Discrete Applied Mathematics 203 (2016) 26–34

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Relating multiway discrepancy and singular values of nonnegative rectangular matrices

Marianna Bolla

Institute of Mathematics, Budapest University of Technology and Economics, Hungary

ARTICLE INFO

Article history:
Received 27 August 2014
Received in revised form 23 June 2015
Accepted 20 September 2015
Available online 23 October 2015

Keywords:
Discrepancy
Normalized matrix
Singular values
Spectral clusters

ABSTRACT

The minimum k-way discrepancy $\mathrm{md}_k(\mathbf{C})$ of a rectangular matrix $\mathbf{C}$ of nonnegative entries is the minimum of the maxima of the within- and between-cluster discrepancies that can be obtained by simultaneous k-clusterings (proper partitions) of its rows and columns. In Theorem 2, irrespective of the size of $\mathbf{C}$, we give the following estimate for the kth largest nontrivial singular value of the normalized matrix: $s_k \le 9\mathrm{md}_k(\mathbf{C})(k + 2 - 9k \ln \mathrm{md}_k(\mathbf{C}))$, provided $0 < \mathrm{md}_k(\mathbf{C}) < 1$ and $k < \mathrm{rank}(\mathbf{C})$. This statement is a certain converse of Theorem 7 of Bolla (2014), and the proof uses some lemmas and ideas of Butler (2006), where the $k = 1$ case is treated. The result naturally extends to the singular values of the normalized adjacency matrix of a weighted undirected or directed graph.

1. Introduction

In many applications, for example when microarrays are analyzed, our data are collected in the form of an $m \times n$ rectangular matrix $\mathbf{C} = (c_{ij})$ of nonnegative real entries. We assume that $\mathbf{C}$ is *non-decomposable* (see Definition A.3.28 of [6]), i.e., $\mathbf{C}\mathbf{C}^T$ (when $m \le n$) or $\mathbf{C}^T\mathbf{C}$ (when $m > n$) is *irreducible*. Consequently, the row-sums $d_{row,i} = \sum_{j=1}^n c_{ij}$ and column-sums $d_{col,j} = \sum_{i=1}^m c_{ij}$ of $\mathbf{C}$ are strictly positive, and the diagonal matrices $\mathbf{D}_{row} = \mathrm{diag}(d_{row,1}, \dots, d_{row,m})$ and $\mathbf{D}_{col} = \mathrm{diag}(d_{col,1}, \dots, d_{col,n})$ are regular. Without loss of generality, we also assume that $\sum_{i=1}^n \sum_{j=1}^m c_{ij} = 1$, since neither our main object, the *normalized matrix*

$$\mathbf{C}_D = \mathbf{D}_{row}^{-1/2}\mathbf{C}\mathbf{D}_{col}^{-1/2}, \tag{1}$$

nor the *multiway discrepancies* to be introduced are affected by the scaling of the entries of $\mathbf{C}$. It is known that the singular values of $\mathbf{C}_D$ are in the [0, 1] interval. The positive ones, enumerated in non-increasing order, are the real numbers

$$1 = s_0 > s_1 \ge \cdots \ge s_{r-1} > 0,$$

where $r = \mathrm{rank}(\mathbf{C}_D) = \mathrm{rank}(\mathbf{C})$. Provided $\mathbf{C}$ is non-decomposable, 1 is a single singular value; it will be called *trivial* and denoted by s_0, since it corresponds to the trivial singular vector pair, which are disregarded in the clustering problems. This is a well-known fact of *correspondence analysis*, for further explanation see [6,7] and Section 3.

In Theorem 2, we will estimate the kth nontrivial singular value s_k of $\mathbf{C}_D$ from above with a (near zero, increasing) function of the *minimum k-way discrepancy* of $\mathbf{C}$ defined herein.

E-mail address: marib@math.bme.hu.

http://dx.doi.org/10.1016/j.dam.2015.09.013
0166-218X/

Definition 1. The multiway discrepancy of the rectangular matrix $\mathbf{C}$ of nonnegative entries in the proper k-partition $R_1, \dots, R_k$ of its rows and $C_1, \dots, C_k$ of its columns is

$$\mathrm{md}(\mathbf{C}; R_1, \dots, R_k, C_1, \dots, C_k) = \max_{\substack{1 \le a,b \le k \\ X \subset R_a,\, Y \subset C_b}} \frac{|c(X,Y) - \rho(R_a, C_b)\mathrm{Vol}(X)\mathrm{Vol}(Y)|}{\sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}}, \tag{2}$$

where $c(X,Y) = \sum_{i \in X} \sum_{j \in Y} c_{ij}$ is the cut between $X \subset R_a$ and $Y \subset C_b$, $\mathrm{Vol}(X) = \sum_{i \in X} d_{row,i}$ is the volume of the row-subset X, $\mathrm{Vol}(Y) = \sum_{j \in Y} d_{col,j}$ is the volume of the column-subset Y, whereas $\rho(R_a, C_b) = \frac{c(R_a, C_b)}{\mathrm{Vol}(R_a)\mathrm{Vol}(C_b)}$ denotes the relative density between R_a and C_b. The minimum k-way discrepancy of $\mathbf{C}$ itself is

$$\mathrm{md}_k(\mathbf{C}) = \min_{\substack{R_1, \dots, R_k \\ C_1, \dots, C_k}} \mathrm{md}(\mathbf{C}; R_1, \dots, R_k, C_1, \dots, C_k).$$

In Section 3, we will extend this notion to an edge-weighted graph G and denote it by $\mathrm{md}_k(G)$. In that setup, $\mathbf{C}$ plays the role of the weighted adjacency matrix (symmetric in the undirected; quadratic, but usually not symmetric in the directed case), when the eigenvalues of the normalized adjacency matrix enter into the estimates, in their decreasing absolute values.

Note that $\mathrm{md}(\mathbf{C}; R_1, \dots, R_k, C_1, \dots, C_k)$ of (2) is the smallest α such that for every R_a, C_b pair and for every $X \subset R_a$, $Y \subset C_b$,

$$|c(X,Y) - \rho(R_a, C_b)\mathrm{Vol}(X)\mathrm{Vol}(Y)| \le \alpha \sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)} \tag{3}$$

holds. Consequently, in the k-partitions of the rows and columns, giving the minimum k-way discrepancy (say, α^*) of $\mathbf{C}$, every R_a, C_b pair is α^**-regular* in terms of the volumes, and α^* is the smallest possible discrepancy that can be attained with proper k-partitions. In the graph case, it resembles the notion of ϵ-regular pairs in the Szemerédi regularity lemma [18], albeit with given number of vertex-clusters, which are usually not equitable; further, with volumes, instead of cardinalities.

Though it is not always called discrepancy, this notion has intensively been used since the 1970s, e.g., in [9] and [18–20]. Thomason [19,20] introduced it in the context of what he called (p, α)-jumbled graphs and proved relations between this and similar notions, related to pseudo-random graphs. Expander graphs and the expander mixing lemma for simple regular graphs are also closely related to this notion, e.g., Alon, Spencer, Hoory, Linial, Widgerson [3,15]. Bollobás and Nikiforov [10] extended the notion of discrepancy to Hermitian matrices. Then they defined two types of discrepancy for graphs and showed that their estimate is valid to both, with due regard to a theorem of Thomason [20]. They also proved that for a large graph G, one type of these discrepancies closely approximates the discrepancy of its adjacency matrix (as a real Hermitian matrix). In Chung, Graham, Wilson [13], the authors used the term quasi-random for simple graphs that satisfy any of some equivalent properties, some of which closely related to discrepancy and eigenvalue separation.

