$, as $n \to \infty$. Let the eigenvalues of $\mathbf{M}_D$, enumerated in decreasing absolute values, be*

$$1 \geq |\mu_1| \geq \cdots \geq |\mu_{k-1}| > \varepsilon \geq |\mu_k| \geq \cdots \geq |\mu_n| = 0.$$

The partition $(V_1, \ldots, V_k)$ of V is defined so that it minimizes the weighted k-variance $S_k^2(\mathbf{X}^*)$ of the optimum vertex representatives – defined in (2) – obtained as row vectors of the $n \times (k-1)$ matrix $\mathbf{X}^*$ of column vectors $\mathbf{D}^{-1/2}\mathbf{u}_i$, where $\mathbf{u}_i$ is the unit-norm eigenvector corresponding to μ_i $(i = 1, \ldots, k-1)$. Assume that there is a constant $0 < K \leq \frac{1}{k}$ such that $|V_i| \geq Kn$, $i = 1, \ldots, k$. With the notation $s = \sqrt{S_k^2(\mathbf{X}^*)}$, the (V_i, V_j) pairs are $\mathcal{O}(\sqrt{2k}s + \varepsilon)$-volume regular $(i \neq j)$ and for the clusters V_i $(i = 1, \ldots, k)$ the following holds: for

all $X, Y \subset V_i$,

$$|w(X,Y) - \rho(V_i)\mathtt{Vol}(X)\mathtt{Vol}(Y)| = \mathcal{O}(\sqrt{2k}s + \varepsilon)\mathtt{Vol}(V_i),$$

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

Note that, in Section 2, we indexed the eigenvalues of $\mathbf{M}_D$ in non-increasing order and denoted them by λ's. The set of all λ_i's is the same as that of all μ_i's. Nonetheless, we need a different notation for the eigenvalues indexed in decreasing order of their absolute values. Recall that 1 cannot be an eigenvalue of $\mathbf{M}_D$ if G is connected. Consequently, $|\mu_1| = 1$ can be if and only if $\mu_1 = -1$, i.e., if G is bipartite. For example, if the conditions of the above theorem hold with $k = 2$ and $\mu_1 = -1$ ($|\mu_i| \leq \varepsilon$, $i \geq 2$), then our graph is a bipartite expander discussed in [1] in details.

For the proof we need the definition of the cut norm of a matrix (see e.g., [16]) and the relation between it and the spectral norm.

Definition 4 *The cut norm of the real matrix* $\mathbf{A}$ *with row-set* Row *and column-set* Col *is*

$$\|\mathbf{A}\|_\square = \max_{R \subset Row,\, C \subset Col} \left| \sum_{i \in R} \sum_{j \in C} a_{ij} \right|.$$

Lemma 5 *For every* $m \times n$ *real matrix* $\mathbf{A}$,

$$\|\mathbf{A}\|_\square \leq \sqrt{mn}\|\mathbf{A}\|,$$

where the right hand side contains the spectral norm, i.e. the largest singular value of $\mathbf{A}$.

PROOF.

$$\begin{aligned}\|\mathbf{A}\|_\square &= \max_{\mathbf{x}\in\{0,1\}^m, \mathbf{y}\in\{0,1\}^n} |\mathbf{x}^T\mathbf{A}\mathbf{y}| = \max_{\mathbf{x}\in\{0,1\}^m, \mathbf{y}\in\{0,1\}^n} \left| (\frac{\mathbf{x}}{\|\mathbf{x}\|})^T \mathbf{A} (\frac{\mathbf{y}}{\|\mathbf{y}\|}) \right| \cdot \|\mathbf{x}\| \cdot \|\mathbf{y}\| | \\ &\leq \sqrt{mn} \max_{\|\mathbf{x}\|=1,\, \|\mathbf{y}\|=1} |\mathbf{x}^T\mathbf{A}\mathbf{y}| = \sqrt{mn}\|\mathbf{A}\|,\end{aligned}$$

since for $\mathbf{x} \in \{0,1\}^m$, $\|\mathbf{x}\| \leq \sqrt{m}$, and for $\mathbf{y} \in \{0,1\}^n$, $\|\mathbf{y}\| \leq \sqrt{n}$. □

The definition of the cut norm and the result of the above lemma naturally extends to symmetric matrices with $m = n$, the spectral norm of which is the maximum of absolute values of their eigenvalues.

