Research Article

Marianna Bolla*, Vilas Winstein, Tao You, Frank Seidl, and Fatma Abdelkhalek

Regularity-based spectral clustering and mapping the Fiedler-carpet

https://doi.org/10.1515/spma-2022-0167
received November 29, 2021; accepted April 13, 2022

Abstract: We discuss spectral clustering from a variety of perspectives that include extending techniques to rectangular arrays, considering the problem of discrepancy minimization, and applying the methods to directed graphs. Near-optimal clusters can be obtained by singular value decomposition together with the weighted k-means algorithm. In the case of rectangular arrays, this means enhancing the method of correspondence analysis with clustering, while in the case of edge-weighted graphs, a normalized Laplacian-based clustering. In the latter case, it is proved that a spectral gap between the $(k-1)$st and kth smallest positive eigenvalues of the normalized Laplacian matrix gives rise to a sudden decrease of the inner cluster variances when the number of clusters of the vertex representatives is 2^{k-1}, but only the first $k-1$ eigenvectors are used in the representation. The ensemble of these eigenvectors constitute the so-called Fiedler-carpet.

Keywords: multiway discrepancy, correspondence analysis, normalized Laplacian, multiway cuts, Fiedler-vector, weighted k-means algorithm

MSC 2020: 05C50, 62H25, 65H10

1 Introduction

The spectral graph theory started developing about 50 years ago (see, e.g., Hoffman [1], Fiedler [2], Cvetković et al. [3], and Chung [4]), with the goal of characterizing certain structural properties of a graph by means of the eigenvalues of its adjacency or Laplacian matrix. In the last decades of the 20th century, the eigenvectors corresponding to the smallest few eigenvalues of the Laplacian matrix were used to obtain clusterings of the vertices into disjoint parts so that the inter-cluster relations are negligible compared to the intra-cluster ones. The motivation for finding such clusterings comes from machine learning, where the classification of data into a small number of highly similar clusters is a common task. In this setup, the famous *Fiedler-vector*, the eigenvector, corresponding to the smallest positive Laplacian eigenvalue was used to classify the vertices into two parts. To find a clustering with $k \geq 2$ parts, one can instead use the $k-1$ eigenvectors corresponding to the $k-1$ smallest positive eigenvalues. It was also noted in [5] that it is better to have the more eigenvectors but no exact estimates between the spectral gap and the quality of the clustering were available, except the

* **Corresponding author: Marianna Bolla,** Deprtement of Stochastic (DS), Budapest University of Technology and Economics (BME), Budapest, Hungary, e-mail: marib@math.bme.hu
Vilas Winstein: Renyi Institute of Mathematics, Budapest, Hungary, e-mail: vilas@renyi.hu
Tao You: Department of Mathematics, Middlebury College, Middlebury, United States, e-mail: tyou@middlebury.edu
Frank Seidl: Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, United States, e-mail: fcseidl@umich.edu
Fatma Abdelkhalek: Faculty of Commerce, Assiut University, Assiut Governorate, Egypt, e-mail: fatma.said@aun.edu.eg

$k = 1$ and $k = 2$ cases; the former case is related to the isoperimetric number and expander graphs (see, e.g., [4,6,7]), while the latter one is related to the sum of the inner variances of 2-clusterings (see [8]).

Since then, the problem has been generalized in several ways: to edge-weighted graphs and rectangular arrays of nonnegative entries (e.g., microarrays in biological genetics and forensic science [9–12]), and to degree-corrected adjacency and Laplacian matrices [13,14]. Starting in the 21st century, physicists and social scientists introduced modularity matrices and investigated the so-called anticommunity structures (where intracluster relations are negligible compared to the intercluster ones) in contrast to the former community structures [15]. By uniting these two approaches, so-called regular cluster pairs with small discrepancy can be defined, where one looks for homogeneous clusters (e.g., in microarrays, one looks for groups of genes that similarly influence the same groups of conditions), see [10]. The existence of such a regular structure is theoretically guaranteed by the Abel-prize winner Szemeredi's regularity lemma [16], which, for any small positive ε, guarantees a number $k = k(\varepsilon)$ of clusters (independent of the numbers of vertices) such that by partitioning the vertices into k parts (sometimes requiring possibly a small exceptional part), the pairs of clusters have discrepancy less than ε. However, this k guaranteed by the regularity lemma can be enormously large and not applicable for practical purposes. Our goal is to give a moderate k, where the sum of the inner variances of 2^{k-1} clusters is estimated as mentioned earlier by the spectral gap between the $(k-1)$st and kth smallest positive normalized Laplacian eigenvalues, even in the worst-case scenario. This is the generalization of a theorem of [8] that was applicable only to the $k = 2$ situation. In that situation, $2^{k-1} = k$, and so the Fiedler vector was used for clustering into $k = 2$ parts. If $k > 2$, then we use $k - 1$ nontrivial eigenvectors, the ensemble of which form what we call *Fiedler-carpet*. In special cases, e.g., in the case of generalized multiclass random or quasirandom graphs, both the objective function of the k-means algorithm and the k-way discrepancy dramatically decrease compared to the values for $k - 1$ [17], where the number of clusters is one more than the number of eigenvectors used in the representation. However, in the generic case, the number of eigenvectors used for the classification is much smaller than the number of clusters, which is promising from the point of view of computational complexity.