Here we rather extend the notion of discrepancy used by Chung and Graham for simple graphs with given degree sequences. In [12], the authors proved that for simple graphs 'small' discrepancy $\mathrm{disc}(G)$ (with our notation, $\mathrm{md}_1(G)$) is caused by eigenvalue 'separation': the second largest singular value (which is also the second largest absolute value eigenvalue), s_1, of the normalized adjacency matrix is 'small', i.e., separated from the trivial singular value $s_0 = 1$, which is the edge of the spectrum. More exactly, they proved $\mathrm{disc}(G) \le s_1$, hence giving some kind of generalization of the *expander mixing lemma for irregular graphs*.

In the backward direction, Bollobás and Nikiforov [10] estimated the second largest singular value of an $n \times n$ Hermitian matrix $\mathbf{A}$ by $C\mathrm{disc}(\mathbf{A}) \log n$, and showed that this is best possible up to a multiplicative constant. Bilu and Linial [4] proved the converse of the expander mixing lemma for simple regular graphs, but their key lemma, producing this statement, goes beyond regular graphs, see Section 3.1 for details. In Alon et al. [2], the authors relaxed the notion of eigenvalue separation to essential eigenvalue separation (by introducing a parameter for it, and requiring the separation only for the eigenvalues of a relatively large part of the graph). Then they proved relations between the constants of this kind of eigenvalue separation and discrepancy.

For a general rectangular matrix $\mathbf{C}$ of nonnegative entries, Butler [11] proved the following forward and backward statement in the $k = 1$ case:

$$\mathrm{disc}(\mathbf{C}) \le s_1 \le 150\mathrm{disc}(\mathbf{C})(1 - 8 \ln \mathrm{disc}(\mathbf{C})), \tag{4}$$

where $\mathrm{disc}(\mathbf{C})$ is our $\mathrm{md}_1(\mathbf{C})$ and, with our notation, s_1 is the largest nontrivial singular value of $\mathbf{C}_D$ (he denoted it with σ_2). Since $s_1 < 1$, the upper estimate makes sense for very small discrepancy, in particular, for $\mathrm{disc}(\mathbf{C}) \le 8.868 \times 10^{-5}$. The lower estimate further generalizes the expander mixing lemma to rectangular matrices.

So far, the overall discrepancy has been considered in the sense, that $\mathrm{disc}(\mathbf{C})$ or $\mathrm{disc}(G)$ measures the largest possible deviation between the actual and expected connectedness of arbitrary (sometimes disjoint) subsets X, Y, where under expected the hypothesis of independence is understood (which corresponds to the rank 1 approximation of the underlying matrix). Our purpose is, in the multicluster scenario, to find similar relations between the minimum k-way discrepancy and the SVD of the normalized matrix, for given k. In the *backward direction*, in Section 2, we will prove the following.

Theorem 2. *For every non-decomposable real matrix* $\mathbf{C}$ *of nonnegative entries and integer* $1 \le k < \mathrm{rank}(\mathbf{C})$,

$$s_k \le 9\mathrm{md}_k(\mathbf{C})(k + 2 - 9k \ln \mathrm{md}_k(\mathbf{C})),$$

provided $0 < \mathrm{md}_k(\mathbf{C}) < 1$, *where* s_k *is the kth largest nontrivial singular value of the normalized matrix* $\mathbf{C}_D$ *introduced in* (1).

Note that $\mathrm{md}_k(\mathbf{C}) = 0$ if $\mathbf{C}$ has a block structure with k row- and column-blocks or if $\mathrm{rank}(\mathbf{C}) = 1$ (case of an independent table, see Section 3.3), in which cases $s_k = 0$ also holds. Likewise, $\mathrm{md}_k(\mathbf{C}) < 1$ is not a peculiar requirement, since in view of $s_k < 1$, the upper bound of the theorem has relevance only for $\mathrm{md}_k(\mathbf{C})$ much smaller than 1; for example, for $\mathrm{md}_1(\mathbf{C}) \le 1.866 \times 10^{-3}$, $\mathrm{md}_2(\mathbf{C}) \le 8.459 \times 10^{-4}$, $\mathrm{md}_3(\mathbf{C}) \le 5.329 \times 10^{-4}$, etc.

Earlier, in the *forward direction*, in [7] we estimated an alternative version of $\mathrm{md}_k(\mathbf{C})$ by means of s_k and another quantity from above, see Section 3.3 for details. Roughly speaking, the two directions together imply that if s_k is 'small' and 'much smaller' than s_{k-1}, then one may expect a simultaneous k-clustering of the rows and columns of $\mathbf{C}$ with small k-way discrepancy.

2. Proof of Theorem 2

Before proving the theorem, we encounter some lemmas of other authors that will be used, possibly with some modifications.

Lemma 3 of Bollobás and Nikiforov [10] is the key to prove their main result. This lemma states that to every $0 < \varepsilon < 1$ and vector $\mathbf{x} \in \mathbf{C}^n$, $\|\mathbf{x}\| = 1$, there exists a vector $\mathbf{y} \in \mathbf{C}^n$ such that its coordinates take no more than $\lceil \frac{8\pi}{\varepsilon} \rceil \lceil \frac{4}{\varepsilon} \log \frac{2n}{\varepsilon} \rceil$ distinct values and $\|\mathbf{x} - \mathbf{y}\| \le \varepsilon$. We will rather use the construction of the following lemma, which is indeed a consequence of Lemma 3 of [10].

Lemma 3 (*Lemma 3 of Butler [11]*)**.** *To any vector* $\mathbf{x} \in \mathbf{C}^n$, $\|\mathbf{x}\| = 1$ *and diagonal matrix* $\mathbf{D}$ *of positive real diagonal entries, one can construct a step-vector* $\mathbf{y} \in \mathbb{C}^n$ *such that* $\|\mathbf{x} - \mathbf{D}\mathbf{y}\| \le \frac{1}{3}$, $\|\mathbf{D}\mathbf{y}\| \le 1$, *and the nonzero entries of* $\mathbf{y}$ *are of the form* $\left(\frac{4}{5}\right)^j e^{\frac{\ell}{29}2\pi i}$ *with appropriate integers j (taking on* $O(\log n)$ *distinct values) and* ℓ *(*$0 \le \ell \le 28$*).*

Note that starting with an $\mathbf{x}$ of real coordinates, we do not need all the 29 values of ℓ, only two of them will show up, as it follows from a better understanding of the construction of [11]. In fact, by the idea of [10], j's come from dividing the coordinates of $\mathbf{D}^{-1}\mathbf{x}/\|\mathbf{D}^{-1}\mathbf{x}\|$ in decreasing absolute values into groups, where the cut-points are powers of $\frac{4}{5}$. With the notation $\mathbf{x} = (x_s)_{s=1}^n$, if x_s is in the jth group, then the corresponding coordinate of the approximating complex vector $\mathbf{y} = (y_s)_{s=1}^n$ is as follows. If $x_s = 0$, then $y_s = 0$, otherwise $y_s = \left(\frac{4}{5}\right)^j e^{\left(\lfloor \frac{29\theta}{2\pi} \rfloor/29\right)2\pi i}$, where θ is the argument of x_s, $0 \le \theta < 2\pi$, and therefore, $\ell = \lfloor \frac{29\theta}{2\pi} \rfloor$ is an integer between 0 and 28. However, when the coordinates of $\mathbf{x}$ are real numbers, then only the values 0 and 14 of ℓ can occur, since θ can take only one of the values 0 or π, depending on whether x_s is positive or negative. We will intensively use this observation in our proof.