PROOF. (Theorem 3). Recall that the spectrum of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ differs from that of $\mathbf{M}_D$ only in the following: it contains the eigenvalue $\mu_0 = 1$ with

corresponding unit-norm eigenvector $\mathbf{u}_0 = \sqrt{\mathbf{d}}$ instead of the eigenvalue 0 of $\mathbf{M}_D$ with the same eigenvector. If G is connected, 1 is a simple eigenvalue. The optimum $(k-1)$-dimensional representatives of the vertices are row vectors of the matrix $\mathbf{X}^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_{k-1}^*)$, where $\mathbf{x}_i^* = \mathbf{D}^{-1/2}\mathbf{u}_i$ $(i = 1, \ldots, k-1)$. The representatives can as well be regarded as k-dimensional ones, as inserting the vector $\mathbf{x}_0^* = \mathbf{D}^{-1/2}\mathbf{u}_0 = \mathbf{1}$ will not change the k-variance $s^2 = S_k^2(\mathbf{X}^*)$. Assume that the minimum k-variance is attained on the k-partition $(V_1, \ldots, V_k)$ of the vertices. By an easy analysis of variance argument (see [5]) it follows that

$$s^2 = \sum_{i=0}^{k-1} \mathtt{dist}^2(\mathbf{u}_i, F), \tag{3}$$

where $F = \mathtt{Span}\,\{\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_k\}$ with the so-called normalized partition vectors $\mathbf{z}_1, \ldots, \mathbf{z}_k$ of coordinates $z_{ji} = \frac{1}{\sqrt{\mathtt{Vol}(V_i)}}$ 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 considerations proved in [5], we can find another orthonormal system $\mathbf{v}_0, \ldots, \mathbf{v}_{k-1} \in F$ such that

$$s^2 \le \sum_{i=0}^{k-1} \|\mathbf{u}_i - \mathbf{v}_i\|^2 \le 2s^2 \tag{4}$$

($\mathbf{v}_0 = \mathbf{u}_0$, since $\mathbf{u}_0 \in F$). We approximate the matrix $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} = \sum_{i=0}^{n-1} \mu_i \mathbf{u}_i \mathbf{u}_i^T$ by the rank k matrix $\sum_{i=0}^{k-1} \mu_i \mathbf{v}_i \mathbf{v}_i^T$ with the following accuracy (in spectral norm):

$$\left\| \sum_{i=0}^{n-1} \mu_i \mathbf{u}_i \mathbf{u}_i^T - \sum_{i=0}^{k-1} \mu_i \mathbf{v}_i \mathbf{v}_i^T \right\| \le \sum_{i=0}^{k-1} |\mu_i| \cdot \left\| \mathbf{u}_i \mathbf{u}_i^T - \mathbf{v}_i \mathbf{v}_i^T \right\| + \left\| \sum_{i=k}^{n-1} \mu_i \mathbf{u}_i \mathbf{u}_i^T \right\| \tag{5}$$

which can be estimated from above with $\sum_{i=0}^{k-1} \sin\alpha_i + \varepsilon \le \sum_{i=0}^{k-1} \|\mathbf{u}_i - \mathbf{v}_i\| + \varepsilon \le \sqrt{2k}s + \varepsilon$, where α_i is the angle between $\mathbf{u}_i$ and $\mathbf{v}_i$, and for it, $\sin\frac{\alpha_i}{2} = \frac{1}{2}\|\mathbf{u}_i - \mathbf{v}_i\|$ holds, $i = 0, \ldots, k-1$.

Based on these considerations and relation between the cut norm and the spectral norm (see Lemma 5), the densities to be estimated in the defining formula 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 $(V_1, \ldots, V_k)$, $i = 0, \ldots, k-1$. The matrix $\sum_{i=0}^{k-1} \lambda_i \mathbf{y}_i \mathbf{y}_i^T$ is therefore a symmetric block-matrix on $k \times k$ blocks belonging to the above partition of the vertices. Let $\hat{w}_{ab}$ denote its entries in the (a, b) block $(a, b = 1, \ldots, k)$. Using (5), the rank k approximation of the matrix $\mathbf{W}$ is performed with the following accuracy of the perturbation $\mathbf{E}$:

$$\|\mathbf{E}\| = \left\| \mathbf{W} - \mathbf{D}\left(\sum_{i=0}^{k-1} \mu_i \mathbf{y}_i \mathbf{y}_i^T\right)\mathbf{D} \right\| = \left\| \mathbf{D}^{1/2}\left(\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} - \sum_{i=0}^{k-1} \mu_i \mathbf{v}_i \mathbf{v}_i^T\right)\mathbf{D}^{1/2} \right\|.$$

Therefore, the entries of $\mathbf{W}$ – for $i \in V_a$, $j \in V_b$ – can be decomposed as $w_{ij} = d_i d_j \hat{w}_{ab} + \eta_{ij}$, where the cut norm of the $n \times n$ symmetric error matrix $\mathbf{E} = (\eta_{ij})$ restricted to $V_a \times V_b$ (otherwise it contains entries all zeros) and denoted by $\mathbf{E}_{ab}$, is estimated as follows:

