SVD, discrepancy, and regular structure of contingency tables

Marianna Bolla *

Institute of Mathematics, Budapest University of Technology and Economics and Center for Telecommunication and Informatics, Debrecen University

Abstract

Factors, obtained by correspondence analysis, are used to find biclustering of a contingency table such that the row–column cluster pairs are regular, i.e., they have small discrepancy. In our main theorem, the constant of the so-called volume-regularity is related to the SVD of the normalized contingency table. This result is applicable to two-way cuts when both the rows and columns are divided into the same number of clusters, thus extending partly the result of Butler for estimating the discrepancy of a contingency table by the largest non-trivial singular value of the normalized table (one-cluster, rectangular case), and partly the result of Bolla for estimating the constant of volume-regularity by the structural eigenvalues and the distances of the corresponding eigen-subspaces of the normalized modularity matrix of an edge-weighted graph (several clusters, symmetric case).

Key words: Normalized contingency table, Normalized two-way cuts, Biclustering, Discrepancy, Directed graphs

1 Introduction

A typical problem of contemporary cluster analysis is to find relatively small number of groups of objects, belonging to rows and columns of a contingency table which exhibit homogeneous behavior with respect to each other and do not differ significantly in size. To make inferences on the separation that can be achieved for a given number of clusters, minimum normalized two-way cuts and discrepancies of the cluster pairs are investigated and related to the SVD of the normalized contingency table.

* Research supported by the TÁMOP-4.2.2.C-11/1/KONV-2012-0001 project.
Email address: marib@math.bme.hu (Marianna Bolla).

Contingency tables are rectangular arrays with nonnegative, real entries, e.g., the keyword–document matrix or microarray gene expression data. In the former one, the matrix entries are associations between documents and words, whereas in the latter one, they are expression levels of genes under different conditions. We also look for a bipartition of the genes and conditions such that genes in the same cluster equally influence conditions of the same cluster. To find so-called biclustering, i.e., simultaneous clustering of the rows and columns, a great variety of algorithms are used.

The first algorithm of this flavor is due to Hartigan [17], who used two-way analysis of variance techniques to find constant valued submatrices within the rectangular array. In [24], applications to microarrays is presented, where biclusters identify subsets of genes sharing similar expression patterns across subsets of conditions, but the authors do not use spectral methods. We will rather concentrate on methods that use the SVD of the original or normalized contingency table.

The Latent Semantic Indexing offers an SVD-based algorithm, which can be generalized in many different ways. For example, in [15], the authors find scoring systems simultaneously for the keywords and documents with respect to the most important topics or factors, and use the singular vector pairs corresponding to the k outstanding singular values of the table.

If a scoring system is endowed with the marginal measures, the problem can be formulated in terms of the correspondence analysis, based on the SVD of the normalized table, see [4,13,16]. We will show, how in possession of the correspondence factor pairs a biclustering can be performed that finds simultaneous clustering of the rows and columns of the table such that certain regularity requirements are met.

A survey of biclustering algorithms in data mining, especially in biological data analysis is given in [8,24]. To find biclustering of a binary table via the k-means algorithm is also discussed in [10], where the author embeds the contingency table into a bipartite graph and uses normalized cut objectives and SVD to obtain the convenient biclustering. To find the SVD for large rectangular matrices, randomized algorithms are favored. A randomized, so-called fast Monte Carlo algorithm for the SVD and its application for clustering large graphs via the k-means algorithm is presented in [12].

The problem is also related to the Page-rank (see [19]). As for the microarray analysis, the authors of [20] use the leading singular values and vector pairs of the normalized contingency table to find a so-called checkerboard pattern in it, but they do not give estimation how this pattern approaches the original table. Some authors, e.g., [21], impose sparsity inducing conditions on the leading singular vector pairs, so that they have piecewise constant structure with many zero coordinates, and so, produce a checkerboard structure.

Though, many papers deal with the SVD-based biclustering of the underlying contingency table (see also [18,22,26]), they just introduce numerical algorithms, possibly with some constraints, which utilize the well-known favorable properties of low-rank approximations. After finding the checkerboard patterns, no inference is made on the homogeneity of the so obtained biclusters by means of the SVD of the table. It is also a drawback that the low-rank approximation, unlike the original table, may have negative entries.

In Section 2, we relate the biclustering problem to normalized two-way cuts of contingency tables, akin to the way normalized cuts of edge-weighted graphs are estimated by the normalized Laplacian spectra, see [10]. The minimization of this objective function favors biclusterings with dense diagonal, and sparse off-diagonal blocks. In terms of microarrays, it finds partition of genes and conditions into the same number of clusters such that to each cluster of conditions we can find a collection of genes responsible for this condition, and vice versa.

In Section 3, more generally, we are looking for so-called volume-regular row–column clusters pairs, such that the association between their row and column subsets is homogeneous, but not necessarily dense or sparse. The minimum of the pairwise discrepancies is related to the so-called structural singular values and corresponding eigen-functions of the normalized table. We use the one-cluster estimation of Butler [9], who estimates the discrepancy of the whole contingency table by the largest non-trivial singular value of the normalized table (one-cluster, rectangular case); moreover, he proves a two-sided relation between this singular value and the discrepancy. Here we extend the forward direction of this estimation for the k-cluster case, where our results also indicate the optimal choice of k. For this purpose, we use the result of Bolla [5] for estimating the constant of volume-regularity by the structural eigenvalues and the distances of the corresponding eigen-subspaces of the normalized modularity matrix of an edge-weighted graph. Since there are several eigenvalues (singular values) responsible for this versatile property, together with the corresponding eigenvectors (singular vector pairs), this statement is more complicated to prove and cannot be simply inverted, akin to the one-cluster case. Nonetheless, this problem has not yet been treated in the literature.

Section 4 is devoted to discussion and possible extension to directed graphs.

2 Normalized two-way cuts of contingency tables

Let $\mathbf{C}$ be a contingency table on row set $Row = \{1,\dots,n\}$ and column set $Col = \{1,\dots,m\}$, where $\mathbf{C}$ is $n \times m$ matrix of entries $c_{ij} \geq 0$. Without loss of generality, we suppose that there are no identically zero rows or columns. Here c_{ij} is some kind of association between the objects behind row i and column j, where 0 means no interaction at all.