Lemma 4 (*Lemma 4 of Butler [11]*)**.** *Let* $\mathbf{M}$ *be a matrix with largest singular value* σ *and corresponding unit-norm singular vector pair* $\mathbf{v}, \mathbf{u}$. *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are vectors such that* $\|\mathbf{x}\| \le 1$, $\|\mathbf{y}\| \le 1$, $\|\mathbf{v} - \mathbf{x}\| \le \frac{1}{3}$, $\|\mathbf{u} - \mathbf{y}\| \le \frac{1}{3}$, *then* $\sigma \le \frac{9}{2}\langle \mathbf{x}, \mathbf{M}\mathbf{y} \rangle$.

Lemma 5 (*Theorem 3 of Thompson [21]*)**.** *Let the* $n \times n$ *matrix have singular values* $\alpha_1 \ge \cdots \ge \alpha_n$ *and* $1 \le k \le n$ *be a fixed integer. Then an* $n \times n$ *matrix* $\mathbf{X}$ *exists with* $\mathrm{rank}(\mathbf{X}) \le k$ *such that* $\mathbf{B} = \mathbf{A} + \mathbf{X}$ *has singular values* $\beta_1 \ge \cdots \ge \beta_n$ *if and only if*

$$\alpha_{i+k} \le \beta_i \le \alpha_{i-k}, \quad i = 1, \ldots, n$$

with the understanding that $\alpha_j = +\infty$ *if* $j \le 0$ *and* $\alpha_j = 0$ *if* $j \ge n$.

Proof. The proof of Theorem 2 is as follows. Assume that $\alpha^* = \mathrm{md}_k(\mathbf{C}) \in (0, 1)$ and it is attained with the proper k-partition $R_1, \ldots, R_k$ of the rows and $C_1, \ldots, C_k$ of the columns of $\mathbf{C}$; i.e., for every R_a, C_b pair and $X \subset R_a$, $Y \subset C_b$ we have

$$|c(X, Y) - \rho(R_a, C_b)\mathrm{Vol}(X)\mathrm{Vol}(Y)| \le \alpha^* \sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}. \tag{5}$$

Our purpose is to put Inequality (5) in a matrix form by using indicator vectors and introducing the $m \times n$ auxiliary matrix

$$\mathbf{F} = \mathbf{C} - \mathbf{D}_{row}\mathbf{R}\mathbf{D}_{col}, \tag{6}$$

where $\mathbf{R} = (\rho(R_a, C_b))$ is the $m \times n$ block-matrix of $k \times k$ blocks with entries equal to $\rho(R_a, C_b)$ over the block $R_a \times C_b$. With the indicator vectors $\mathbf{1}_X$ and $\mathbf{1}_Y$ of $X \subset R_a$ and $Y \subset C_b$, Inequality (5) has the following equivalent form:

$$|\langle \mathbf{1}_X, \mathbf{F}\mathbf{1}_Y \rangle| \le \alpha^* \sqrt{\langle \mathbf{1}_X, \mathbf{C}\mathbf{1}_n \rangle \langle \mathbf{1}_m, \mathbf{C}\mathbf{1}_Y \rangle}, \tag{7}$$

where $\mathbf{1}_n$ denotes the all 1's vector of size n. At the same time, Eq. (6) yields

$$\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2} = \mathbf{D}_{row}^{-1/2}\mathbf{C}\mathbf{D}_{col}^{-1/2} - \mathbf{D}_{row}^{1/2}\mathbf{R}\mathbf{D}_{col}^{1/2} = \mathbf{C}_D - \mathbf{D}_{row}^{1/2}\mathbf{R}\mathbf{D}_{col}^{1/2}.$$

Since the rank of the matrix $\mathbf{D}_{row}^{1/2}\mathbf{R}\mathbf{D}_{col}^{1/2}$ is at most k, by the upper estimate of Lemma 5 (with the role-cast $\mathbf{A}=\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2}$, $\mathbf{B}=\mathbf{C}_D$, $\mathbf{X}=\mathbf{D}_{row}^{1/2}\mathbf{R}\mathbf{D}_{col}^{1/2}$, and $i=k+1$)[1] we obtain the following upper estimate for s_k, that is the $(k+1)$th largest (including the trivial 1) singular value of $\mathbf{C}_D$:

$$s_k \le s_{max}(\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2}) = \|\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2}\|,$$

where $\|.\|$ denotes the spectral norm.

Let $\mathbf{v}\in\mathbb{R}^m$ be the left and $\mathbf{u}\in\mathbb{R}^n$ be the right unit-norm singular vector corresponding to the maximal singular value of $\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2}$, i.e.,

$$|\langle\mathbf{v},(\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2})\mathbf{u}\rangle| = \|\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2}\|.$$

In view of Lemma 3, there are step-vectors $\mathbf{x}\in\mathbb{C}^m$ and $\mathbf{y}\in\mathbb{C}^n$ such that $\|\mathbf{v}-\mathbf{D}_{row}^{1/2}\mathbf{x}\|\le\frac{1}{3}$ and $\|\mathbf{u}-\mathbf{D}_{col}^{1/2}\mathbf{y}\|\le\frac{1}{3}$; further, $\|\mathbf{D}_{row}^{1/2}\mathbf{x}\|\le 1$ and $\|\mathbf{D}_{col}^{1/2}\mathbf{y}\|\le 1$. Then Lemma 4 yields

$$\|\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2}\| \le \frac{9}{2}\left|\langle(\mathbf{D}_{row}^{1/2}\mathbf{x}),(\mathbf{D}_{row}^{-1/2}\mathbf{F}\mathbf{D}_{col}^{-1/2})(\mathbf{D}_{col}^{1/2}\mathbf{y})\rangle\right| = \frac{9}{2}|\langle\mathbf{x},\mathbf{F}\mathbf{y}\rangle|.$$

Now we will use the construction of the proof of Lemma 3 in the special case when the vectors $\mathbf{v}=(v_s)_{s=1}^m$ and $\mathbf{u}=(u_s)_{s=1}^n$, to be approximated, have real coordinates. Therefore, only the following three types of coordinates of the approximating complex vectors $\mathbf{x}=(x_s)_{s=1}^m$ and $\mathbf{y}=(y_s)_{s=1}^n$ will appear. If $v_s=0$, then $x_s=0$; if $v_s>0$, then $x_s=(\frac{4}{5})^j$ with some integer j; if $v_s<0$, then $x_s=(\frac{4}{5})^j e^{\frac{28}{29}\pi i}$ with some integer j. Likewise, if $u_s=0$, then $y_s=0$; if $u_s>0$, then $y_s=(\frac{4}{5})^\ell$ with some integer ℓ; if $u_s<0$, then $y_s=(\frac{4}{5})^\ell e^{\frac{28}{29}\pi i}$ with some integer ℓ. With these observations, the step-vectors $\mathbf{x}$ and $\mathbf{y}$ can be written as the following finite sums with respect to the integers j and ℓ:

$$\mathbf{x}=\sum_j\left(\frac{4}{5}\right)^j\mathbf{x}^{(j)},\quad \mathbf{x}^{(j)}=\sum_{a=1}^k(\mathbf{1}_{\mathcal{X}_{ja1}}+e^{\frac{28}{29}\pi i}\mathbf{1}_{\mathcal{X}_{ja2}}),\ \text{where}$$
$$\mathcal{X}_{ja1}=\{s:\ v_s>0,\ s\in R_a\}\quad\text{and}\quad\mathcal{X}_{ja2}=\{s:\ v_s<0,\ s\in R_a\};$$

likewise,

$$\mathbf{y}=\sum_\ell\left(\frac{4}{5}\right)^\ell\mathbf{y}^{(\ell)},\quad \mathbf{y}^{(\ell)}=\sum_{b=1}^k(\mathbf{1}_{\mathcal{Y}_{\ell b1}}+e^{\frac{28}{29}\pi i}\mathbf{1}_{\mathcal{Y}_{\ell b2}}),\ \text{where}$$
$$\mathcal{Y}_{\ell b1}=\{s:\ u_s>0,\ s\in C_b\}\quad\text{and}\quad\mathcal{Y}_{\ell b2}=\{s:\ u_s<0,\ s\in C_b\}.$$