The organization of this article is as follows. In Section 2, we define the most important notions and facts concerning normalized Laplacian spectra and multiway discrepancy of rectangular arrays with nonnegative entries [10,11,13,17,18]. In Section 3, the main theorem is stated and proved. The proof is based on the energy minimizing representation, and we analyze the structure of the vertex representatives and the underlying spectral subspaces that are mapped in a convenient way. In Appendix A, a few technical details, simulation results, and plots of the so-obtained Fiedler-carpet are presented. In Section 4, an application to the directed graph of a migration dataset is discussed. This discussion is supplemented with more figures in Appendix C. Appendix B contains pseudocode for the algorithms, which are discussed. In Section 5, the main contributions are summarized and conclusions are drawn.

2 Minimal and regular cuts versus spectra

Let $\boldsymbol{W}$ be the $n \times n$ edge-weight matrix of a graph G on n vertices. It is symmetric and has 0 diagonal and nonnegative entries. In the case of a simple graph, $\boldsymbol{W} = (w_{ij})$ is the usual adjacency matrix. The generalized degrees are $d_i = \sum_{j=1}^{n} w_{ij}$, for $i = 1, \dots, n$. Assume that d_is are all positive and the diagonal degree-matrix $\boldsymbol{D}$ contains them in its main diagonal. The *Laplacian* of G is $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, while its *normalized Laplacian* is

$$\boldsymbol{L_D} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}.$$

Because of the normalization, $\boldsymbol{L_D}$ is not affected by the scaling of the edge-weights. Therefore, we can assume that $\sum_{i=1}^{n} d_i = 1$. $\boldsymbol{L_D}$ is positive semidefinite, and if $\boldsymbol{W}$ is irreducible (G is connected), then its eigenvalues are

$$0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{n-1} \leq 2$$

with unit-norm pairwise orthogonal eigenvectors $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{n-1}$. In particular, $\mathbf{u}_0 = (\sqrt{d_1}, \dots, \sqrt{d_n})^T =: \sqrt{\mathbf{d}}^T$.

For $1 < d < n$, the row vectors of the $n \times d$ matrix $\boldsymbol{X}^* = (\boldsymbol{D}^{-1/2}\mathbf{u}_1, \dots, \boldsymbol{D}^{-1/2}\mathbf{u}_d)$ are optimal d-dimensional representatives $\mathbf{r}_1^*, \dots, \mathbf{r}_n^*$ of the vertices. More precisely, $\boldsymbol{X}^*$ minimizes the *energy* function:

$$Q_d(\boldsymbol{X}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \|\mathbf{r}_i - \mathbf{r}_j\|^2 = \operatorname{tr}(\boldsymbol{X}^T \boldsymbol{L} \boldsymbol{X}), \tag{1}$$

where the general vertex representatives $\mathbf{r}_1, \dots, \mathbf{r}_n \in \mathbb{R}^d$ are row vectors of the $n \times d$ matrix $\boldsymbol{X}$, and they are required to satisfy the constraints $\boldsymbol{X}^T \boldsymbol{D} \boldsymbol{X} = \boldsymbol{I}_d$ and $\sum_{i=1}^n d_i \mathbf{r}_i = \mathbf{0}$. Intuitively, minimizing the energy $Q_d(\boldsymbol{X})$ forces representatives of vertices connected with large edge-weights to be close to each other. The minimum of $Q_d(\boldsymbol{X})$, attained at $\boldsymbol{X}^*$, is the sum of the d smallest positive eigenvalues of $\boldsymbol{L_D}$.

The *weighted k-variance* of these representatives is defined as follows:

$$S_k^2(\boldsymbol{X}) = \min_{(V_1, \dots, V_k) \in \mathcal{P}_k} \sum_{i=1}^{k} \sum_{j \in V_i} d_j \|\mathbf{r}_j - \mathbf{c}_i\|^2, \tag{2}$$

where $\operatorname{Vol}(U) = \sum_{j \in U} d_j$, $\mathbf{c}_i = \frac{1}{\operatorname{Vol}(V_i)} \sum_{j \in V_i} d_j \mathbf{r}_j$ is the weighted center of the cluster V_i, and the minimization is over the set $\mathcal{P}_k$ of proper k-partitions $P_k = (V_1, \dots, V_k)$ of the vertex set.