$$\begin{aligned}
\|\mathbf{E}_{ab}\|_\square &\le n\|\mathbf{E}_{ab}\| \le n \cdot \|\mathbf{D}_a^{1/2}\| \cdot (\sqrt{2k}s + \varepsilon) \cdot \|\mathbf{D}_b^{1/2}\| \\
&\le n \cdot \sqrt{c_1 \frac{\mathtt{Vol}(V_a)}{|V_a|}} \cdot \sqrt{c_1 \frac{\mathtt{Vol}(V_b)}{|V_b|}} \cdot (\sqrt{2k}s + \varepsilon) \\
&= c_1 \cdot \sqrt{\frac{n}{|V_a|}} \cdot \sqrt{\frac{n}{|V_b|}} \cdot \sqrt{\mathtt{Vol}(V_a)}\sqrt{\mathtt{Vol}(V_b)}(\sqrt{2k}s + \varepsilon) \\
&\le c_1 \cdot \frac{1}{K}\sqrt{\mathtt{Vol}(V_a)}\sqrt{\mathtt{Vol}(V_b)}(\sqrt{2k}s + \varepsilon) \\
&= c\sqrt{\mathtt{Vol}(V_a)}\sqrt{\mathtt{Vol}(V_b)}(\sqrt{2k}s + \varepsilon).
\end{aligned}$$

Here the diagonal matrix $\mathbf{D}_a$ contains the diagonal part of $\mathbf{D}$ restricted to V_a, otherwise zeros, and the constant c does not depend on n. Consequently, for $a, b = 1, \ldots, k$ and $X \subseteq V_a$, $Y \subseteq V_b$:

$$\begin{aligned}
&|w(X,Y) - \rho(V_a,V_b)\mathtt{Vol}(X)\mathtt{Vol}(Y)| = \\
&\left|\sum_{i\in X}\sum_{j\in Y}(d_i d_j \hat{w}_{ab} + \eta_{ij}) - \frac{\mathtt{Vol}(X)\mathtt{Vol}(Y)}{\mathtt{Vol}(V_a)\mathtt{Vol}(V_b)}\sum_{i\in V_a}\sum_{j\in V_b}(d_i d_j \hat{w}_{ab} + \eta_{ij})\right| = \\
&\left|\sum_{i\in X}\sum_{j\in Y}\eta_{ij} - \frac{\mathtt{Vol}(X)\mathtt{Vol}(Y)}{\mathtt{Vol}(V_a)\mathtt{Vol}(V_b)}\sum_{i\in V_a}\sum_{j\in V_b}\eta_{ij}\right| \le 2c(\sqrt{2k}s + \varepsilon)\sqrt{\mathtt{Vol}(V_a)\mathtt{Vol}(V_b)},
\end{aligned}$$

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

Note that in the $k = 2$ special case, due to a theorem proved in [5], the 2-variance of the optimum 1-dimensional representatives can be directly estimated from above by the gap between the two largest absolute value eigenvalues of $\mathbf{M}_D$, and hence, the statement of Theorem 3 simplifies, see [8]. For a general k, we can make the following considerations.

Assume that the normalized modularity spectrum (with decreasing absolute values) of $G = (V, \mathbf{W})$ satisfies

$$1 \ge |\mu_1| \ge \cdots \ge |\mu_{k-1}| \ge \theta > \varepsilon \ge |\mu_k| \ge \cdots \ge |\mu_n| = 0.$$

Our purpose is to estimate s with the gap $\delta := \theta - \varepsilon$. We will use the notation of the proof of Theorem 3 and apply the results of [4] for the perturbation of

spectral subspaces of the symmetric matrices

$$\mathbf{A} = \sum_{i=0}^{n-1} \mu_i \mathbf{u}_i \mathbf{u}_i^T \quad \text{and} \quad \mathbf{B} = \sum_{i=0}^{k-1} \mu_i \mathbf{v}_i \mathbf{v}_i^T$$

in the following situation. The subsets $S_1 = \{\mu_k, \ldots, \mu_{n-1}\}$ and $S_2 = \{\mu_0, \ldots, \mu_{k-1}\}$ of the eigenvalues of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ are separated by an annulus, where $\mathtt{dist}(S_1, S_2) = \delta > 0$. Denote by $\mathbf{P}_A$ and $\mathbf{P}_B$ the projections onto the spectral subspaces of $\mathbf{A}$ and $\mathbf{B}$ spanned by the eigenvectors corresponding to the eigenvalues in S_1 and S_2, respectively:

$$\mathbf{P}_A(S_1) = \sum_{j=k}^{n-1} \mathbf{u}_j \mathbf{u}_j^T, \quad \mathbf{P}_B(S_2) = \sum_{i=0}^{k-1} \mathbf{v}_i \mathbf{v}_i^T.$$

Then Theorem VII.3.4 of [4] implies that

$$\|\mathbf{P}_A \mathbf{P}_B\|_F \leq \frac{1}{\delta} \|\mathbf{P}_A(\mathbf{A} - \mathbf{B})\mathbf{P}_B\|_F, \tag{6}$$

where $\|.\|_F$ denotes the Frobenius norm. On the left hand side, $\|\mathbf{P}_A \mathbf{P}_B\|_F = \sqrt{\sum_{i=0}^{k-1} \sin^2 \alpha_i}$, and in view of $\|\mathbf{u}_i - \mathbf{v}_i\| = 2 \sin \frac{\alpha_i}{2}$ and (4), this is between $\frac{\sqrt{3}}{2}s$ and s. On the right hand side,