It is important that the $2k$ indicator vectors appearing in the decomposition of any $\mathbf{x}^{(j)}$ or $\mathbf{y}^{(\ell)}$ are disjointly supported, and so, all the coordinates of these vectors are of absolute value 1. These considerations give rise to the following estimation.

$$\begin{aligned}
|\langle\mathbf{x}^{(j)},\mathbf{F}\mathbf{y}^{(\ell)}\rangle| &\le \sum_{a=1}^k\sum_{p=1}^2\sum_{b=1}^k\sum_{q=1}^2\left|\langle\mathbf{1}_{\mathcal{X}_{jap}},\mathbf{F}\mathbf{1}_{\mathcal{Y}_{\ell bq}}\rangle\right| \\
&\overset{(7)}{\le} \sum_{a=1}^k\sum_{p=1}^2\sum_{b=1}^k\sum_{q=1}^2\alpha^*\sqrt{\langle\mathbf{1}_{\mathcal{X}_{jap}},\mathbf{C}\mathbf{1}_n\rangle\langle\mathbf{1}_m,\mathbf{C}\mathbf{1}_{\mathcal{Y}_{\ell bq}}\rangle} \\
&\le \alpha^*2k\sqrt{\sum_{a=1}^k\sum_{p=1}^2\sum_{b=1}^k\sum_{q=1}^2\langle\mathbf{1}_{\mathcal{X}_{jap}},\mathbf{C}\mathbf{1}_n\rangle\langle\mathbf{1}_m,\mathbf{C}\mathbf{1}_{\mathcal{Y}_{\ell bq}}\rangle} \\
&= 2k\alpha^*\sqrt{\left\langle\sum_{a=1}^k\sum_{p=1}^2\mathbf{1}_{\mathcal{X}_{jap}},\mathbf{C}\mathbf{1}_n\right\rangle\left\langle\mathbf{1}_m,\mathbf{C}\sum_{b=1}^k\sum_{q=1}^2\mathbf{1}_{\mathcal{Y}_{\ell bq}}\right\rangle} \\
&= 2k\alpha^*\sqrt{\langle|\mathbf{x}^{(j)}|,\mathbf{C}\mathbf{1}_n\rangle\langle\mathbf{1}_m,\mathbf{C}|\mathbf{y}^{(\ell)}|\rangle},
\end{aligned}\tag{8}$$

where in the first inequality we used the triangle inequality and $|e^{\frac{28}{29}\pi i}|=1$, in the second one we used (7), while in the third one, the Cauchy–Schwarz inequality with $4k^2$ terms.

[1] Actually, Lemma 5 is about square matrices, but in the possession of a rectangular one, we can supplement it with zero rows or columns to make it quadratic; further, the nonzero singular values of the so obtained square matrix are the same as those of the rectangular one, supplemented with additional zero singular values that will not alter the shifted interlacing facts.

In the last step we exploited that the indicator vectors composing $\mathbf{x}^{(j)}$ and $\mathbf{y}^{(\ell)}$ are disjointly supported. We also introduced the notation $|\mathbf{z}| = (|z_s|)_{s=1}^n$ for the real vector, the coordinates of which are the absolute values of the corresponding coordinates of the (possibly complex) vector $\mathbf{z}$. (Note that the so introduced $|\mathbf{z}|$ is a vector, unlike $\|\mathbf{z}\| = (\sum_{s=1}^n |z_s|^2)^{1/2}$.) In the same spirit, let $|\mathbf{M}|$ denote the matrix whose entries are the absolute values of the corresponding entries of $\mathbf{M}$ (we will use this only for real matrices). With this formalism, this is the right moment to prove the following inequalities that will be used soon to finish the proof:

$$\sum_\ell |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle| \le 2\langle |\mathbf{x}^{(j)}|, \mathbf{C}\mathbf{1}_n\rangle, \qquad \sum_j |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle| \le 2\langle \mathbf{1}_m, \mathbf{C}|\mathbf{y}^{(\ell)}|\rangle. \tag{9}$$

Since the two inequalities are of the same flavor, it suffices to prove only the first one. Note that it is here, where we use the exact definition of $\mathbf{F}$ as follows.

$$\begin{aligned}\sum_\ell |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle| &\le \left\langle |\mathbf{x}^{(j)}|, |\mathbf{F}| \sum_\ell |\mathbf{y}^{(\ell)}| \right\rangle \\ &\le \langle |\mathbf{x}^{(j)}|, (\mathbf{C} + \mathbf{D}_{row}\mathbf{R}\mathbf{D}_{col})\mathbf{1}_n\rangle| = 2\langle |\mathbf{x}^{(j)}|, \mathbf{C}\mathbf{1}_n\rangle\end{aligned}$$

because $|\mathbf{y}^{(\ell)}|$ is a 0–1 vector and $\mathbf{C} + \mathbf{D}_{row}\mathbf{R}\mathbf{D}_{col}$ is a (real) matrix of nonnegative entries. We also used that the ith coordinate of the vector $(\mathbf{C} + \mathbf{D}_{row}\mathbf{R}\mathbf{D}_{col})\mathbf{1}_n$ for $i \in R_a$ is

$$d_{row,i}\left(1 + \sum_{b=1}^k \rho(R_a, C_b)\mathrm{Vol}(C_b)\right) = 2d_{row,i}$$

(here we utilized that the sum of the entries of $\mathbf{C}$ is 1), and therefore,

$$(\mathbf{C} + \mathbf{D}_{row}\mathbf{R}\mathbf{D}_{col})\mathbf{1}_n = 2\mathbf{C}\mathbf{1}_n.$$

Finally, we will finish the proof with similar considerations as in [11]. Let us further estimate

$$\langle \mathbf{x}, \mathbf{F}\mathbf{y}\rangle = \sum_j \sum_\ell \left\langle \left(\frac{4}{5}\right)^j \mathbf{x}^{(j)}, \mathbf{F}\left(\frac{4}{5}\right)^\ell \mathbf{y}^{(\ell)} \right\rangle.$$

Put $\gamma := \log_{4/5}\alpha^*$; in view of $\alpha^* < 1$, $\gamma > 0$ holds. Then we divide the above summation into three parts as follows.

$$\begin{aligned}|\langle \mathbf{x}, \mathbf{F}\mathbf{y}\rangle| &\le \sum_j \sum_\ell \left(\frac{4}{5}\right)^{j+\ell} |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle| \\ &= \underbrace{\sum_{|j-\ell|\le\gamma} \left(\frac{4}{5}\right)^{j+\ell} |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle|}_{(a)} + \underbrace{\sum_{j-\ell>\gamma} \left(\frac{4}{5}\right)^{j+\ell} |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle|}_{(b)} + \underbrace{\sum_{j-\ell<-\gamma} \left(\frac{4}{5}\right)^{j+\ell} |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle|}_{(c)}.\end{aligned}$$