We will use the weighted k-means algorithm to approach this minimum. In [19], it is stated that if the data satisfy the k-clusterable criterion ($S_k^2 \le \varepsilon^2 S_{k-1}^2$ with a small enough ε), then there is a PTAS (polynomial time approximation scheme) for the k-means problem. This is the situation we usually encounter.

It is well known that the sum of the k bottom eigenvalues of the normalized Laplacian matrix estimates from below the k-way *normalized cut* of G, which is $f_k(G) = \min_{P_k \in \mathcal{P}_k} f(P_k, G)$, where

$$f(P_k, G) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left(\frac{1}{\operatorname{Vol}(V_a)} + \frac{1}{\operatorname{Vol}(V_b)} \right) w(V_a, V_b) = \sum_{a=1}^{k} \frac{w(V_a, \overline{V}_a)}{\operatorname{Vol}(V_a)} = k - \sum_{a=1}^{k} \frac{w(V_a, V_a)}{\operatorname{Vol}(V_a)}.$$

Here, $w(V_a, V_b) = \sum_{i \in V_a} \sum_{j \in V_b} w_{ij}$ is the weighted cut between the cluster pairs. Since $\sum_{i=1}^{k-1} \lambda_i$ is the overall minimum of Q_k (on the orthogonality constraints) and $f_k(G)$ is the minimum over partition vectors (having stepwise constant coordinates over the parts of P_k), the relation

$$\sum_{i=1}^{k-1} \lambda_i \le f_k(G) \tag{3}$$

is easy to prove. This estimate is sharper if the eigensubspace spanned by the corresponding eigenvectors is closer to that of the partition vectors in the convenient k-partition of the vertices, the one produced by the *weighted k-means algorithm*. Therefore, $S_k^2(\boldsymbol{X}_{k-1}^*)$ indicates the quality of the k-clustering based on the $k-1$ bottom eigenvectors (excluding the trivial one). Later, it will be used that neither Q_k nor $S_k^2(\boldsymbol{X}_{k-1}^*)$ is affected by the orientation of the orthonormal eigenvectors.

In [8], it is proved that $S_2^2(\boldsymbol{X}_1^*) \le \frac{\lambda_1}{\lambda_2}$, so the larger the gap after the first positive eigenvalue of $\boldsymbol{L_D}$, the sharper the estimate in (3) is. Here, this statement is generalized to the gap between λ_{k-1} and λ_k, but instead of k clusters, we must consider 2^{k-1} ones, which are based on $(k-1)$-dimensional vertex representatives. This contrasts with the typical situation in the literature (see, e.g., [20–22]), where the number of clusters is the same as the number of eigenvectors used for the classification. The message of Theorem 1 of Section 3 is that the number of clusters is much higher in the generic case than the dimension of the representatives, at least in the minimum multiway cut problems. With the discrepancy objective, the famous Szemerédi's regularity lemma [16] also suggests this phenomenon.

Now we consider the discrepancy view. The theory can be better illustrated by considering rectangular arrays of nonnegative entries; adjacency matrices of simple, edge-weighted, and directed graphs are special cases of such matrices.

In many applications, for example, when microarrays are analyzed, the data are collected in the form of an $m \times n$ rectangular matrix $\boldsymbol{C} = (c_{ij})$ of nonnegative real entries. (If the entries are integer frequency counts, then the array is called contingency table in statistics.) We assume that $\boldsymbol{C}$ is *nondegenerate*, i.e., $\boldsymbol{C}\boldsymbol{C}^T$ (when $m \le n$) or $\boldsymbol{C}^T\boldsymbol{C}$ (when $m > n$) is *irreducible*. Consequently, the row-sums $d_{\text{row},i} = \sum_{j=1}^n c_{ij}$ and

column-sums $d_{\text{col},j} = \sum_{i=1}^{m} c_{ij}$ of $\boldsymbol{C}$ are strictly positive, and the diagonal matrices $\boldsymbol{D}_{\text{row}} = \text{diag}(d_{\text{row},1}, \dots, d_{\text{row},m})$ and $\boldsymbol{D}_{\text{col}} = \text{diag}(d_{\text{col},1}, \dots, d_{\text{col},n})$ are regular. Without the 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 contingency table*

$$\boldsymbol{C_D} = \boldsymbol{D}_{\text{row}}^{-1/2}\boldsymbol{C}\boldsymbol{D}_{\text{col}}^{-1/2}, \tag{4}$$

nor the *multiway discrepancies* to be introduced are affected by any uniform scaling of the entries of $\boldsymbol{C}$. It is known that the singular values of $\boldsymbol{C_D}$ are in the $[0, 1]$ interval. The positive singular values, enumerated in nonincreasing order, are the real numbers denoted by

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

where $r = \text{rank}(\boldsymbol{C}_D) = \text{rank}(\boldsymbol{C})$. Provided $\boldsymbol{C}$ is nondegenerate, 1 is a single singular value; it is called the *trivial* singular value and is denoted by s_0 since it corresponds to the trivial pair of singular vectors, which are disregarded in clustering problems. This is a well-known fact of *correspondence analysis*; for further explanation, see [13,23] and the subsequent paragraph.