Let the row- and column-sums of $\mathbf{C}$ be

$$d_{row,i} = \sum_{j=1}^{m} c_{ij} \quad (i = 1,\dots,n) \quad \text{and} \quad d_{col,j} = \sum_{i=1}^{n} c_{ij} \quad (j = 1,\dots,m)$$

which are collected in the main diagonals of the $n \times n$ and $m \times m$ diagonal matrices $\mathbf{D}_{row}$ and $\mathbf{D}_{col}$, respectively. The matrix

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

is called the *correspondence matrix (normalized contingency table)* belonging to the table $\mathbf{C}$, see [4]. If we multiply all the entries of $\mathbf{C}$ with the same positive constant, the correspondence matrix $\mathbf{C}_{corr}$ will not change. Therefore, without the loss of generality, $\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} = 1$ will be assumed in the sequel.

Given an integer k ($0 < k \leq \texttt{rank}\,(\mathbf{C})$), we want to simultaneously partition the rows and columns of $\mathbf{C}$ into disjoint, nonempty subsets

$$Row = R_1 \cup \dots \cup R_k, \quad Col = C_1 \cup \dots \cup C_k$$

such that we impose conditions on the *cuts* $c(R_a, C_b) = \sum_{i \in R_a}\sum_{j \in C_b} c_{ij}$ ($a, b = 1,\dots,k$) between the row-column cluster pairs. For this purpose, the following so-called *normalized two-way cut* of the contingency table with respect to the above k-partitions $P_{row} = (R_1,\dots,R_k)$ and $P_{col} = (C_1,\dots,C_k)$ of its rows and columns and the collection of signs σ is defined as follows:

$$\nu_k(P_{row}, P_{col}, \sigma) = \sum_{a=1}^{k}\sum_{b=1}^{k}\left(\frac{1}{\texttt{Vol}(R_a)} + \frac{1}{\texttt{Vol}(C_b)} + \frac{2\sigma_{ab}\delta_{ab}}{\sqrt{\texttt{Vol}(R_a)\texttt{Vol}(C_b)}}\right) c(R_a, C_b),$$

where

$$\texttt{Vol}(R_a) = \sum_{i \in R_a} d_{row,i} = \sum_{i \in R_a}\sum_{j=1}^{m} c_{ij}, \quad \texttt{Vol}(C_b) = \sum_{j \in C_b} d_{col,j} = \sum_{j \in C_b}\sum_{i=1}^{n} c_{ij}$$

are volumes of the clusters, δ_{ab} is the Kronecker delta-symbol, and the sign σ_{ab} is equal to 1 or -1 (it only has relevance in the $a = b$ case, when it helps balancing between the volumes of the same index row and column clusters), $\sigma := (\sigma_{11},\dots,\sigma_{kk})$. We want to minimize the above normalized two-way cut with respect to all possible k-partitions $\mathcal{P}_{row,k}$ and $\mathcal{P}_{col,k}$ of the rows and columns, further, to σ, simultaneously. The objective function penalizes row- and column clusters of extremely different volumes in the $a \neq b$ case, whereas in the $a = b$ case σ_{aa} moderates the balance between $\texttt{Vol}(R_a)$ and $\texttt{Vol}(C_a)$.

Definition 1 *The normalized two-way cut of the contingency table* $\mathbf{C}$ *is*

$$\nu_k(\mathbf{C}) = \min_{P_{row}, P_{col}, \sigma} \nu_k(P_{row}, P_{col}, \sigma).$$

Theorem 2 *Let* $1 = s_1 > s_2 \cdots \geq s_r$ *be the positive singular values of the correspondence matrix belonging to the contingency table* $\mathbf{C}$ *of rank* r *(we assume that* $\mathbf{C}\mathbf{C}^T$ *is irreducible). For any positive integer* $k \leq r$,

$$\nu_k(\mathbf{C}) \geq 2k - \sum_{i=1}^{k} s_i.$$

PROOF. We will show that $\nu_k(P_{row}, P_{col}, \sigma)$ is the value of the quadratic objective function Q_k, introduced below, taken on partition vectors belonging to P_{row} and P_{col}, respectively.

For a given integer $1 \leq k \leq \min\{n, m\}$, we are looking for k-dimensional representatives $\mathbf{r}_1, \ldots, \mathbf{r}_n$ of the rows and $\mathbf{c}_1, \ldots, \mathbf{c}_m$ of the columns such that they minimize the objective function

$$Q_k = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} \|\mathbf{r}_i - \mathbf{c}_j\|^2 \tag{2}$$

subject to

$$\sum_{i=1}^{n} d_{row,i} \mathbf{r}_i \mathbf{r}_i^T = \mathbf{I}_k, \quad \sum_{j=1}^{m} d_{col,j} \mathbf{c}_j \mathbf{c}_j^T = \mathbf{I}_k. \tag{3}$$

When minimized, the objective function Q_k favors k-dimensional placement of the rows and columns such that representatives of highly associated rows and columns are forced to be close to each other. This is equivalent to the problem of correspondence analysis. Indeed, let us put both the objective function and the constraints in a more favorable form. Let $\mathbf{X}$ be the $n \times k$ matrix of rows $\mathbf{r}_1^T, \ldots, \mathbf{r}_n^T$; let $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ denote the columns of $\mathbf{X}$, for which fact we use the notation $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$. Similarly, let $\mathbf{Y}$ be the $m \times k$ matrix of rows $\mathbf{c}_1^T, \ldots, \mathbf{c}_m^T$; let $\mathbf{y}_1, \ldots, \mathbf{y}_k \in \mathbb{R}^m$ denote the columns of $\mathbf{Y}$, i.e., $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_k)$. Hence, the constraints (3) can be formulated like

$$\mathbf{X}^T \mathbf{D}_{row} \mathbf{X} = \mathbf{I}_k, \quad \mathbf{Y}^T \mathbf{D}_{col} \mathbf{Y} = \mathbf{I}_k.$$

With this notation, the objective function (2) is

$$Q_k = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\|\mathbf{r}_i - \mathbf{c}_j\|^2 = \sum_{i=1}^{n} d_{row,i}\|\mathbf{r}_i\|^2 + \sum_{j=1}^{m} d_{col,j}\|\mathbf{c}_j\|^2 - \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\mathbf{r}_i^T\mathbf{c}_j$$
$$= 2k - \texttt{tr}\,\mathbf{X}^T\mathbf{C}\mathbf{Y} = 2k - \texttt{tr}\,(\mathbf{D}_{row}^{1/2}\mathbf{X})^T\mathbf{C}_{corr}(\mathbf{D}_{col}^{1/2}\mathbf{Y}). \tag{4}$$