The three terms are estimated separately. Term (a) can be bounded from above as follows:

$$\begin{aligned}\sum_{|j-\ell|\le\gamma} \left(\frac{4}{5}\right)^{j+\ell} |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle| &\overset{(8)}{\le} 2k\alpha^* \sum_{|j-\ell|\le\gamma} \sqrt{\left(\frac{4}{5}\right)^{2j} \langle |\mathbf{x}^{(j)}|, \mathbf{C}\mathbf{1}_n\rangle \left(\frac{4}{5}\right)^{2\ell} \langle \mathbf{1}_m, \mathbf{C}|\mathbf{y}^{(\ell)}|\rangle} \\ &\overset{(*)}{\le} k\alpha^* \sum_{|j-\ell|\le\gamma} \left[\left(\frac{4}{5}\right)^{2j} \langle |\mathbf{x}^{(j)}|, \mathbf{C}\mathbf{1}_n\rangle + \left(\frac{4}{5}\right)^{2\ell} \langle \mathbf{1}_m, \mathbf{C}|\mathbf{y}^{(\ell)}|\rangle\right] \\ &\overset{(**)}{\le} k\alpha^*(2\gamma+1)\left[\sum_j \left(\frac{4}{5}\right)^{2j} \langle |\mathbf{x}^{(j)}|, \mathbf{C}\mathbf{1}_n\rangle + \sum_\ell \left(\frac{4}{5}\right)^{2\ell} \langle \mathbf{1}_m, \mathbf{C}|\mathbf{y}^{(\ell)}|\rangle\right], \\ &\overset{(***)}{\le} 2k\alpha^*(2\gamma+1),\end{aligned}$$

where in the first inequality, the estimate of (8), and in (*), the geometric–arithmetic mean inequality were used; (**) comes from the fact that in the second line, the first term depends merely on j, while the second one merely on ℓ, and so, for fixed j or ℓ, any term can show up at most $2\gamma + 1$ times; (***) is due to the easy observation that

$$\sum_j \left(\frac{4}{5}\right)^{2j} \langle |\mathbf{x}^{(j)}|, \mathbf{C}\mathbf{1}_n\rangle = \|\mathbf{D}_{row}^{1/2}\mathbf{x}\|^2 \le 1, \qquad \sum_\ell \left(\frac{4}{5}\right)^{2\ell} \langle \mathbf{1}_m, \mathbf{C}|\mathbf{y}^{(\ell)}|\rangle = \|\mathbf{D}_{col}^{1/2}\mathbf{y}\|^2 \le 1. \tag{10}$$

Terms (b) and (c) are of similar appearance (the role of j and ℓ is symmetric in them), therefore, we will estimate only (b). Here $j - \ell > \gamma$, yielding $j + \ell > 2\ell + \gamma$. Therefore,

$$\sum_{j-\ell>\gamma} \left(\frac{4}{5}\right)^{j+\ell} |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle| \le \sum_{\ell} \left(\frac{4}{5}\right)^{2\ell+\gamma} \sum_{j} |\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)}\rangle| \overset{(9)}{\le} \sum_{\ell} \left(\frac{4}{5}\right)^{2\ell+\gamma} 2\langle \mathbf{1}_m, \mathbf{C}|\mathbf{y}^{(\ell)}|\rangle$$
$$= 2\left(\frac{4}{5}\right)^{\gamma} \sum_{\ell} \left(\frac{4}{5}\right)^{2\ell} \langle \mathbf{1}_m, \mathbf{C}|\mathbf{y}^{(\ell)}|\rangle \overset{(10)}{\le} 2\left(\frac{4}{5}\right)^{\gamma}$$

where, in the second and third inequalities, (9) and (10) were used. Consequently, (c) can also be estimated from above with $2\left(\frac{4}{5}\right)^{\gamma}$.

Collecting the so obtained estimates together, we get

$$s_k \le \frac{9}{2}|\langle \mathbf{x}, \mathbf{F}\mathbf{y}\rangle| \le \frac{9}{2}\left[2k\alpha^*(2\gamma+1) + 4\left(\frac{4}{5}\right)^{\gamma}\right] = 9\alpha^*\left[2k\frac{\ln\alpha^*}{\ln\frac{4}{5}} + k + 2\right]$$
$$\le 9\alpha^*[2k(-4.5)\ln\alpha^* + k + 2] = 9\alpha^*(k + 2 - 9k\ln\alpha^*),$$

that was to be proved.

Note that for $k = 1$, our upper bound is tighter than that of (4), see Theorem 2 of [11].

3. Special cases and conclusions

3.1. Undirected graphs

The notion of the multiway discrepancy naturally extends to edge-weighted graphs. A weighted undirected graph $G = (V, \mathbf{W})$ on n vertices is uniquely characterized by its $n \times n$ *weighted adjacency matrix* $\mathbf{W}$, which is symmetric of nonnegative entries and zero diagonal. $\mathbf{D} = \mathrm{diag}(d_1, \dots, d_n)$ is the diagonal *degree-matrix* ($d_i = \sum_{j=1}^{n} w_{ij}$), $\mathrm{Vol}(U) = \sum_{i\in U} d_i$ is the volume of $U \subset V$, and for simplicity we assume that $\sum_{i=1}^{n} d_i = 1$; it does not hurt the generality, because neither the *normalized weighted adjacency matrix* $\mathbf{W}_D = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, nor the multiway discrepancies are affected by the scaling of $\mathbf{W}$.

Definition 6. The multiway discrepancy of the undirected, edge-weighted graph $G = (V, \mathbf{W})$ in the proper k-partition $V_1, \dots, V_k$ of its vertices is

$$\mathrm{md}(G; V_1, \dots, V_k) = \max_{\substack{1\le a\le b\le k\\ X\subset V_a,\, Y\subset V_b}} \frac{|w(X, Y) - \rho(V_a, V_b)\mathrm{Vol}(X)\mathrm{Vol}(Y)|}{\sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}}.$$

The minimum k-way discrepancy of G is

$$\mathrm{md}_k(G) = \min_{V_1,\dots,V_k} \mathrm{disc}(G; V_1, \dots, V_k).$$

A result, similar to that of Theorem 2 can now be proved in terms of the eigenvalues of $\mathbf{W}_D$, the absolute values of which are the singular values:

$$1 = \mu_0 \ge |\mu_1| \ge \dots \ge |\mu_{n-1}|.$$

Theorem 7. *Let $G = (V, \mathbf{W})$ be an edge-weighted, undirected graph with the non-decomposable weighted adjacency matrix $\mathbf{W}$. Then*

$$|\mu_k| \le 9\mathrm{md}_k(G)(k + 2 - 9k\ln\mathrm{md}_k(G)), \tag{11}$$

where μ_k is the kth largest absolute value eigenvalue (excluding the trivial 1) of the normalized weighted adjacency matrix $\mathbf{W}_D$ ($k = 1, \dots, n-1$).

Proof. The proof follows the same considerations as the proof of Theorem 2 with the difference that here we use symmetric matrices. In particular, $\mathbf{R} = (\rho(V_a, V_b))$ is an $n \times n$ symmetric block-matrix of $k \times k$ blocks corresponding to the partition $V_1, \dots, V_k$ of the vertices for which $\alpha^* = \mathrm{disc}(G) = \mathrm{disc}(G; V_1, \dots, V_k)$; consequently, the matrix $\mathbf{F} = \mathbf{W} - \mathbf{DRD}$ is also symmetric. Therefore, in accord with (7) and Definition 6: for every V_a, V_b pair and $X \subset V_a$, $Y \subset V_b$ ($1 \le a \le b \le k$) we have

$$|\langle \mathbf{1}_X, \mathbf{F}\mathbf{1}_Y\rangle| \le \alpha^*\sqrt{\langle \mathbf{1}_X, \mathbf{W}\mathbf{1}_n\rangle\langle \mathbf{1}_n, \mathbf{W}\mathbf{1}_Y\rangle}. \tag{12}$$