For a given integer $1 \leq k \leq \min\{m, n\}$, we are looking for k-dimensional representatives $\mathbf{r}_1, \dots, \mathbf{r}_m \in \mathbb{R}^k$ of the rows and $\mathbf{q}_1, \dots, \mathbf{q}_n \in \mathbb{R}^k$ of the columns of $\boldsymbol{C}$ such that they minimize the energy function:

$$Q_k = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}\|\mathbf{r}_i - \mathbf{q}_j\|^2 \tag{5}$$

subject to

$$\sum_{i=1}^{m} d_{\text{row},i}\mathbf{r}_i\mathbf{r}_i^T = \boldsymbol{I}_k \quad \text{and} \quad \sum_{j=1}^{n} d_{\text{col},j}\mathbf{q}_j\mathbf{q}_j^T = \boldsymbol{I}_k. \tag{6}$$

When minimized, the objective function Q_k favors k-dimensional placement of the rows and columns such that representatives of columns and rows with large association (c_{ij}) are close to each other. It is easy to prove that the minimum is obtained by the singular value decomposition (SVD):

$$\boldsymbol{C_D} = \sum_{k=0}^{r-1} s_k\mathbf{v}_k\mathbf{u}_k^T, \tag{7}$$

where $r \leq \min\{n, m\}$ is the rank of $\boldsymbol{C_D}$. The constrained minimum of Q_k is $2k - \sum_{i=0}^{k-1} s_i$, and it is attained with row- and column-representatives that are row vectors of the matrices $\boldsymbol{D}_{\text{row}}^{-1/2}(\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{k-1})$ and $\boldsymbol{D}_{\text{col}}^{-1/2}(\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{k-1})$, respectively.

Note that if the entries of $\boldsymbol{C}$ are frequency counts and their sum (N) is the sample size, then the χ^2 statistic, which measures the deviation from *independence*, is

$$\chi^2 = N\sum_{i=1}^{r-1} s_i^2. \tag{8}$$

If the χ^2 test based on this statistic indicates significant deviance from independence (i.e., from the rank 1 approximation of $\boldsymbol{C}$), then one may look for the optimal rank k approximation ($1 < k < r = \text{rank}(\boldsymbol{C})$), which is constructed using the first k singular vector pairs.

When bi-clustering the rows and columns of $\boldsymbol{C}$ one may also look for subtables close to independent ones. This is measured by the *discrepancy*. In [10], the multiway discrepancy of the rectangular matrix $\boldsymbol{C}$ 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 defined as follows:

$$\begin{aligned} \text{md}(\boldsymbol{C};\, R_1, \dots, R_k, C_1, \dots, C_k) &= \max_{\substack{1\leq a,b\leq k \\ X\subset R_a, Y\subset C_b}} \frac{|c(X, Y) - \rho(R_a, C_b)\text{Vol}(X)\text{Vol}(Y)|}{\sqrt{\text{Vol}(X)\text{Vol}(Y)}} \\ &= \max_{\substack{1\leq a,b\leq k \\ X\subset R_a, Y\subset C_b}} |\rho(X, Y) - \rho(R_a, C_b)|\sqrt{\text{Vol}(X)\text{Vol}(Y)}\,, \end{aligned} \tag{9}$$

where $c(X, Y) = \sum_{i\in X}\sum_{j\in Y} c_{ij}$ is the cut between $X \subset R_a$ and $Y \subset C_b$, $\mathrm{Vol}(X) = \sum_{i\in X} d_{\mathrm{row},i}$ is the volume of the row-subset X, and $\mathrm{Vol}(Y) = \sum_{j\in Y} d_{\mathrm{col},j}$ 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 $\boldsymbol{C}$ itself is expressed as follows:

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

In [10], the following is proved:

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

provided $0 < \mathrm{md}_k(\boldsymbol{C}) < 1$. In the forward direction, the following is established in [13]. Given the $m \times n$ contingency table $\boldsymbol{C}$, 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 to the $(k-1)$-dimensional row- and column-representatives. Let $S^2_{k,\mathrm{row}}$ and $S^2_{k,\mathrm{col}}$ denote the minima of the weighted k-means algorithm, applied to the rows and columns, respectively. Then, under some balancing conditions for the margins and for the cluster sizes, $\mathrm{md}_k(\boldsymbol{C}) \le B(\sqrt{2k}(S_{k,\mathrm{row}} + S_{k,\mathrm{col}}) + s_k)$, where B is some constant that depends only on the balancing conditions, and does not depend on m and n. 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 $\boldsymbol{C}$ with small k-way discrepancy. This is the case for generalized random and quasirandom graphs [17].