The correspondence matrix (1) has SVD

$$\mathbf{C}_{corr} = \sum_{i=1}^{r} s_i\mathbf{v}_i\mathbf{u}_i^T, \tag{5}$$

where $r \le \min\{n,m\}$ is the rank of $\mathbf{C}_{corr}$, or equivalently (since there are not identically zero rows or columns), the rank of $\mathbf{C}$. Here $1 = s_1 > s_2 \ge \cdots \ge s_r > 0$ are the non-zero singular values of $\mathbf{C}_{corr}$, and 1 is a single singular value if the matrix $\mathbf{C}\mathbf{C}^T$ is irreducible. In this case, $\mathbf{v}_1 = (\sqrt{d_{row,1}},\dots,\sqrt{d_{row,n}})^T$ and $\mathbf{u}_1 = (\sqrt{d_{col,1}},\dots,\sqrt{d_{col,m}})^T$.

In view of (4), we have to maximize

$$\texttt{tr}\,(\mathbf{D}_{row}^{1/2}\mathbf{X})^T\mathbf{C}_{corr}(\mathbf{D}_{col}^{1/2}\mathbf{Y})$$

under the given constraints. By a simple linear algebra (see [1,25]) it follows that the minimum of (2) subject to (3) is $2k - \sum_{i=1}^{k} s_i$ and it is attained with the optimum row representatives $\mathbf{r}_1^*,\dots,\mathbf{r}_n^*$ and column representatives $\mathbf{c}_1^*,\dots,\mathbf{c}_m^*$, the transposes of which are row vectors of $\mathbf{X}^* = \mathbf{D}_{row}^{-1/2}(\mathbf{v}_1,\dots,\mathbf{v}_k)$ and $\mathbf{Y}^* = \mathbf{D}_{col}^{-1/2}(\mathbf{u}_1,\dots,\mathbf{u}_k)$, respectively. Since 1 is a single singular value, the first vector components are the constantly $\mathbf{1}$ vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and hence, the k-dimensional representation is realized in a $(k-1)$-dimensional hyperplane of $\mathbb{R}^k$.

Therefore, the statement of the theorem follows, as the overall minimum of Q_k is $2k - \sum_{i=1}^{k} s_i$, whereas the following special representation yields $\nu_k(P_{row}, P_{col}, \sigma)$ with given $P_{row} = (R_1,\dots,R_k)$, $P_{col} = (C_1,\dots,C_k)$, and σ. Indeed, let the ith coordinate of the left vector component $\mathbf{x}_a$ be

$$x_{ia} := \frac{1}{\sqrt{\texttt{Vol}(R_a)}} \quad \text{if} \quad i \in R_a,\ a = 1,\dots k;$$

similarly, let the jth coordinate of the right vector component $\mathbf{y}_b$ be

$$y_{jb} = \sigma_{bb}\frac{1}{\sqrt{\texttt{Vol}(C_b)}} \quad \text{if} \quad j \in C_b,\ b = 1,\dots,k,$$

otherwise the coordinates are zeros. With this, the matrices $\mathbf{X}$ and $\mathbf{Y}$ satisfy the conditions imposed on the representatives, further

$$\|\mathbf{r}_i - \mathbf{c}_j\|^2 = \frac{1}{\mathtt{Vol}(R_a)} + \frac{1}{\mathtt{Vol}(C_b)} + \frac{2\sigma_{bb}\delta_{ab}}{\sqrt{\mathtt{Vol}(R_a)\mathtt{Vol}(C_b)}}, \quad \text{if} \quad i \in R_a,\ j \in C_b.$$

Therefore, the objective function (2) becomes $\nu_k(P_{row}, P_{col}, \sigma)$, which is at least $2k - \sum_{i=1}^k s_i$. Consequently, $\nu_k(\mathbf{C}) \geq 2k - \sum_{i=1}^k s_i$.

We remark that Ding et al. [11] treat this problem for two row- and column-clusters and minimize another objective function such that it favors 2-partitions where $c(R_1, C_2)$ and $c(R_2, C_1)$ are small compared to $c(R_1, C_1)$ and $c(R_2, C_2)$. The solution is also given by the transformed $\mathbf{v}_2, \mathbf{u}_2$ pair. However, it is the objective function $\nu_k(\mathbf{C})$ which best complies with the SVD of the correspondence matrix, and hence, gives the continuous relaxation of the normalized cut minimization problem. Dhillon [10] also suggests a multipartition algorithm that runs the k-means algorithm simultaneously for the row and column representatives.

3 Regular row-column cluster pairs

Let us start with the one-cluster case. Let $\mathbf{C}$ be an $n \times m$ contingency table and $\mathbf{C}_{corr}$ be the correspondence matrix belonging to it. The Expander Mixing Lemma for edge-weighted graphs naturally extends to this situation, see the following result of [9].

Proposition 3 *Let $\mathbf{C}$ be a contingency table ($\mathbf{C}\mathbf{C}^T$ is irreducible) on row set Row and column set Col, and of total volume 1. Then for all $R \subset Row$ and $C \subset Col$*

$$|c(R, C) - \mathtt{Vol}(R)\mathtt{Vol}(C)| \leq s_2\sqrt{\mathtt{Vol}(R)\mathtt{Vol}(C)},$$

where s_2 is the largest but 1 singular value of the normalized contingency table $\mathbf{C}_{corr}$.

Since the spectral gap of $\mathbf{C}_{corr}$ is $1 - s_2$, in view of the above Expander Mixing Lemma, 'large' spectral gap is an indication that the weighted cut between any row and column subset of the contingency table is near to what is expected in a random table. The following notion of *discrepancy* just measures the deviation from this random situation. The discrepancy (see [9]) of the contingency table $\mathbf{C}$ of total volume 1 is the smallest $\alpha > 0$ such that for all $R \subset Row$ and $C \subset Col$

$$|c(R, C) - \mathtt{Vol}(R)\mathtt{Vol}(C)| \leq \alpha\sqrt{\mathtt{Vol}(R)\mathtt{Vol}(C)}.$$

In view of this, the result of Proposition 3 can be interpreted as follows: α singular value separation causes α discrepancy, where the singular value separation is the second largest singular value of the normalized contingency table, which is the smaller, the bigger the separation between the largest singular value (the 1) of the normalized contingency table and the other singular values is. Based on the ideas of [2] and [6], Butler [9] proves the converse of the Expander Mixing Lemma for contingency tables, namely that

$$s_2 \leq 150\alpha(1 - 8\log\alpha).$$

Now we extend the notion of discrepancy to volume-regular cluster pairs.