The left and right singular vectors ($\mathbf{v}, \mathbf{u} \in \mathbb{R}^n$) corresponding to the maximal singular value of the real symmetric matrix $\mathbf{D}^{-1/2}\mathbf{F}\mathbf{D}^{-1/2}$ satisfy $\mathbf{u} = \pm\mathbf{v}$ (the sign is the same as the sign of the eigenvalue of the maximal absolute value). If $\mathbf{u} = \mathbf{v}$, then $\mathcal{Y}_{\ell bq} = \mathcal{X}_{\ell bq}$ for every $\ell, b = 1, \dots, k$, and $q = 1, 2$. If $\mathbf{u} = -\mathbf{v}$, then $\mathcal{Y}_{\ell b1} = \mathcal{X}_{\ell b2}$ and $\mathcal{Y}_{\ell b2} = \mathcal{X}_{\ell b1}$ for every ℓ and $b = 1, \dots, k$. Consequently, in the estimates of (8), when we use the absolute values of the coordinates of the vectors $\mathbf{x}^{(j)}$ and $\mathbf{y}^{(\ell)}$, constructed in Section 2, the inequalities remain valid. Namely,

$$|\langle \mathbf{x}^{(j)}, \mathbf{F}\mathbf{y}^{(\ell)} \rangle| \le \sum_{a=1}^{k} \sum_{p=1}^{2} \sum_{b=1}^{k} \sum_{q=1}^{2} \left| \langle \mathbf{1}_{\mathcal{X}_{jap}}, \mathbf{F}\mathbf{1}_{\mathcal{X}_{\ell bq}} \rangle \right| .$$

Here the summation is for every $1 \le a, b \le k$ pair. However, if $a \le b$, then by (12) we get

$$\left| \langle \mathbf{1}_{\mathcal{X}_{jap}}, \mathbf{F}\mathbf{1}_{\mathcal{X}_{\ell bq}} \rangle \right| \le \alpha^* \sqrt{\langle \mathbf{1}_{\mathcal{X}_{jap}}, \mathbf{W}\mathbf{1}_n \rangle \langle \mathbf{1}_n, \mathbf{W}\mathbf{1}_{\mathcal{X}_{\ell bq}} \rangle};$$

whereas, if $a > b$, then by the symmetry of $\mathbf{F}$:

$$\begin{aligned} \left| \langle \mathbf{1}_{\mathcal{X}_{jap}}, \mathbf{F}\mathbf{1}_{\mathcal{X}_{\ell bq}} \rangle \right| &= \left| \langle \mathbf{1}_{\mathcal{X}_{\ell bq}}, \mathbf{F}\mathbf{1}_{\mathcal{X}_{jap}} \rangle \right| \le \alpha^* \sqrt{\langle \mathbf{1}_{\mathcal{X}_{lbq}}, \mathbf{W}\mathbf{1}_n \rangle \langle \mathbf{1}_n, \mathbf{W}\mathbf{1}_{\mathcal{X}_{jap}} \rangle} \\ &= \sqrt{\langle \mathbf{1}_{\mathcal{X}_{jap}}, \mathbf{W}\mathbf{1}_n \rangle \langle \mathbf{1}_n, \mathbf{W}\mathbf{1}_{\mathcal{X}_{\ell bq}} \rangle}. \end{aligned}$$

Therefore, for $a \ne b$, the same term appears twice, and all the subsequent estimates remain valid by substituting $\mathbf{W}$ for $\mathbf{C}$ and $\mathbf{D}$ for both $\mathbf{D}_{row}$ and $\mathbf{D}_{col}$. This completes the proof.

Recall that Bilu and Linial [4] proved the following converse of the expander mixing lemma for simple d-regular graphs on n vertices. Assume that for any disjoint vertex-subsets S, T: $|e(S,T) - \frac{|S||T|d}{n}| \le \alpha\sqrt{|S||T|}$. Then all but the largest adjacency eigenvalue of G are bounded (in absolute value) by $O(\alpha(1+\log\frac{d}{\alpha}))$. Note that for a d-regular graph the adjacency eigenvalues are d times larger than the normalized adjacency ones, and the deviation between $e(S,T)$ and the one what is expected in a random d-regular graph, is also proportional to our (1-way) discrepancy in terms of the volumes (note that $\mathrm{Vol}(S)$ is also proportional to $|S|$). Though they use disjoint subsets S, T, their upper estimate for the absolute value of the second largest (in absolute value) eigenvalue with the (1-way) discrepancy α is $C\alpha(1 - A\log\alpha)$ with some absolute constants A, C. Hence, the upper estimate of (4) or that of (11) in the $k = 1$ case is reminiscent of this.

In the forward direction, for an arbitrary k (between 1 and rank $\mathbf{W}$), in Theorem 3 of [8] we proved that under some balancing conditions for the degrees and the cluster sizes (when $n \to \infty$) and denoting by $V_1, \dots, V_k$ the clusters obtained by spectral clustering (see the forthcoming explanation), the (V_a, V_b) pairs are $O(\sqrt{2k}S_k + |\mu_k|)$-volume regular ($a \ne b$) and similar statement holds for the subgraphs induced by V_a's. In this direction, we merely assume that $\mathbf{W}$ is irreducible, so 1 is an eigenvalue of $\mathbf{W}_D$. However, the singular value 1 of $\mathbf{W}_D$ can have multiplicity greater than one. For example, if G is bipartite, then $\mathbf{W}$ is irreducible, but decomposable, hence -1 is also an eigenvalue of $\mathbf{W}_D$, and so, the singular value 1 has multiplicity two. It will be important in the forthcoming examples.

In fact, inspired by [2], in [8] we used a bit different notation and concept of α-volume regular pairs, namely, for every $X \subseteq V_a$, $Y \subseteq V_b$ we required

$$|w(X,Y) - \rho(V_a, V_b)\mathrm{Vol}(X)\mathrm{Vol}(Y)| \le \alpha\sqrt{\mathrm{Vol}(V_a)\mathrm{Vol}(V_b)}.$$

In the above formula, the right hand side contains the squareroots of the volumes of the clusters, unlike (3), which contains the squareroots of the volumes of X and Y. However, in the spirit of the Szemerédi regularity lemma [18], if we require (3) to hold only for X, Y's satisfying $\mathrm{Vol}(X) \ge \varepsilon\mathrm{Vol}(V_i)$, $\mathrm{Vol}(Y) \ge \varepsilon\mathrm{Vol}(V_j)$ with some fixed ε, then the so modified k-way discrepancy is $O(\sqrt{2k}S_k + |\mu_k|)$, and so does $\mathrm{md}_k(G)$. Here the partition $V_1, \dots, V_k$ is defined so that it minimizes the weighted k-variance S_k^2 of the vertex representatives $\mathbf{r}_1, \dots, \mathbf{r}_n \in \mathbb{R}^{k-1}$ obtained as the row vectors of the $n \times (k-1)$ matrix of column vectors $\mathbf{D}^{-1/2}\mathbf{u}_i$, where $\mathbf{u}_i$ is the unit-norm eigenvector corresponding to $\mu_i (i = 1, \dots, k-1)$. The k-variance of the representatives is defined as

$$S_k^2 = \min_{(V_1,\dots,V_k)} \sum_{a=1}^{k} \sum_{j \in V_a} d_j \|\mathbf{r}_j - \mathbf{c}_a\|^2, \tag{13}$$

where $\mathbf{c}_a = \frac{1}{\mathrm{Vol}(V_a)} \sum_{j\in V_a} d_j \mathbf{r}_j$ is the weighted center of cluster V_a. It is the weighted k-means algorithm that gives this minimum, and the point is that the optimum S_k is just the minimum distance between the eigensubspace corresponding to $\mu_0, \dots \mu_{k-1}$ and the one of the suitably transformed step-vectors over the k-partitions of V. In [8] we also discussed that, in view of subspace perturbation theorems, the larger the gap between $|\mu_{k-1}|$ and $|\mu_k|$, the smaller S_k is. So the message is, that here the eigenvectors corresponding to the largest absolute value eigenvalues have to be used, unlike usual spectral clustering methods which automatically use the bottom eigenvalues of the Laplacian or normalized Laplacian matrix (latter one is just $\mathbf{I} - \mathbf{W}_D$). The clusters or cluster-pairs of small discrepancy behave like expanders or bipartite expanders. In another context, they resemble the generalized random or quasirandom graphs of Lovász, Sós, Simonovits [16,17].