This notion can be extended to an edge-weighted graph G and denoted by $\mathrm{md}_k(G)$. In that setup, $\boldsymbol{C}$ plays the role of the weighted adjacency matrix. Here, the singular values of the normalized adjacency matrix are the absolute values of the eigenvalues, which are encountered in the decreasing order.

At this point, we introduce some new matrices, originally defined by physicists (see [15]). The *modularity matrix* of an edge-weighted graph G is defined as $\boldsymbol{M} = \boldsymbol{W} - \mathbf{d}\mathbf{d}^T$, where the entries of $\boldsymbol{W}$ sum to 1. The *normalized modularity matrix* of G (see [24]) is expressed as follows:

$$\boldsymbol{M_D} = \boldsymbol{D}^{-1/2}\boldsymbol{M}\boldsymbol{D}^{-1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T = \boldsymbol{W_D} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T.$$

The normalized modularity matrix is the normalized edge-weight matrix deprived of the trivial dyad. Obviously, $\boldsymbol{L_D} = \boldsymbol{I} - \boldsymbol{W_D} = \boldsymbol{I} - \boldsymbol{M_D} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T$, see [13].

Therefore, the $k-1$ largest singular values of $\boldsymbol{M_D}$ are the absolute values of the $k-1$ largest absolute value eigenvalues of $\boldsymbol{W_D}$ (except the trivial 1). These, in turn, are 1 minus the $k-1$ positive eigenvalues of $\boldsymbol{L_D}$, which are farthest from 1. If those are all less than 1, then these are $1-\lambda_1,\dots, 1-\lambda_{k-1}$. In this case, the regularity-based spectral clustering boils down to the minimum cut objective.

Otherwise, for a $1 < k < n$ integer fixed, in the modularity-based spectral clustering, we look for the proper k-partition $V_1,\dots, V_k$ of the vertices such that the within- and between-cluster discrepancies are minimized. To motivate the introduction of the exact discrepancy measure, observe that the ij entry of $\boldsymbol{M}$ is $w_{ij} - d_id_j$, which is the difference between the actual edge-weight between the vertices i and j and the edge-weight that is expected under independent attachment of them with probabilities d_i and d_j, respectively. Consequently, the difference between the actual and the expected total edge-weight between the subsets $X, Y \subset V$ is expressed as follows:

$$\sum_{i\in X}\sum_{j\in Y}(w_{ij} - d_id_j) = w(X, Y) - \mathrm{Vol}(X)\mathrm{Vol}(Y).$$

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

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

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

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

In [10], it is proved that

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

where s_k is the kth largest nontrivial singular value of the normalized weighted adjacency matrix $\boldsymbol{W_D} = \boldsymbol{D}_{\mathrm{in}}^{-1/2}\boldsymbol{W}\boldsymbol{D}_{\mathrm{out}}^{-1/2}$. In Section 4, we apply the SVD of $\boldsymbol{W_D}$ to find migration patterns, i.e., emigration and immigration trait clusters.

3 Mapping the Fiedler-carpet: more clusters than eigenvectors

Theorem 1. *Let $G = (V, \boldsymbol{W})$ be connected edge-weighted graph with generalized degrees $d_1,\dots, d_n$ and assume that $\sum_{i=1}^n d_i = 1$. Let $0 = \lambda_0 < \lambda_1 \le\cdots\le \lambda_{n-1} \le 2$ denote the eigenvalues of the normalized Laplacian matrix $\boldsymbol{L_D}$ of G. Then, for the weighted 2^{k-1}-variance of the optimal $(k-1)$-dimensional vertex representatives, comprising row vectors of the matrix $\boldsymbol{X}_{k-1}^*$, the following upper estimate holds*:

$$S_{2^{k-1}}^2(\boldsymbol{X}_{k-1}^*) \le \frac{\sum_{j=1}^{k-1}\lambda_j}{\lambda_k},$$

provided $\lambda_{k-1} < \lambda_k$.