Definition 4 *The row–column cluster pair $R \subset Row$, $C \subset Col$ of the contingency table $\mathbf{C}$ of total volume 1 is γ-volume regular if for all $X \subset R$ and $Y \subset C$ the relation*

$$|c(X,Y) - \rho(R,C)\mathtt{Vol}(X)\mathtt{Vol}(Y)| \leq \gamma\sqrt{\mathtt{Vol}(R)\mathtt{Vol}(C)} \tag{6}$$

holds, where $\rho(R,C) = \frac{c(R,C)}{\mathtt{Vol}(R)\mathtt{Vol}(C)}$ is the relative inter-cluster density of the row–column pair R, C.

We will show that for given k, if the clusters are formed via applying the weighted k-means algorithm for the optimal row- and column representatives, respectively, then the so obtained row–column cluster pairs are homogeneous in the sense that they form equally dense parts of the contingency table. More precisely, the constant γ of the volume regularity of the pairs will be related to the SVD of $\mathbf{C}_{corr}$. To this end, we introduce the following notion.

The weighted k-variance of the k-dimensional row representatives is defined by

$$S_k^2(\mathbf{X}) = \min_{(R_1,\dots,R_k)} \sum_{a=1}^{k} \sum_{j \in R_a} d_{row,j} \|\mathbf{r}_j - \bar{\mathbf{r}}_a\|^2, \tag{7}$$

where $\bar{\mathbf{r}}_a = \frac{1}{\mathtt{Vol}(R_a)} \sum_{j \in R_a} d_{row,j}\mathbf{r}_j$ is the weighted center of cluster R_a ($a = 1, \dots, k$). Similarly, the weighted k-variance of the k-dimensional column representatives is

$$S_k^2(\mathbf{Y}) = \min_{(C_1,\dots,C_k)} \sum_{a=1}^{k} \sum_{j \in C_a} d_{col,j} \|\mathbf{c}_j - \bar{\mathbf{c}}_a\|^2, \tag{8}$$

where $\bar{\mathbf{c}}_a = \frac{1}{\mathtt{Vol}(C_a)} \sum_{j \in C_a} d_{col,j}\mathbf{c}_j$ is the weighted center of cluster C_a ($a = 1, \dots, k$). Observe, that the trivial vector components can be omitted, and the k-variance of the so obtained $(k-1)$-dimensional representatives will be the same.

We need the following definition of the cut-norm of a matrix (see [7,14,23]).

Definition 5 *The cut-norm of the rectangular 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 6 *For the cut-norm of the $n \times m$ real matrix $\mathbf{A}$*

$$\|\mathbf{A}\|_{\square} \le \sqrt{nm} \|\mathbf{A}\|$$

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

PROOF.

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

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

Theorem 7 *Let $\mathbf{C}$ be a contingency table of n-element row set Row and m-element column set Col, with row- and column sums $d_{row,1}, \dots, d_{row,n}$ and $d_{col,1}, \dots, d_{col,m}$, respectively. Assume that $\mathbf{C}\mathbf{C}^T$ is irreducible, $\sum_{i=1}^n \sum_{j=1}^m c_{ij} = 1$, and there are no dominant rows and columns, i.e., $d_{row,i} = \Theta(1/n)$, $(i = 1, \dots, n)$ and $d_{col,j} = \Theta(1/m)$, $(j = 1, \dots, m)$ as $n, m \to \infty$. Let the singular values of $\mathbf{C}_{corr}$ be*

$$1 = s_1 > s_2 \ge \dots \ge s_k > \varepsilon \ge s_i, \quad i \ge k+1.$$

The partition $(R_1, \dots, R_k)$ of Row and $(C_1, \dots, C_k)$ of Col are defined so that they minimize the weighted k-variances $S_k^2(\mathbf{X})$ and $S_k^2(\mathbf{Y})$ of the row and column representatives defined in (7) and (8), respectively. Suppose that there are constants $0 < K_1, K_2 \le \frac{1}{k}$ such that $|R_i| \ge K_1 n$ and $|C_i| \ge K_2 m$ $(i = 1, \dots, k)$, respectively. Then the R_i, C_j pairs are $\mathcal{O}(\sqrt{2k}(S_k(\mathbf{X}) S_k(\mathbf{Y})) + \varepsilon)$-volume regular $(i, j = 1, \dots, k)$.

PROOF. Recall that, provided $\mathbf{C}\mathbf{C}^T$ is irreducible, the largest singular value $s_1 = 1$ of $\mathbf{C}_{corr}$ is single with corresponding singular vector pair $\mathbf{v}_1 = \mathbf{D}_{row}^{1/2} \mathbf{1}$ and $\mathbf{u}_1 = \mathbf{D}_{col}^{1/2} \mathbf{1}$ with the constantly $\mathbf{1}$ vectors of appropriate size. The optimal k-dimensional representatives of the rows and columns are row vectors of the matrices $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ and $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_k)$, where $\mathbf{x}_i = \mathbf{D}_{row}^{-1/2} \mathbf{v}_i$ and $\mathbf{y}_i = \mathbf{D}_{col}^{-1/2} \mathbf{u}_i$, respectively $(i = 1, \dots, k)$, using the SVD (5) of $\mathbf{C}_{corr}$. Assume that the minimum k-variance is attained on the k-partition $(R_1, \dots, R_k)$ of the rows and $(C_1, \dots, C_k)$ of the columns. By an easy analysis of variance argument of [3] it follows that

$$S_k^2(\mathbf{X}) = \sum_{i=1}^{k} \mathtt{dist}^2(\mathbf{v}_i, F), \quad S_k^2(\mathbf{Y}) = \sum_{i=1}^{k} \mathtt{dist}^2(\mathbf{u}_i, G),$$