In some special cases $S_k = 0$, and so, $\mathrm{md}_k(G) \le B|\mu_k| = Bs_k$ follows from the result of [8]. In particular, $S_k = 0$ whenever the vectors $\mathbf{D}^{-1/2}\mathbf{u}_1, \dots, \mathbf{D}^{-1/2}\mathbf{u}_{k-1}$ are step-vectors over the same proper k-partition of the vertices. Some examples:

- If $k = 2$ and G is *bipartite*, then $\mu_1 = -1$, $s_1 = 1$, and S_2^2, i.e., the 2-variance of the coordinates of the transformed eigenvector corresponding to μ_1 can be small if $|\mu_2|$ is separated from $|\mu_1| = 1$ (see also the bipartite expanders of [1]).
- Let $k = 2$ and G be *bipartite, biregular* on the independent vertex-subsets V_1, V_2. That is, all the edge-weights within V_1 or V_2 are zeros, and the 0–1 weights between vertices of V_1 and V_2 are such that $d_i = k_1$ if $i \in V_1$ and $d_i = k_2$ if $i \in V_2$ with the understanding that $|V_1|k_1 = |V_2|k_2$ (both are the total number of edges in G). It is easy to see that the unit-norm eigenvector corresponding to the eigenvalue $\mu_1 = -1$ is $\mathbf{u}_1 = \mathbf{D}^{1/2}\mathbf{1}_{V_1} - \mathbf{D}^{1/2}\mathbf{1}_{V_2}$, and $\mathbf{D}^{-1/2}\mathbf{u}_1 = \mathbf{1}_{V_1} - \mathbf{1}_{V_2}$. Therefore, the representatives of vertices of V_1 are all 1's, and those of V_2 are -1's, so $S_2 = 0$. Consequently, $\mathrm{md}_2(G) \le B|\mu_2|$, with some absolute constant B. Up to the constant, this was another proof of Lemma 3.2 of Evra et al. [14]. They call their result expander mixing lemma for bipartite graphs, and use cardinalities instead of volumes, but in this special case, these cardinalities are proportional to the volumes, both within V_1 and V_2.
- Let $G(n, \mathbf{P})$ be a generalized random graph with n vertices over the symmetric $k \times k$ pattern matrix $\mathbf{P} = (p_{ab})$; i.e., there is a proper k-partition, $V_1, \dots, V_k$, of its vertices such that $|V_a| = n_a$ $(a = 1, \dots, k)$, $\sum_{a=1}^k n_a = n$, and for any $1 \le a \le b \le k$, vertices $i \in V_a$ and $j \in V_b$ are connected independently, with the same probability p_{ab}. This is the k-cluster generalization of the classical Erdős–Rényi random graph, see also [16] for their generalized quasirandom counterparts. In [5] we characterized the adjacency and normalized Laplacian spectra of such graphs, that extends to their normalized adjacency spectra as follows: both $|\mu_k| = s_k$ and S_k tend to zero almost surely when $n \to \infty$, under some balancing conditions for the cluster sizes ($\frac{n_a}{n} \ge c$ with some constant c, for $a = 1, \dots, k$). By the forward statement of [8], it also holds for the k-way discrepancy in the clustering $V_1, \dots, V_k$.

Summarizing, in the $k = 1$ case: when the second singular value $|\mu_1| = s_1$ is 'small' (much smaller than $s_0 = 1$), then the overall discrepancy is 'small'. However, for $k > 1$, a 'small' s_k is necessary, but not sufficient for a 'small' k-way discrepancy. In addition, S_k should be 'small' too. With subspace perturbation theorems, it is 'small' if s_k is 'much smaller' than s_{k-1}. Hence, a gap in the normalized spectrum may be an indication for the choice of k.

3.2. Directed graphs

A directed weighted graph $G = (V, \mathbf{W})$ is described by its quadratic, but usually not symmetric weighted adjacency matrix $\mathbf{W} = (w_{ij})$ of zero diagonal, where w_{ij} is the nonnegative weight of the $i \to j$ edge ($i \ne j$). The row-sums $d_{out,i} = \sum_{j=1}^n w_{ij}$ and column-sums $d_{in,j} = \sum_{i=1}^n w_{ij}$ of $\mathbf{W}$ are the *out- and in-degrees*, while $\mathbf{D}_{out} = \mathrm{diag}(d_{out,1}, \dots, d_{out,n})$ and $\mathbf{D}_{in} = \mathrm{diag}(d_{in,1}, \dots, d_{in,n})$ are the diagonal *out- and in-degree matrices*, respectively. Now Definition 1 can be formulated as follows.

Definition 8. The multiway discrepancy of the directed, edge-weighted graph $G = (V, \mathbf{W})$ in the in-clustering $V_{in,1}, \dots, V_{in,k}$ and out-clustering $V_{out,1}, \dots, V_{out,k}$ of its vertices is

$$\mathrm{md}(G; V_{in,1}, \dots, V_{in,k}, V_{out,1}, \dots, V_{out,k}) = \max_{\substack{1\le a,b\le k\\ X\subset V_{out,a},\, Y\subset V_{in,b}}} \frac{|w(X,Y) - \rho(V_{out,a}, V_{in,b})\mathrm{Vol}_{out}(X)\mathrm{Vol}_{in}(Y)|}{\sqrt{\mathrm{Vol}_{out}(X)\mathrm{Vol}_{in}(Y)}},$$

where $w(X, Y)$ is the sum of the weights of the $X \to Y$ edges, whereas $\mathrm{Vol}_{out}(X) = \sum_{i\in X} d_{out,i}$ and $\mathrm{Vol}_{in}(Y) = \sum_{j\in Y} d_{in,j}$ are the out- and in-volumes, respectively. The minimum k-way discrepancy of the directed weighted graph $G = (V, \mathbf{W})$ is

$$\mathrm{md}_k(G) = \min_{\substack{V_{in,1},\dots,V_{in,k}\\ V_{out,1},\dots,V_{out,k}}} \mathrm{md}(G; V_{in,1}, \dots, V_{in,k}, V_{out,1}, \dots, V_{out,k}).$$

Butler [11] treats the $k = 1$ case, and for a general k, Theorem 2 implies the following.

Proposition 9. *Let $G = (V, \mathbf{W})$ be a directed edge-weighted graph with non-decomposable weighted adjacency matrix $\mathbf{W}$. Then*

$$s_k \le 9\mathrm{md}_k(G)(k + 2 - 9k \ln \mathrm{md}_k(G)),$$

where s_k is the kth largest nontrivial singular value of the normalized weighted adjacency matrix $\mathbf{W}_D = \mathbf{D}_{out}^{-1/2}\mathbf{W}\mathbf{D}_{in}^{-1/2}$.

We applied the SVD based algorithm to find migration patterns in the set of 75 countries, and found 3 underlying immigration and emigration trait clusters. Since the algorithm is the same as for rectangular matrices, we will describe it in Section 3.3.