Proof. Recall that, with the notation of Section 2, $\boldsymbol{X}_{k-1}^* = (\boldsymbol{D}^{-1/2}\mathbf{u}_1,\dots,\boldsymbol{D}^{-1/2}\mathbf{u}_{k-1})$, where the trivial $\boldsymbol{D}^{-1/2}\mathbf{u}_0 = \mathbf{1}$ vector is disregarded, and $\mathbf{u}_0, \mathbf{u}_1,\dots \mathbf{u}_{k-1}$ are unit-norm, pairwise orthogonal eigenvectors corresponding to the eigenvalues $0 = \lambda_0 < \lambda_1 \le \cdots\le \lambda_{k-1}$ of $\boldsymbol{L}_D$, respectively. As the trivial eigenvector is disregarded, we only use the coordinates of the vectors $\mathbf{x}_j \coloneqq \boldsymbol{D}^{-1/2}\mathbf{u}_j = (x_{j1},\dots,x_{jn})^T$ for $j = 1,\dots, k-1$.

Since $\mathbf{u}_1,\dots \mathbf{u}_{k-1}$ form an orthonormal system and they are orthogonal to the $\mathbf{u}_0 = \sqrt{\mathbf{d}}$ vector, for the coordinates of $\mathbf{x}_j$, the following relations hold:

$$\sum_{i=1}^n d_i x_{ji} = 0, \quad \sum_{i=1}^n d_i x_{ji}^2 = 1\ (j = 1,\dots,k-1), \quad \sum_{i=1}^n d_i x_{ji}x_{li} = 0\ (j \ne l). \tag{10}$$

Now we will find a "witness," i.e., a vector $\mathbf{y} = (y_1,\dots,y_n)^T$ such that for it, the conditions

$$\sum_{i=1}^n d_i y_i = 0 \tag{11}$$

and

$$\sum_{i=1}^n d_i x_{ji} y_i = 0, \quad j = 1,\dots, k-1 \tag{12}$$

hold. Moreover, we will find $\mathbf{y}$ with coordinates in the following form:

$$y_i \coloneqq \sum_{j=1}^{k-1} |x_{ji} - a_j| - b, \quad i = 1,\dots, n, \tag{13}$$

where $a_1, \dots, a_{k-1}$ and b are appropriate real numbers. We will show that there exist such real numbers so that y_i's defined by them satisfy conditions (11) and (12).

Indeed, when we already have $a_1, \dots, a_{k-1}$, the aforementioned conditions together with $\sum_{i=1}^n d_i = 1$ yield

$$b = \sum_{j=1}^{k-1} b_j, \quad \text{where } b_j = \sum_{i=1}^n d_i |x_{ji} - a_j|, \quad j = 1, \dots, k-1. \tag{14}$$

With this choice of b, the fulfillment of (12) means that for $j = 1, \dots, k-1$:

$$\sum_{i=1}^n d_i x_{ji} y_i = \sum_{i=1}^n d_i x_{ji} \left(\sum_{l=1}^{k-1} |x_{li} - a_l| - b_l \right) = 0.$$

But (10) implies that

$$\sum_{i=1}^n d_i x_{ji} b_l = 0$$

for $l = 1, \dots, k-1$. This provides the following system of equations for $a_1, \dots, a_{k-1}$:

$$f_j = \sum_{i=1}^n d_i x_{ji} \sum_{l=1}^{k-1} |x_{li} - a_l| = 0, \quad j = 1, \dots, k-1. \tag{15}$$

We are looking for the root of the $f = (f_1, \dots, f_{k-1}) : \mathbb{R}^{k-1} \to \mathbb{R}^{k-1}$ function of stepwise linear coordinate functions. To prove that f has a root, we will use the multi-dimensional generalization of the Bolzano theorem: a continuous map between two normed metric spaces of the same dimensions takes a connected set into a connected one. Because of symmetry considerations, the range contains the origin, see Appendix A for further details.

Now let us define the two cluster centers for the jth coordinates by

$$c_{j1} = a_j - b_j \quad \text{and} \quad c_{j2} = a_j + b_j.$$

Observe that

$$|x_{ji} - a_j| - b_j = \begin{cases} c_{j1} - x_{ji} & \text{if } x_{ji} < a_j \\ x_{ji} - c_{j2} & \text{if } x_{ji} \geq a_j; \end{cases}$$

therefore,

$$|x_{ji} - a_j| - b_j = \min\{|x_{ji} - c_{j1}|, |x_{ji} - c_{j2}|\} \tag{16}$$

holds for $i = 1, \dots, n$; $j = 1, \dots, k-1$. For $j = 1, \dots, k-1$, they form 2^{k-1} centers in $k-1$ dimensions.