$$\mathbf{P}_A\mathbf{A}\mathbf{P}_B - \mathbf{P}_A\mathbf{B}\mathbf{P}_B = (\mathbf{P}_A\mathbf{A})\mathbf{P}_B - \mathbf{P}_A(\mathbf{P}_B\mathbf{B}) = \sum_{i=0}^{k-1} \sum_{j=k}^{n-1} (\mu_j - \mu_i) \mathbf{u}_j^T (\mathbf{u}_i - \mathbf{v}_i) \mathbf{u}_j \mathbf{v}_i^T,$$

where the Frobenius norm of the rank 1 matrices $\mathbf{u}_j \mathbf{v}_i^T$ is 1, and the inner product $\mathbf{u}_j^T(\mathbf{u}_i - \mathbf{v}_i)$ is the smaller if the $\mathbf{u}_i$'s and the $\mathbf{v}_i$'s are the closer ($i = 1, \ldots, k-1$). Therefore, by the inequality (6), s is the smaller if δ is the larger and the $|\mu_j - \mu_i|$ differences for $i = 0, \ldots, k-1$; $j = k, \ldots, n-1$ are closer to δ. If $|\mu_k| = \varepsilon$ is small, then $|\mu_1|, \ldots, |\mu_{k-1}|$ should be close to each other ($\mu_0 = 1$ does not play an important role because of $\mathbf{u}_0 = \mathbf{v}_0$).

4 Testability of the normalized modularity spectrum and eigensubspaces

Authors of [12] defined the testability of simple graph parameters and proved equivalent notions of this testability. They also anticipated that their results remain valid if they consider weighted graph sequences (G_n) with edge-weights in the [0,1] interval and no dominant vertex-weights $\alpha_i(G_n) > 0$ ($i = 1, \ldots, n$), i.e., $\max_i \frac{\alpha_i(G_n)}{\alpha_{G_n}} \to 0$ as $n \to \infty$, where $\alpha_{G_n} = \sum_{i=1}^{n} \alpha_i(G_n)$. To this end, in [11], we slightly modified the definition of a testable graph parameter for weighted graphs in the following way.

Definition 6 *A weighted graph parameter f is testable if for every $\varepsilon > 0$ there is a positive integer $m < n$ such that if G_n satisfies $\max_i \frac{\alpha_i(G_n)}{\alpha_{G_n}} \leq \frac{1}{m}$, then*

$$\mathbb{P}(|f(G_n) - f(\eta(m, G_n))| > \varepsilon) \leq \varepsilon,$$

where $\eta(m, G_n)$ is a random simple graph on m vertices selected randomly from G_n in the following manner: m vertices of G_n are selected with replacement, with respective probabilities proportional to the vertex-weights; given the selected vertex-subset, the edges come into existence conditionally independently, with probabilities of the edge-weights.

By the above definition, a testable weighted graph parameter can be consistently estimated based on a fairly large sample. Based on the results of [12] for simple graphs, in [11], we established equivalent statements of this testability, from among which we will use the following.

Fact 7 *Let f be a testable weighted graph parameter. Then for every convergent weighted graph sequence (G_n), with no dominant vertex-weights, $f(G_n)$ is also convergent as $n \to \infty$.*

The notion of the convergence of a weighted graph sequence is defined in [12], where the authors also describe the limit object as a symmetric, measurable function $W : [0,1] \times [0,1] \to [0,1]$, called *graphon*. The so-called *cut distance* between the graphons W and U is $\delta_\square(W, U) = \inf_\nu \|W - U^\nu\|_\square$, where the cut norm of the graphon W is defined by

$$\|W\|_\square = \sup_{S,T \subset [0,1]} |\int_{S \times T} W(x, y)\, dx\, dy|,$$

and the above infimum is taken over all measure preserving bijections $\nu : [0,1] \to [0,1]$, while U^ν denotes the transformed U after performing the same measure preserving bijection ν on both sides of the unit square. Graphons are considered modulo measure preserving maps, and under graphon the whole equivalence class is understood. In this way, to a convergent weighted graph sequence (G_n), there is a unique limit graphon W such that $\delta_\square(G_n, W) \to 0$ as $n \to \infty$, where $\delta_\square(G_n, W)$ is defined as $\delta_\square(W_{G_n}, W)$ with the step-function graphon W_{G_n} assigned to G_n in the following way: the sides of the unit square are divided into intervals $I_1, \ldots, I_n$ of lengths $\alpha_1(G_n)/\alpha_{G_n}, \ldots, \alpha_n(G_n)/\alpha_{G_n}$, and over the rectangle $I_i \times I_j$ the stepfunction takes on the value $w_{ij}(G_n)$.