3.3. Back to rectangular matrices

In multivariate statistics, sometimes our data are collected in an $m \times n$ matrix $\mathbf{C}$, where the entries are frequency counts corresponding to the joint distribution of two categorized random variables (taking on m and n distinct values, respectively). Such a $\mathbf{C}$ is called *contingency table* in statistical language, and the data are popularly said to be cross-tabulated. The χ^2 statistic, which measures the deviation from *independence*, is $N\sum_{i=1}^{r-1} s_i^2$ with the notation of Section 1, where N is the (usually 'large') sample size. However, the second factor can be 'small' if s_1 is 'small', and this corresponds to the existence of a good rank 1 approximation of $\mathbf{C}$ (when the two underlying random variables are 'nearly' independent). This fact is also supported by the $\mathrm{disc}(\mathbf{C}) = \mathrm{md}_1(\mathbf{C}) \le s_1$ relation. Otherwise, one may ask, whether there exists a 'good' rank k approximation for some integer $1 < k < r = \mathrm{rank}(\mathbf{C})$, which problem is treated in the *correspondence analysis* by the first k dyads of the SVD of $\mathbf{C}_D$. However, there it is not made exact how s_k is estimated by $\mathrm{md}_k(\mathbf{C})$. Our Theorem 2 says that if the minimum k-way discrepancy is very 'small', i.e., the sub-tables $R_a \times C_b$ behave like independent tables in the optimal k-partitions of the rows and columns, then s_k is 'small' too.

In the forward direction, in [7], we proved the following. Given the $m \times n$ contingency table $\mathbf{C}$, consider the spectral clusters $R_1, \dots, R_k$ of its rows and $C_1, \dots, C_k$ of its columns, obtained by applying the weighted k-means algorithm for the $(k-1)$-dimensional row- and column representatives, defined as the row vectors of the matrices of column vectors $(\mathbf{D}_{row}^{-1/2}\mathbf{v}_1, \dots, \mathbf{D}_{row}^{-1/2}\mathbf{v}_{k-1})$ and $(\mathbf{D}_{col}^{-1/2}\mathbf{u}_1, \dots, \mathbf{D}_{col}^{-1/2}\mathbf{u}_{k-1})$, respectively, where $\mathbf{v}_i, \mathbf{u}_i$ is the unit norm singular vector pair corresponding to $s_i (i = 1, \dots, k-1)$. In fact, these partitions minimize the weighted k-variances $S_{k,row}^2$ and $S_{k,col}^2$ of these row- and column-representatives introduced in (13). Then, under some balancing conditions for the margins and for the cluster sizes, we proved that $\mathrm{md}_k(\mathbf{C}) \le B(\sqrt{2k}(S_{k,row} + S_{k,col}) + s_k)$, with some constant B, which depends only on the constants of the balancing conditions, and does not depend on m and n. This is the base of our algorithm, with fixed k.

We remark that the *correspondence analysis* uses the above $(k-1)$-dimensional row- and column-representatives for simultaneously plotting the row- and column-categories in $\mathbb{R}^{k-1}$ ($k = 2, 3$ or 4 in most applications), and hence, the practitioner can draw conclusions from their mutual positions. For example, in microarray analysis we can plot the genes and conditions together, and the biclusters obtained by k-clustering the row- and column-representatives give clusters of the genes and the conditions such that, every gene-cluster and condition-cluster pair behaves like a random weighted bipartite graph in the sense, that genes and conditions of the same cluster nearly independently influence each other, which fact may have importance for practitioners.

In the possession of networks or microarrays, practitioners want to find a fairly small k, such that there is a k-cluster structure behind the matrix or the graph in the sense that the subgraphs and bipartite subgraphs have 'small' discrepancy. It depends on the matrix or the graph that how small discrepancy can be attained and with what k. The above theory tells that we have to inspect the normalized spectra together with spectral subspaces, since the leading ones carry a lot of information about the smallest attainable multiway discrepancy.

Acknowledgments

The author wishes to thank Gergely Kiss and Zoltán Miklós Sándor for discussions on the topic. Parts of the research were done under the auspices of the Budapest Semesters of Mathematics program, in the framework of an undergraduate research course on spectral clustering with the participation of US students James Drain, Cristina Mata, Matthew Willian, and in particular, Calvin Cheng whose computer processing of real-world data helped in formulating the main theorem (in 2014). Formerly (in 2012), in the framework of the same course, Max Del Giudice and Joan Wang applied the biclustering algorithm to directed graphs for finding migration patterns, based on sociological data.

References

[1] N. Alon, 1986 Eigenvalues and expanders, Combinatorica 6 (1986) 83–96.
[2] N. Alon, A. Coja-Oghlan, H. Han, M. Kang, V. Rödl, M. Schacht, Quasi-randomness and algorithmic regularity for graphs with general degree distributions, Siam J. Comput. 39 (2010) 2336–2362.
[3] N. Alon, J.H. Spencer, The Probabilistic Method, Wiley, 2000.
[4] Y. Bilu, N. Linial, Lifts, discrepancy and nearly optimal spectral gap, Combinatorica 26 (2006) 495–519.
[5] M. Bolla, Noisy random graphs and their Laplacians, Phys. Rev. E 84 (1) (2011) 016108–4230.
[6] M. Bolla, Spectral Clustering and Biclustering. Learning Large Graphs and Contingency Tables, Wiley, 2013.
[7] M. Bolla, SVD, discrepancy, and regular structure of contingency tables, Discrete Appl. Math. 176 (2014) 3–11.
[8] M. Bolla, Modularity spectra, eigen-subspaces and structure of weighted graphs, European J. Combin. 35 (2014) 105–116.
[9] B. Bollobás, Random Graphs, second ed., Cambridge Univ. Press, Cambridge, 2001.
[10] B. Bollobás, V. Nikiforov, Hermitian matrices and graphs: singular values and discrepancy, Discret. Math. 285 (2004) 17–32.
[11] S. Butler, Using discrepancy to control singular values for nonnegative matrices, Linear Algebra Appl. 419 (2006) 486–493.
[12] F. Chung, R. Graham, Quasi-random graphs with given degree sequences, Random Structures Algorithms 12 (2008) 1–19.
[13] F. Chung, R. Graham, R.K. Wilson, Quasi-random graphs, Combinatorica 9 (1989) 345–362.
[14] S. Evra, K. Golubev, A. Lubotzky, Mixing properties and the chromatic number of Ramanujan complexes, 2014. arXiv:1407.7700 [math.CO].
[15] S. Hoory, N. Linial, A. Widgerson, Expander graphs and their applications, Bull. Amer. Math. Soc. (N. S.) 43 (2006) 439–561.
[16] L. Lovász, V. T -Sós, Generalized quasirandom graphs, J. Combin. Theory Ser. B 98 (2008) 146–163.
[17] M. Simonovits, V. T -Sós, Szemerédi's partition and quasi-randomness, Random Structures Algorithms 2 (1991) 1–10.
[18] E. Szemerédi, Regular partitions of graphs, in: J-C Bermond, J-C Fournier, M. Las Vergnas, D. Sotteau (Eds.), Colloque Inter. CNRS. No. 260, Problémes Combinatoires et Théorie Graphes, 1976, pp. 399–401.
[19] A. Thomason, Pseudo-random graphs, in: M. Karónski (Ed.), Proc. Random Graphs, Poznán (1985), in: Annals of Discrete Math., vol. 33, 1987, pp. 307–331.
[20] A. Thomason, Dense expanders and pseudo-random bipartite graphs, Discrete Math. 75 (1989) 381–386.
[21] R.C. Thompson, The behavior of eigenvalues and singular values under perturbations of restricted rank, Linear Algebra Appl. 13 (1976) 69–78.