Let

$$\sigma^2(\mathbf{y}) = \sum_{i=1}^n d_i y_i^2$$

be the variance of the coordinates of $\mathbf{y}$ with respect to the discrete measure $d_1, \dots, d_n$. Due to (16), $\sigma^2(\mathbf{y}) \geq S^2_{2^{k-1}}(\boldsymbol{X}^*_{k-1})$. Define the normalized vector $\mathbf{y}$, i.e., the vector $\mathbf{z} \in \mathbb{R}^n$ of the following coordinates:

$$z_i = \frac{y_i}{\sigma(\mathbf{y})}, \quad i = 1, \dots, n;$$

obviously, $\sum_{i=1}^n d_i z_i^2 = 1$. Then,

$$\max_{i \neq m} \frac{|z_i - z_m|}{\sum_{j=1}^{k-1} |x_{ji} - x_{jm}|} \leq \frac{1}{\sigma(\mathbf{y})},$$

since, by the definition of y_i, the relation

$$|y_i - y_m| \le \sum_{j=1}^{k-1} |x_{ji} - x_{jm}| \quad (i \neq m)$$

holds.

Let $\boldsymbol{X}_k = (\boldsymbol{X}^*_{k-1}, \mathbf{z}) = (\mathbf{x}_1, \dots \mathbf{x}_{k-1}, \mathbf{z})$ be $n \times k$ matrix, containing valid k-dimensional representatives $\mathbf{r}_1, \dots, \mathbf{r}_n$ of the vertices in its rows; recall that the $n \times k$ matrix $\boldsymbol{X}^*_k = (\mathbf{x}_1, \dots \mathbf{x}_{k-1}, \boldsymbol{D}^{-1/2}\mathbf{u}_k)$ contains the optimal k-dimensional representatives in its rows. Observe that $\boldsymbol{X}_k$ and $\boldsymbol{X}^*_k$ differ only in their last columns. Let $\mathbf{r}^*_i$ denote the vector comprised of the first $k-1$ coordinates of $\mathbf{r}_i$, $i = 1, \dots, n$. These are optimal $(k-1)$-dimensional representatives of the vertices. By the optimality of the k-dimensional representation and using equation (1),

$$\begin{aligned}
\frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots \lambda_{k-1}} &= \frac{\operatorname{tr}(\boldsymbol{X}^{*T}_k \boldsymbol{L} \boldsymbol{X}^*_k)}{\operatorname{tr}(\boldsymbol{X}^{*T}_{k-1} \boldsymbol{L} \boldsymbol{X}^*_{k-1})} \\
&\le \frac{\operatorname{tr}(\boldsymbol{X}^T_k \boldsymbol{L} \boldsymbol{X}_k)}{\operatorname{tr}(\boldsymbol{X}^{*T}_{k-1} \boldsymbol{L} \boldsymbol{X}^*_{k-1})} \\
&= \frac{\sum_{i=1}^{n-1} \sum_{m=i+1}^{n} w_{im} \|\mathbf{r}_i - \mathbf{r}_m\|^2}{\sum_{i=1}^{n-1} \sum_{m=i+1}^{n} w_{im} \|\mathbf{r}^*_i - \mathbf{r}^*_m\|^2} \\
&= \frac{\sum_{i=1}^{n-1} \sum_{m=i+1}^{n} w_{im} [\|\mathbf{r}^*_i - \mathbf{r}^*_m\|^2 + (z_i - z_m)^2]}{\sum_{i=1}^{n-1} \sum_{m=i+1}^{n} w_{im} \|\mathbf{r}^*_i - \mathbf{r}^*_m\|^2} \\
&\le 1 + \max_{i \neq m} \frac{(z_i - z_m)^2}{\|\mathbf{r}^*_i - \mathbf{r}^*_m\|^2} \le 1 + \frac{1}{\sigma^2(\mathbf{y})} \le 1 + \frac{1}{S^2_{2^{k-1}}(\boldsymbol{X}^*_{k-1})},
\end{aligned}$$

which, by subtracting 1 from both the left- and right-hand sides and taking the reciprocals, finishes the proof. □

Note that only if $\lambda_{k-1} < \lambda_k$, $\mathbf{u}_k$ and $\mathbf{x}_k$ are not in the subspace spanned by $\mathbf{u}_1, \dots, \mathbf{u}_{k-1}$. Theorem 1 indicates the following clustering property of the $(k-1)$st and kth smallest normalized Laplacian eigenvalues: the greater the gap between them, the better the optimal k-dimensional representatives of the vertices can be classified into 2^{k-1} clusters.

Figure 1 shows a graph with three well-separated clusters, as well as the image of the corresponding map $f : \mathbb{R}^2 \to \mathbb{R}^2$ of the Fiedler-carpet as used in the proof of Theorem 1.