In [11], we proved the testability of some normalized and unnormalized balanced multiway cut densities such that we imposed balancing conditions on the cluster volumes. Under similar conditions, for fixed number of clusters k, the unnormalized and normalized multiway cuts and modularities are also testable, provided our edge-weighted graph has no dominant vertices. The proofs rely on statistical physics notions of [13], utilizing the fact that the

graph convergence implies the convergence of the ground state energy (minimum of the energy function over the set of k-partitions of vertices). In [22], the authors showed that the Newman-Girvan modularity is an energy function (Hamiltonian), and hence, testability of the maximum/minimum normalized modularities, under appropriate balancing conditions, can be shown analogously. Here we rather discuss the testability of spectra and k-variances, because in spectral clustering methods these provide us with polynomial time algorithms, though only approximate solutions are obtained as analyzed in Section 2.

In Theorem 6.6 of [13], the authors prove that the normalized spectrum of a convergent graph sequence also converges in the following sense. Let W be a graphon and (G_n) be a sequence of weighted graphs with uniformly bounded edge-weights tending to W. (For simplicity, we assume that $|V(G_n)| = n$). Let $|\lambda_{n,1}| \geq |\lambda_{n,2}| \geq \cdots \geq |\lambda_{n,n}|$ be the adjacency eigenvalues of G_n indexed by their decreasing absolute values, and let $\mu_{n,i} = \lambda_{n,i}/n$ $(i = 1, \ldots, n)$ be the normalized eigenvalues. Further, let T_W be the $L^2[0,1] \to L^2[0,1]$ integral operator corresponding to W:

$$T_W f(x) = \int_0^1 W(x,y) f(y)\, dy.$$

It is well-known that his operator is self-adjoint and compact, and hence, it has a discrete real spectrum, whose only possible point of accumulation is the 0. Let $\mu_i(W)$ denote the ith largest absolute value eigenvalue of T_W. Then for every $i \geq 1$, $\mu_{n,i} \to \mu_i(W)$ as $n \to \infty$. In fact, the authors prove a bit more (see Theorem 6.7 of [13]): if a sequence W_n of uniformly bounded graphons converges to a graphon W, then for every $i \geq 1$, $\mu_i(W_n) \to \mu_i(W)$ as $n \to \infty$. Note that the spectrum of W_G is the normalized spectrum of G, together with countably infinitely many 0's. Therefore, the convergence of the spectrum of (G_n) is the consequence of that of (W_{G_n}).

We will prove that in the absence of dominant vertices, the normalized modularity spectrum is testable. To this end, both the modularity matrix and the graphon are related to kernels of special integral operators, described herein. Let (ξ, ξ') be a pair of identically distributed real-valued random variables defined over the product space $\mathcal{X} \times \mathcal{X}$ having a symmetric joint distribution $\mathbb{W}$ with equal margins $\mathbb{P}$. Assume that the dependence between ξ and ξ' is regular, i.e., their joint distribution $\mathbb{W}$ is absolutely continuous with respect to the product measure $\mathbb{P} \times \mathbb{P}$, and let w denote its Radon–Nikodym derivative, see [23]. Let $H = L^2(\xi)$ and $H' = L^2(\xi')$ be the Hilbert spaces of random variables which are functions of ξ and ξ' and have zero expectation and finite variance with respect to $\mathbb{P}$. Observe that H and H' are isomorphic Hilbert spaces with the covariance as inner product; further, they are embedded as subspaces into the L^2-space defined similarly over the product space. (Here H and H' are also isomorphic in the sense that for any $\psi \in H$ there exists a

$\psi' \in H'$ and vice versa, such that ψ and ψ' are identically distributed.)

Consider the linear operator taking conditional expectation between H' and H with respect to the joint distribution. It is an integral operator and will be denoted by $P_{\mathbb{W}} : H' \to H$ as it is a projection restricted to H' and projects onto H. To $\psi' \in H'$ the operator $P_{\mathbb{W}}$ assigns $\psi \in H$ such that $\psi = \mathbb{E}_{\mathbb{W}}(\psi' \,|\, \xi)$, i.e.,

$$\psi(x) = \int_{\mathcal{Y}} w(x,y)\psi'(y)\,\mathbb{P}(dy), \quad x \in \mathcal{X}.$$

If

$$\int_{\mathcal{X}}\int_{\mathcal{X}} w^2(x,y)\mathbb{P}(dx)\mathbb{P}(dy) < \infty,$$

then $P_{\mathbb{W}}$ is a Hilbert–Schmidt operator, therefore compact and has spectral decomposition

$$P_{\mathbb{W}} = \sum_{i=1}^{\infty} \lambda_i \langle ., \psi_i' \rangle_{H'} \psi_i,$$

where for the eigenvalues $|\lambda_i| \leq 1$ holds and the eigenvalue–eigenfunction equation looks like

$$P_{\mathbb{W}}\psi_i' = \lambda_i \psi_i \quad (i = 1, 2, \dots),$$

where ψ_i and ψ_i' are identically distributed, whereas their joint distribution is $\mathbb{W}$. It is easy to see that $P_{\mathbb{W}}$ is self-adjoint and it takes the constantly 1 random variable of H' into the constantly 1 random variable of H; however, the $\psi_0 = 1, \psi_0' = 1$ pair is not regarded as a function pair with eigenvalue $\lambda_0 = 1$, since they have no zero expectation. More precisely, the kernel is reduced to $w(x,y) - 1$.