where $F = \mathtt{Span}\,\{\mathbf{D}_{row}^{1/2}\mathbf{w}_1, \dots, \mathbf{D}_{row}^{1/2}\mathbf{w}_k\}$ and $G = \mathtt{Span}\,\{\mathbf{D}_{col}^{1/2}\mathbf{z}_1, \dots, \mathbf{D}_{col}^{1/2}\mathbf{z}_k\}$ with the so-called normalized row partition vectors $\mathbf{w}_1, \dots, \mathbf{w}_k$ of coordinates $w_{ji} = \frac{1}{\sqrt{\mathtt{Vol}(R_i)}}$ if $j \in R_i$ and 0, otherwise; and column partition vectors $\mathbf{z}_1, \dots, \mathbf{z}_k$ of coordinates $z_{ji} = \frac{1}{\sqrt{\mathtt{Vol}(C_i)}}$ if $j \in C_i$ and 0, otherwise $(i = 1, \dots, k)$. Note that the vectors $\mathbf{D}_{row}^{1/2}\mathbf{w}_1, \dots, \mathbf{D}_{row}^{1/2}\mathbf{w}_k$ and $\mathbf{D}_{col}^{1/2}\mathbf{z}_1, \dots, \mathbf{D}_{col}^{1/2}\mathbf{z}_k$ form orthonormal systems in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively (but they are, usually, not complete). Using a statement of [3], we can find orthonormal systems $\tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_k \in F$ and $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_k \in G$ such that

$$S_k^2(\mathbf{X}) \le \sum_{i=1}^{k} \|\mathbf{v}_i - \tilde{\mathbf{v}}_i\|^2 \le 2S_k^2(\mathbf{X}), \quad S_k^2(\mathbf{Y}) \le \sum_{i=1}^{k} \|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2 \le 2S_k^2(\mathbf{Y}).$$

We approximate the matrix $\mathbf{C}_{corr} = \sum_{i=1}^{r} s_i \mathbf{v}_i \mathbf{u}_i^T$ by the rank k matrix $\sum_{i=1}^{k} s_i \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T$ with the following accuracy (in spectral norm):

$$\left\| \sum_{i=1}^{r} s_i \mathbf{v}_i \mathbf{u}_i^T - \sum_{i=1}^{k} s_i \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T \right\| \le \sum_{i=1}^{k} s_i \left\| \mathbf{v}_i \mathbf{u}_i^T - \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T \right\| + \left\| \sum_{i=k+1}^{r} s_i \mathbf{v}_i \mathbf{u}_i^T \right\|, \qquad (9)$$

where the spectral norm of the last term is at most ε, and the the individual terms of the first one are estimated from above in the following way.

$$\begin{aligned}
s_i \|\mathbf{v}_i \mathbf{u}_i^T - \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T\| &\le \|(\mathbf{v}_i \mathbf{u}_i^T - \tilde{\mathbf{v}}_i \mathbf{u}_i^T) + (\tilde{\mathbf{v}}_i \mathbf{u}_i^T - \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T)\| \\
&\le \|(\mathbf{v}_i - \tilde{\mathbf{v}}_i)\mathbf{u}_i^T\| + \|\tilde{\mathbf{v}}_i(\mathbf{u}_i - \tilde{\mathbf{u}}_i)^T\| \\
&= \sqrt{\|(\mathbf{v}_i - \tilde{\mathbf{v}}_i)\mathbf{u}_i^T \mathbf{u}_i (\mathbf{v}_i - \tilde{\mathbf{v}}_i)^T\|} + \sqrt{\|(\mathbf{u}_i - \tilde{\mathbf{u}}_i)\tilde{\mathbf{v}}_i^T \tilde{\mathbf{v}}_i (\mathbf{u}_i - \tilde{\mathbf{u}}_i)^T\|} \\
&= \sqrt{(\mathbf{v}_i - \tilde{\mathbf{v}}_i)^T (\mathbf{v}_i - \tilde{\mathbf{v}}_i)} + \sqrt{(\mathbf{u}_i - \tilde{\mathbf{u}}_i)^T (\mathbf{u}_i - \tilde{\mathbf{u}}_i)} \\
&= \|\mathbf{v}_i - \tilde{\mathbf{v}}_i\| + \|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|,
\end{aligned}$$

where we exploited that the spectral norm (i.e., the largest singular value) of an $n \times m$ matrix $\mathbf{A}$ is equal to either the squareroot of the largest eigenvalue of the matrix $\mathbf{A}\mathbf{A}^T$ or equivalently, that of $\mathbf{A}^T\mathbf{A}$. In the above calculations all of these matrices are of rank 1, hence, the largest eigenvalue of the symmetric, positive semidefinite matrix under the squareroot is the only non-zero eigenvalue of it, therefore, it is equal to its trace; finally, we used the commutativity of the trace, and in the last line we have the usual vector norm.

Therefore the first term in (9) can be estimated from above by

$$\sum_{i=1}^{k} \|\mathbf{v}_i \mathbf{u}_i^T - \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T\| \le \sqrt{k} \sqrt{\sum_{i=1}^{k} \|\mathbf{v}_i - \tilde{\mathbf{v}}_i\|^2} + \sqrt{k} \sqrt{\sum_{i=1}^{k} \|\mathbf{u}_i - \tilde{\mathbf{u}}_i\|^2}$$
$$\le \sqrt{k}(\sqrt{2S_k^2(\mathbf{X})} + \sqrt{2S_k^2(\mathbf{Y})}) = \sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})).$$

Based on these considerations and relation between the cut norm and the spectral norm (see Lemma 6), the densities to be estimated in the defining formula (6) of volume regularity can be written in terms of stepwise constant vectors in the following way. The vectors $\hat{\mathbf{v}}_i := \mathbf{D}_{row}^{-1/2} \tilde{\mathbf{v}}_i$ are stepwise constants on the partition $(R_1, \dots, R_k)$ of the rows; whereas the vectors $\hat{\mathbf{u}}_i := \mathbf{D}_{col}^{-1/2} \tilde{\mathbf{u}}_i$ are stepwise constants on the partition $(C_1, \dots, C_k)$ of the columns, $i = 1, \dots, k$. The matrix

$$\sum_{i=1}^{k} s_i \hat{\mathbf{v}}_i \hat{\mathbf{u}}_i^T$$

is therefore an $n \times m$ block-matrix on $k \times k$ blocks belonging to the above partition of the rows and columns. Let $\hat{c}_{ab}$ denote its entries in the a, b block $(a, b = 1, \dots, k)$. Using (9), the rank k approximation of the matrix $\mathbf{C}$ is performed with the following accuracy of the perturbation $\mathbf{E}$ in spectral norm:

$$\|\mathbf{E}\| = \left\| \mathbf{C} - \mathbf{D}_{row} (\sum_{i=1}^{k} s_i \hat{\mathbf{v}}_i \hat{\mathbf{u}}_i^T) \mathbf{D}_{col} \right\| = \left\| \mathbf{D}_{row}^{1/2} (\mathbf{C}_{corr} - \sum_{i=1}^{k} s_i \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T) \mathbf{D}_{col}^{1/2} \right\|.$$