Theorem 8 *Let $G_n = (V_n, \mathbf{W}_n)$ be the general entry of a convergent sequence of connected edge-weighted graphs whose edge-weights are in [0,1] and the vertex-weights are the generalized degrees. Assume that there are no dominant vertices. Let W denote the limit graphon of the sequence (G_n), and let*

$$1 \geq |\mu_{n,1}| \geq |\mu_{n,2}| \geq \cdots \geq |\mu_{n,n}| = 0$$

be the normalized modularity spectrum of G_n (the eigenvalues are indexed by their decreasing absolute values). Further, let $\mu_i(P_{\mathbb{W}})$ is the ith largest absolute value eigenvalue of the integral operator $P_{\mathbb{W}} : L^2(\xi') \to L^2(\xi)$ taking conditional expectation with respect to the joint measure $\mathbb{W}$ embodied by the normalized limit graphon W, and ξ, ξ' are identically distributed random variables with the marginal distribution of their symmetric joint distribution $\mathbb{W}$. Then for every $i \geq 1$,

$$\mu_{n,i} \to \mu_i(P_{\mathbb{W}}) \quad \textit{as} \quad n \to \infty.$$

PROOF. In case of a finite $\mathcal{X}$ (vertex set) we have a weighted graph, and we will show that the operator taking conditional expectation with respect

to the joint distribution determined by the edge-weights corresponds to its normalized modularity matrix.

Indeed, let $\mathcal{X} = V$, $|V| = n$, and $G_n = (V, \mathbf{W})$ be an edge-weighted graph on the $n \times n$ weight matrix of the edges $\mathbf{W}$ with entries W_{ij}'s; now, they do not necessarily sum up to 1. (For the time being, n is kept fixed, so – for the sake of simplicity – we do not denote the dependence of $\mathbf{W}$ on n). Let the vertices be also weighted with special weights $\alpha_i(G_n) := \sum_{j=1}^n W_{ij}$, $i = 1, \dots, n$. Then the step-function graphon W_{G_n} is such that $W_{G_n}(x, y) = W_{ij}$ whenever $x \in I_i$ and $y \in I_j$, where the (not necessarily contiguous) intervals $I_1, \dots, I_n$ form a partition of [0,1] such that the length of I_i is $\alpha_i(G_n)/\alpha_{G_n}$ $(i = 1, \dots, n)$.

Let us transform $\mathbf{W}$ into a symmetric joint distribution $\mathbb{W}_n$ over $V \times V$. The entries $w_{ij} = W_{ij}/\alpha_{G_n}$ $(i, j = 1, \dots, n)$ embody this discrete joint distribution of random variables ξ and ξ' which are identically distributed with marginal distribution $d_1, \dots, d_n$, where $d_i = \alpha_i(G_n)/\alpha_{G_n}$ $(i = 1, \dots, n)$. With the previous notation $H = L^2(\xi)$, $H' = L^2(\xi')$, the operator $P_{\mathbb{W}_n} : H' \to H$ taking conditional expectation is an integral operator with now discrete kernel $K_{ij} = \frac{w_{ij}}{d_i d_j}$. The fact that ψ, ψ' is an eigenfunction pair of $P_{\mathbb{W}_n}$ with eigenvalue λ means that

$$\frac{1}{d_i} \sum_{j=1}^n w_{ij} \psi'(j) = \sum_{j=1}^n \frac{w_{ij}}{d_i d_j} \psi'(j) d_j = \lambda \psi(i), \tag{7}$$

where $\psi(j) = \psi'(j)$ denotes the value of ψ or ψ' taken on with probability d_i (recall that ψ and ψ' are identically distributed). The above equation is equivalent to

$$\sum_{j=1}^n \frac{w_{ij}}{\sqrt{d_i}\sqrt{d_j}} \sqrt{d_j} \psi(j) = \lambda \sqrt{d_i} \psi(i),$$

therefore the vector of coordinates $\sqrt{d_i}\psi(i)$ $(i = 1, \dots, n)$ is a unit-norm eigenvector of the normalized modularity matrix with eigenvalue λ (note that the normalized modularity spectrum does not depend on the scale of the edge-weights, it is the same whether we use W_{ij}'s or w_{ij}'s as edge-weights). Consequently, the eigenvalues of the conditional expectation operator are the same as the eigenvalues of the normalized modularity matrix, and the possible values taken on by the eigenfunctions of the conditional expectation operator are the same as the coordinates of the transformed eigenvectors of the normalized modularity matrix forming the column vectors of the matrix $\mathbf{X}^*$ of the optimal $(k-1)$-dimensional representatives, see Section 2 (a).