Therefore, the entries of $\mathbf{C}$ – for $i \in R_a$, $j \in C_b$ – can be decomposed as

$$c_{ij} = d_{row,i} d_{col,j} \hat{c}_{ab} + \eta_{ij},$$

where the cut norm of the $n \times m$ error matrix $\mathbf{E} = (\eta_{ij})$ restricted to $R_a \times C_b$ (otherwise it contains entries all zeroes) and denoted by $\mathbf{E}_{ab}$, is estimated as follows:

$$\|\mathbf{E}_{ab}\|_\square \le \sqrt{mn} \|\mathbf{E}_{ab}\| \le \sqrt{nm} \cdot \|\mathbf{D}_{row,a}^{1/2}\| \cdot (\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon) \cdot \|\mathbf{D}_{col,b}^{1/2}\|$$
$$\le \sqrt{nm} \sqrt{c_1 \frac{\mathtt{Vol}(R_a)}{|R_a|}} \cdot \sqrt{c_2 \frac{\mathtt{Vol}(C_b)}{|C_b|}} (\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon)$$
$$= \sqrt{c_1 c_2} \cdot \sqrt{\frac{n}{|R_a|}} \cdot \sqrt{\frac{m}{|C_b|}} \cdot \sqrt{\mathtt{Vol}(R_a)} \sqrt{\mathtt{Vol}(C_b)} (\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon)$$
$$\le \sqrt{\frac{c_1 c_2}{K_1 K_2}} \sqrt{\mathtt{Vol}(R_a)} \sqrt{\mathtt{Vol}(C_b)} (\sqrt{2k} s + \varepsilon)$$
$$= c \sqrt{\mathtt{Vol}(R_a)} \sqrt{\mathtt{Vol}(C_b)} (\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon),$$

where the $n \times n$ diagonal matrix $\mathbf{D}_{row,a}$ inherits $\mathbf{D}_{row}$'s diagonal entries over R_a; whereas the $m \times m$ diagonal matrix $\mathbf{D}_{col,b}$ inherits $\mathbf{D}_{col}$'s diagonal entries over C_b, otherwise they are zeros. Further, the constants c_1, c_2 are due to the fact that there are no dominant rows and columns, while K_1, K_2 are derived from the cluster size balancing conditions. Hence, the constant c does not depend on n and m. Consequently, for $a, b = 1, \dots, k$ and $X \subset R_a$, $Y \subset C_b$:

$$|c(X,Y) - \rho(R_a, C_b)\mathtt{Vol}(X)\mathtt{Vol}(Y)| =$$
$$\left| \sum_{i\in X}\sum_{j\in Y}(d_{row,i}d_{col,j}\hat{c}_{ab} + \eta_{ij}) - \frac{\mathtt{Vol}(X)\mathtt{Vol}(Y)}{\mathtt{Vol}(R_a)\mathtt{Vol}(C_b)}\sum_{i\in R_a}\sum_{j\in C_b}(d_{row,i}d_{col,j}\hat{c}_{ab} + \eta_{ij}) \right| =$$
$$\left| \sum_{i\in X}\sum_{j\in Y}\eta_{ij} - \frac{\mathtt{Vol}(X)\mathtt{Vol}(Y)}{\mathtt{Vol}(R_a)\mathtt{Vol}(C_b)}\sum_{i\in R_a}\sum_{j\in C_b}\eta_{ij} \right| \le 2\|\mathbf{E}_{ab}\|_\square$$
$$\le 2c(\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon)\sqrt{\mathtt{Vol}(R_a)\mathtt{Vol}(C_b)},$$

that gives the required statement for $a, b = 1, \dots, k$.

Note that when we use Definition 4 of γ-volume regularity for the row–column cluster pairs R_i, C_j $(i, j = 1, \dots, k)$, then we may say that the *k-way discrepancy* of the underlying contingency table is the minimum γ for which all the row–column cluster pairs are γ-volume regular. With this nomenclature, Theorem 7 states that the k-way discrepancy of a contingency table can be estimated from above by the the kth largest non-trivial singular value of the correspondence matrix and the k-variance of the clusters obtained by the left and right singular vectors corresponding to the k largest singular values of this matrix. Hence, SVD based representation is applicable to find volume regular cluster pairs for a k, where k is the number of structural (protruding) singular values. Unfortunately, we are not able to invert this statement, as there are too many elements to be taken into consideration, not just one non-trivial singular value and corresponding vector pair.

We also remark that when we perform a low-rank approximation of a contingency table of nonnegative entries, the entries of this approximating matrix will usually have both positive or negative entries. Nonetheless, the entries $\hat{c}_{ij}$'s of the block-matrix constructed in the proof of Theorem 7 will already be positive provided the weighted k-variances $S_k^2(\mathbf{X})$ and $S_k^2(\mathbf{Y})$ are 'small' enough. Indeed, with the notation used in the proof, denote by ab in the lower index the matrix restricted to the $R_a \times C_b$ block (otherwise it has zero entries). Then for the squared Frobenius norm of the rank k approximation of $\mathbf{D}_{row}^{-1}\mathbf{C}\mathbf{D}_{col}^{-1}$, restricted to the ab block, we have that

$$\begin{aligned} \left\| \mathbf{D}_{row,a}^{-1}\mathbf{C}_{ab}\mathbf{D}_{col,b}^{-1} - (\sum_{i=1}^k s_i\hat{\mathbf{v}}_i\hat{\mathbf{u}}_i^T)_{ab} \right\|_2^2 &= \sum_{i\in R_a}\sum_{j\in C_b}(\frac{c_{ij}}{d_{row,i}d_{col,j}} - \hat{c}_{ab})^2 \\ = \sum_{i\in R_a}\sum_{j\in C_b}(\frac{c_{ij}}{d_{row,i}d_{col,j}} - \bar{c}_{ab})^2 &+ |R_a||C_b|(\bar{c}_{ab} - \hat{c}_{ab})^2 \end{aligned} \tag{10}$$