Let f be a stepwise constant function on [0,1], taking on value $\psi(i)$ on I_i. Then $\mathtt{Var}\, \psi = 1$ is equivalent to $\int_0^1 f^2(x)\, dx = 1$. Let K_{G_n} be the stepwise constant graphon defined as $K_{G_n}(x, y) = K_{ij}$ for $x \in I_i$ and $y \in I_j$. With this, the eigenvalue–eigenvector equation (7) looks like

$$\lambda f(x) = \int_0^1 K_{G_n}(x, y) f(y)\, dy.$$

The spectrum of K_{G_n} is the normalized modularity spectrum of G_n together with countably infinitely many 0's (it is of finite rank, and therefore, trivially compact), and because of the convergence of the weighted graph sequence G_n, in lack of dominant vertices, the sequence of graphons K_{G_n} also converges. Indeed, the $W_{G_n} \to W$ convergence in the cut metric means the convergence of the induced discrete distributions $\mathbb{W}_n$'s to the continuous $\mathbb{W}$. Since K_{G_n} and K are so-called copula transformations of those distributions, in lack of dominant vertices (this causes the convergence of the margins) they also converge, which in turn implies the $K_{G_n} \to K$ convergence in the cut metric.

Let K denote the limit graphon of K_{G_n} ($n \to \infty$). This will be the kernel of the integral operator taking conditional expectation with respect to the joint distribution $\mathbb{W}$. It is easy to see that this operator is also a Hilbert–Schmidt operator, and therefore, compact. With these considerations the remainder of the proof is analogous to the proof of Theorem 6.7 of [13], where the authors prove that if the sequence (W_{G_n}) of graphons converges to the limit graphon W, then both ends of the spectra of the integral operators, induced by W_{G_n}'s as kernels, converge to the ends of the spectrum of the integral operator induced by W as kernel. We apply this argument for the spectra of the integral operators induced by the kernels K_{G_n}'s and K. □

Note that in [19], kernel operators are also discussed, but not with our normalization.

Remark 9 *By Fact 7, provided there are no dominant vertices, Theorem 8 implies that for any fixed positive integer k, the $(k-1)$-tuple of the largest absolute value eigenvalues of the normalized modularity matrix is testable.*

Theorem 10 *Assume that there are constants $0 < \varepsilon < \theta \leq 1$ such that the normalized modularity spectrum (with decreasing absolute values) of any G_n satisfies*

$$1 \geq |\mu_{n,1}| \geq \cdots \geq |\mu_{n,k-1}| \geq \theta > \varepsilon \geq |\mu_{n,k}| \geq \cdots \geq |\mu_{n,n}| = 0.$$

With the notions of Theorem 8, and assuming that there are no dominant vertices of G_n's, the subspace spanned by the transformed eigenvectors $\mathbf{D}^{-1/2}\mathbf{u}_1$, ...,$\mathbf{D}^{-1/2}\mathbf{u}_{k-1}$ belonging to the $k-1$ largest absolute value eigenvalues of the normalized modularity matrix of G_n also converges to the corresponding $(k-1)$-dimensional subspace of $P_{\mathbb{W}}$. More precisely, if $\mathbf{P}_{n,k-1}$ denotes the projection onto the subspace spanned by the transformed eigenvectors belonging to $k-1$ largest absolute value eigenvalues of the normalized modularity matrix of G_n, and $\mathbf{P}_{k-1}$ denotes the projection onto the corresponding eigen-subspace of $P_{\mathbb{W}}$, then $\|\mathbf{P}_{n,k-1} - \mathbf{P}_{k-1}\| \to 0$ as $n \to \infty$ (in spectral norm).

PROOF. If we apply the convergence fact $\mu_{n,i} \to \mu_i(P_{\mathbb{W}})$ for indices $i = k-1$ and k, we get that there will be a gap of order $\theta - \varepsilon - o(1)$ between $|\mu_{k-1}(P_{\mathbb{W}})|$ and $|\mu_k(P_{\mathbb{W}})|$ too.

Let $P_{\mathbb{W},n}$ denote the n-rank approximation of $P_{\mathbb{W}}$ (keeping its n largest absolute value eigenvalues, together with the corresponding eigenfunctions) in spectral norm. The projection $\mathbf{P}_{k-1}$ $(k < n)$ operates on the eigen-subspace spanned by the eigenfunctions belonging to the $k-1$ largest absolute value eigenvalues of $P_{\mathbb{W},n}$ in the same way as on the corresponding $(k-1)$-dimensional subspace determined by $P_{\mathbb{W}}$. With these considerations, we apply the perturbation theory of eigen-subspaces with the following unitary invariant norm: the Schatten 4-norm of the Hilbert–Schmidt operator A is $\|A\|_4 = (\sum_{i=1}^{\infty} \lambda_i^4(A))^{1/4}$. Our argument with the finite $(k-1)$ rank projections is the following. Denoting by $P_{\mathbb{W}_n}$ the integral operator belonging to the normalized modularity matrix of G_n (with kernel K_{G_n} introduced in the proof of Theorem 8),

$$\begin{aligned}\|\mathbf{P}_{n,k-1} - \mathbf{P}_{k-1}\| = \|\mathbf{P}_{n,k-1}^{\perp}\mathbf{P}_{k-1}\| \leq \|\mathbf{P}_{n,k-1}^{\perp}\mathbf{P}_{k-1}\|_4 \\ \leq \frac{c}{\theta - \varepsilon - o(1)}\|P_{\mathbb{W}_n} - P_{\mathbb{W},n}\|_4\end{aligned}$$

with constant c that is at most $\pi/2$ (Theorem VII.3.2 of [4]). But

$$\|P_{\mathbb{W}_n} - P_{\mathbb{W},n}\|_4 \leq \|P_{\mathbb{W}_n} - P_{\mathbb{W}}\|_4 + \|P_{\mathbb{W}} - P_{\mathbb{W},n}\|_4,$$

where the last term tends to 0 as $n \to \infty$, since the tail of the spectrum (taking the fourth power of the eigenvalues) of a Hilbert–Schmidt operator converges. For the convergence of the first term we use Lemma 7.1 of [12], which states that the Schatten 4-norm of an integral operator can be estimated from above by four times the cut norm of the corresponding kernel. But the convergence in the cut distance of the corresponding kernels to zero follows from the considerations made in the proof of Theorem 8. This finishes the proof. □

Remark 11 *As the k-variance depends continuously on the above subspaces (see the expansion (3) of s^2 in the proof of Theorem 3), Theorem 10 implies the testability of the k-variance as well.*

5 Summary

The above results suggest that in the absence of dominant vertices, even the normalized modularity matrix of a smaller part of the underlying weighted graph, selected at random with an appropriate procedure, is able to reveal its cluster structure. Hence, the gain regarding the computational time of this spectral clustering algorithm is twofold: we only use a smaller part of the graph

and the spectral decomposition of its normalized modularity matrix runs in polynomial time in the reduced number of the vertices. Under the vertex- and cluster-balance conditions this method can give quite good approximations for the multiway cuts and helps us to find the number of clusters and identify the cluster structure. In addition, taking into account both the positive and negative, large absolute value eigenvalues together with eigenvectors, regular cuts can also be detected, as the investigated spectral characteristics give good estimates for the volume regularity's constant of the cluster pairs by Theorem 3. Such regular cuts are of importance in social or biological networks, e.g., if we want to find equally functioning synapses of the brain.

Acknowledgements

We thank the anonymous referee for his/her constructive comments.

References

[1] Alon, N., 1986 Eigenvalues and expanders, *Combinatorica* **6** (1986), 83-96.

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

[3] Azran, A. and Ghahramani, Z., Spectral methods for automatic multiscale data clustering, in *Proceedings of the CVPR Conference* (2006), pp. 190-197.

[4] Bhatia, R., *Matrix Analysis*, Springer, New York, 1997.

[5] Bolla, M. and Tusnády, G., Spectra and optimal partitions of weighted graphs, *Discret. Math.* **128** (1994), 1-20.

[6] Bolla, M. and Molnár-Sáska, G., Isoperimetric properties of weighted graphs related to Laplacian spectrum and canonical correlations, *Studia Sci. Math. Hun.* **39** (2002), 425-441.

[7] Bolla, M., Recognizing linear structure in noisy matrices, *Lin. Alg. Appl.* **402** (2005), 228-244.

[8] Bolla, M., Beyond the expanders, *International Journal of Combinatorics*, Paper 787596 (2011).

[9] Bolla, M., Penalized versions of the Newman–Girvan modularity and their relation to multiway cuts and k-means clustering, *Physical Review E* **84**, 016108 (2011).

[10] Bolla, M., Spectra and structure of weighted graphs, *Electronic Notes in Discret. Math.* **38** (2011), 149-154.

[11] Bolla, M., Kói, T., Krámli, A., Testability of minimum balanced multiway cut densities, *Discret. Appl. Math.* **160** (2012), 1019–1027.

[12] Borgs, C., Chayes, J. T., Lovász, L., T.-Sós, V., and Vesztergombi, K., Convergent Sequences of Dense Graphs I: Subgraph Frequencies, Metric Properties and Testing, *Advances in Math.* **219** (2008), 1801-1851.

[13] Borgs, C., Chayes, J. T., Lovász, L., T.-Sós, V., and Vesztergombi, K., Convergent Sequences of Dense Graphs II: Multiway Cuts and Statistical Physics, *Annals of Math.* **176**, 151–219.

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

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

[16] Frieze, A. and Kannan, R., Quick approximation to matrices and applications. *Combinatorica* **19** (1999), 175–220.

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

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

[19] Lovász L, L. and Szegedy, B., Finitely forcible graphons, *J. Comb. Theory B.* **101** (2011), 269–301.

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

[21] Newman, M. E. J., Finding community structure in networks using the eigenvectors of matrices, *Physical Review E* **74**, 036104 (2006).

[22] Reichardt, J. and Bornholdt, S., Partitioning and modularity of graphs with arbitrary degree distribution, *Physical Review E* **76**, 015102(R) (2007).

[23] Rényi, A., On measures of dependence, *Acta Math. Acad. Sci. Hungar.* **10** (1959), 441-451.