where we used the Steiner equality with the average $\bar{c}_{ab}$ of the entries of $\mathbf{D}_{row}^{-1}\mathbf{C}\mathbf{D}_{col}^{-1}$ in the ab block. Now we estimate the above Frobenius norm by a constant multiple of the spectral norm, where for the spectral norm

$$\left\| \mathbf{D}_{row,a}^{-1}\mathbf{C}_{ab}\mathbf{D}_{col,b}^{-1} - (\sum_{i=1}^{k} s_i \hat{\mathbf{v}}_i \hat{\mathbf{u}}_i^T)_{ab} \right\| = \left\| \mathbf{D}_{row,a}^{-1/2}(\mathbf{C}_{corr} - \sum_{i=1}^{k} s_i \tilde{\mathbf{v}}_i \tilde{\mathbf{u}}_i^T)_{ab}\mathbf{D}_{col,b}^{-1/2} \right\|$$
$$\leq \max_{i\in R_a} \frac{1}{\sqrt{d_{row,i}}} \cdot \max_{j\in C_b} \frac{1}{\sqrt{d_{col,j}}} \cdot [\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon]$$

holds. Therefore,

$$\left\| \mathbf{D}_{row,a}^{-1}\mathbf{C}_{ab}\mathbf{D}_{col,b}^{-1} - (\sum_{i=1}^{k} s_i \hat{\mathbf{v}}_i \hat{\mathbf{u}}_i^T)_{ab} \right\|_2^2$$
$$\leq \min\{|R_a|, |C_b|\} \cdot \max_{i\in R_a} \frac{1}{d_{row,i}} \cdot \max_{j\in C_b} \frac{1}{d_{col,j}} \cdot [\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon]^2.$$

Consequently, in view of (10),

$$(\bar{c}_{ab} - \hat{c}_{ab})^2 \leq \frac{1}{\max\{|R_a|, |C_b|\}} \cdot \max_{i\in R_a} \frac{1}{d_{row,i}} \cdot \max_{j\in C_b} \frac{1}{d_{col,j}} \cdot [\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon]^2.$$

But using the conditions on the block sizes and the row- and column-sums of Theorem 7, provided

$$\sqrt{2k}(S_k(\mathbf{X}) + S_k(\mathbf{Y})) + \varepsilon) = \mathcal{O}\left(\frac{1}{(\min\{m,n\})^{\frac{1}{2}+\tau}} \right)$$

holds with some 'small' $\tau > 0$, the relation $\bar{c}_{ab} - \hat{c}_{ab} \to 0$ also holds as $n, m \to \infty$. Therefore, both $\hat{c}_{ab}$ and $\hat{c}_{ab} d_{row,i} d_{col,j}$ are positive over blocks that are not constantly zero in the original table when m and n are large enough.

4 Discussion and extension to directed graphs

In the ideal k-cluster case, we consider the following generalized random binary contingency table model: given the partition $(R_1, \dots, R_k)$ of the rows and $(C_1, \dots, C_k)$ of the columns, the entry in the row $i \in R_a$ and column $j \in C_b$ is 1 with probability p_{ab}, and 0 otherwise, independently of other rows of R_a and columns of C_b, $1 \leq a, b \leq k$. We can think of the probability p_{ab} as the inter-cluster density of the row–column cluster pair R_a, C_b. Since generalized contingency tables can be viewed as block-matrices (with $k \times k$ blocks) burdened with a general random noise, in [4], we gave the following spectral characterization of them. Fixing k, and tending with n and m to infinity in such a way that the cluster sizes grow at the same rate and also n

and m subpolynomially, there exists a positive number $\theta \le 1$, independent of n and m, such that for every $0 < \tau < 1/2$ there are exactly k singular values of $\mathbf{C}_{corr}$ greater than $\theta - \max\{n^{-\tau}, m^{-\tau}\}$, while all the others are at most $\max\{n^{-\tau}, m^{-\tau}\}$; further, the weighted k-variance of the row and column representatives constructed by the k transformed structural left and right singular vectors is $\mathcal{O}(\max\{n^{-\tau}, m^{-\tau}\})$, respectively. For non-random contingency tables, our results imply that the existence of k singular values of $\mathbf{C}_{corr}$, separated from 0 by ε, is indication of a k-cluster structure, while the singular values accumulating around 0 are responsible for the pairwise regularities. The clusters themselves can be recovered by applying the k-means algorithm for the row and column representatives obtained via the left and right singular vectors corresponding to the structural singular values. In [4], we allowed the number of row- and column-clusters to be larger than k and not necessarily be the same when making inferences on the residual spectral norm. However, in the context of the discrepancy, we cannot see a direct way how the proof of Theorem 7 could be adopted to this situation.

We can consider quadratic, but not symmetric contingency tables with zero diagonal as edge-weight matrices of directed graphs. The $n \times n$ edge-weight matrix $\mathbf{W}$ of a directed graph has zero diagonal, but is usually not symmetric: w_{ij} is the weight of the $i \to j$ edge ($i, j = 1, \dots, n$; $i \ne j$). In this setup, the generalized in- and out-degrees are

$$d_{out,i} = \sum_{j=1}^{n} w_{ij} \quad (i = 1, \dots, n) \quad \text{and} \quad d_{in,j} = \sum_{i=1}^{n} w_{ij} \quad (j = 1, \dots, n);$$

further, $\mathbf{D}_{in} = \mathtt{diag}\,(d_{in,1}, \dots, \mathrm{d}_{in,n})$ and $\mathbf{D}_{out} = \mathtt{diag}\,(d_{out,1}, \dots, \mathrm{d}_{out,n})$ are the in- and out-degree matrices. Suppose that there are no sources and sinks (i.e. no zero out- and in-degrees), further, that $\mathbf{W}\mathbf{W}^T$ is irreducible. Then the correspondence matrix belonging to $\mathbf{W}$ is

$$\mathbf{W}_{corr} = \mathbf{D}_{out}^{-1/2} \mathbf{W} \mathbf{D}_{in}^{-1/2},$$

and its SVD is used to minimize the normalized two-way cut of $\mathbf{W}$ as a contingency table, see Section 2. Butler [9] generalized the Expander Mixing Lemma for this situation. We can further generalize it to obtain regular in- and out-vertex cluster pairs, for a given k, in the following sense. The V_{in}, V_{out} in- and out-vertex cluster pair of the directed graph (with sum of the weights of directed edges 1) is γ-volume regular if for all $X \subset V_{out}$ and $Y \subset V_{in}$ the relation

$$|w(X, Y) - \rho(V_{out}, V_{in}) \mathtt{Vol}_{out}(X) \mathtt{Vol}_{in}(Y)| \le \gamma \sqrt{\mathtt{Vol}_{out}(V_{out}) \mathtt{Vol}_{in}(V_{in})}$$

holds, where the *directed cut* $w(X, Y)$ is the sum the weights of the $X \to Y$ edges, $\mathtt{Vol}_{out}(X) = \sum_{i \in X} d_{out,i}$, $\mathtt{Vol}_{in}(Y) = \sum_{j \in Y} d_{in,j}$, and $\rho(V_{out}, V_{in}) = \frac{w(V_{out}, V_{in})}{\mathtt{Vol}_{out}(V_{out}) \mathtt{Vol}_{in}(V_{in})}$ is the relative inter-cluster density of the out–in cluster pair

V_{out}, V_{in}. The clustering $(V_{in,1}, \dots, V_{in,k})$ and $(V_{out,1}, \dots, V_{out,k})$ of the columns and rows – guaranteed by Theorem 7 – corresponds to in- and out-clusters of the same vertex set such that the directed information flow $V_{out,a} \to V_{in,b}$ is as homogeneous as possible for all $a, b = 1, \dots, k$ pairs.

Acknowledgements

I am indebted to the anonymous referees for many useful suggestions.

References

[1] Bhatia, R., *Matrix Analysis*, Springer (1996).

[2] Bilu, Y. and Linial, N., Lifts, discrepancy and nearly optimal spectral gap, *Combinatorica* **26** (2006), 495–519.

[3] Bolla, M., Tusnády, G., Spectra and optimal partitions of weighted graphs, *Discrete Mathematics* **128** (1994), 1-20.

[4] Bolla, M., Friedl, K., Krámli, A., Singular value decomposition of large random matrices (for two-way classification of microarrays), *Journal of Multivariate Analysis* **101** (2010), 434–446.

[5] Bolla, M., Modularity spectra, eigen-subspaces and structure of weighted graphs, *European Journal of Combinatorics* **35** (2014), 105-116.

[6] Bollobás, B., Nikiforov, V., Hermitian matrices and graphs: singular values and discrepancy, *Discrete Mathematics* **285** (2004), 17–32.

[7] Borgs, C., Chayes, J. T., Lovász, L., T.-Sós, V., Vesztergombi, K., Convergent graph sequences I: Subgraph Frequencies, metric properties, and testing. *Advances in Math.* **219** (2008), 1801–1851.

[8] Busygin, S., Prokopyev, O., Pardalos, P. M., Biclustering in data mining, *Computers and Operations Research* **35** (2008), 2964–2987.

[9] Butler, S., Using discrepancy to control singular values for nonnegative matrices, *Lin. Alg. Appl.* **419** (2006), 486–493.

[10] Dhillon, I. S., Co-clustering documents and words using bipartite spectral graph partitioning. In *Proc. ACM Int. Conf. Knowledge Disc. Data Mining (KDD 2001)* (Provost, F. J. and Srikant, R. eds), Association for Computer Machinery, New York (2001), 269–274.

[11] Ding, C., He, X., Zha, H., Gu, M., Simon, H. D., A minmax cut spectral method for data clustering and data partitioning, Lawrence Berkeley National Laboratory Tech. Rep. 54111, 2003.

[12] Drineas, P., Frieze, A., Kannan, R., Vempala, S., Vinay, V., Clustering large graphs via the singular value decomposition, *Mach. Learn.* **56** (2004), 9–33.

[13] Faust, K., Using correspondence analysis for joint displays of affiliation networks. In *Models and Methods in Social Network Analysis* **7** (Carrington, P. J. and Scott, J. eds), Cambridge Univ. Press, Cambridge (2005), 117–147.

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

[15] Frieze, A., Kannan, R., Vempala, S., Fast Monte-Carlo Algorithms for finding low-rank approximations. In *Proc. of the 39th Annual IEEE Symposium on Foundations of Computer Science (FOCS)* (1998), 370–386.

[16] Greenacre, M. J., *Theory and Applications of Correspondence Analysis.* Academic Press, New York (1984).

[17] Hartigan, J. A., Direct clustering of a data matrix, *J. Am. Stat. Assoc.* **67** (1972), 123–129.

[18] Higham, D. J., Kalna, G., Vass, J. K., Analysis of the singular value decomposition as a tool for processing microarray expression data, *Proc. of ALGORITMY* (2005) 250–259.

[19] Kleinberg, J., Authoritative sources in hyperlinked environment, IBM Research Report RJ 10076 (91892), 1997.

[20] Kluger, Y., Basri, R., Chang, J. T., Gerstein, M., Spectral biclustering of microarray data: clustering genes and conditions, *Genome Research* **13** (2003), 703–716.

[21] Lee, M., Shen, H, Huang, J. Z., Marron, J. S., Biclustering via sparse singular value decomposition, *Biometrics* **66** (2010), 1087–1095.

[22] Liu, L., Hawkins, D. M., Ghosh, S., Young, S. S., Robust singular value decomposition analysis of microarray data, *Proc. Natl. Acad. Sci. USA* **100** (2003), 13167–13172.

[23] Lovász, L., Szegedy, B., Limits of dense graph sequences, *J. Comb. Theory B* **96** (2006), 933–957.

[24] Madeira, S. C., Oliveira, A. L., Biclustering algorithms for biological data analysis: A survey, *IEEE-ACM Trans. Comput. Biol. Bioinform.* **1** (2004), 24–45.

[25] Rao, C. R., Separation theorems for singular values of matrices and their applications in multivariate analysis, *J. Multivariate Analysis* **9** (1979), 362–377.

[26] Yang, W., Dai, D., Yan, H., Biclustering of microarray data based on singular value decomposition, *Engineering Technologies in Knowledge Discovery and Data Mining*, 194–205.