The image of f contains the origin, and we can find by inspection a pair (a_1, a_2) for which $f(a_1, a_2)$ is approximately zero; namely, choosing $a_1 = -0.19099$ and $a_2 = -0.35688$ gives $f(a_1, a_2) \approx (-0.000002, 0.00000001)$. The two-dimensional representative of vertex i is the point (x_{1i}, x_{2i}) for $i = 1, \dots n$; these are

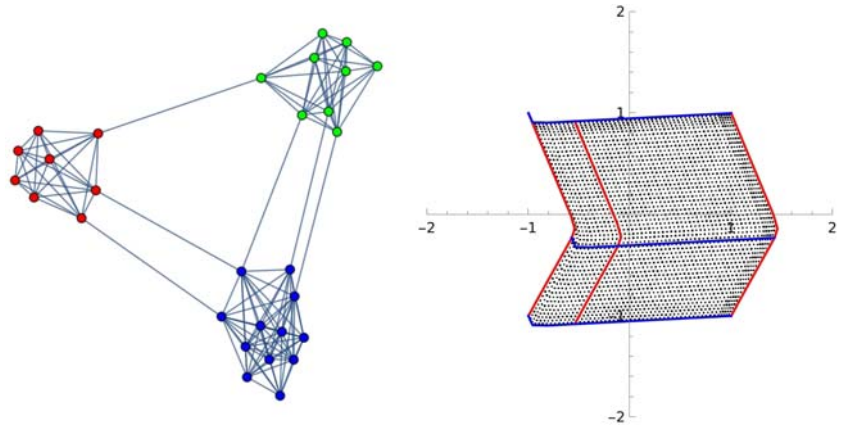

Figure 1: A graph with three well-separated clusters, and the image of its Fiedler-carpet.

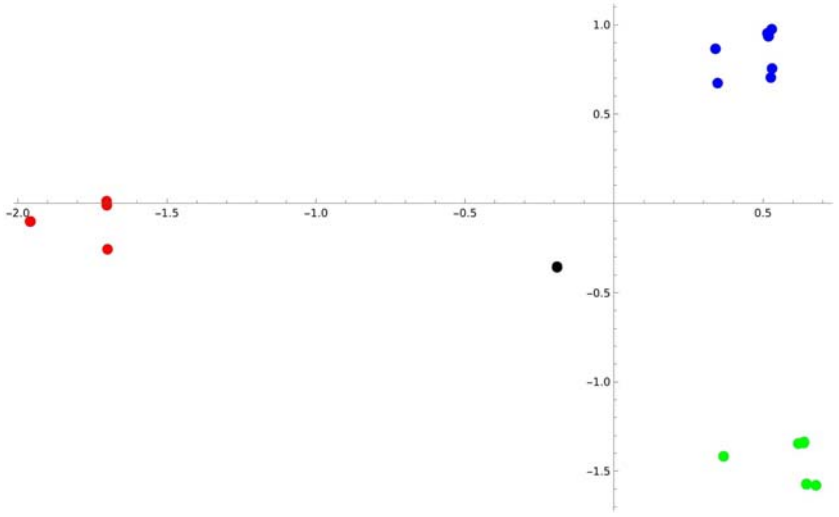

Figure 2: Two-dimensional vertex representatives of the graph of Figure 1. The black point denotes the approximate root of f.

plotted in Figure 2, where the colors indicate the cluster memberships and the black point denotes the approximate root of f.

Some remarks are in order:

- Theorem 1 is the generalization of Theorem 2.2.3 of [8]. In the $k = 2$ case, $2^{k-1} = k$, but in general, the number of clusters is much larger than that of the relevant eigenvectors.
- The statement of the theorem has relevance, since for any $k > 0$, the relation $\sum_{j=1}^{k-1}\lambda_j < (k-1)\lambda_k$ holds; but with analysis of variance considerations, $S^2_{2^{k-1}} \le k-1$ also holds. In particular, when $k = 2$, only one eigenvector is used for the representation. The total inertia of the coordinates is 1, and it can be divided into the sum of nonnegative within- and between-cluster inertias. The within-cluster inertia is the sum of the inner variances of the two clusters, which is S_2^2, and it is at most 1. (Here, the variances are calculated with respect to the discrete distribution $d_1, \dots, d_n$.) When $(k-1)$-dimensional representatives are used, then the total inertia is $\mathrm{tr}(\boldsymbol{X}_{k-1}^T \boldsymbol{D}\boldsymbol{X}_{k-1}) = \mathrm{tr}(\boldsymbol{I}_{k-1}) = k-1$, and the sum of the inner variances is again at most $k-1$, but it is further bounded with $\frac{\sum_{j=1}^{k-1}\lambda_j}{\lambda_k}$ by Theorem 1.
- The vector $\mathbf{u}_1$ is called Fiedler-vector of the (nonnormalized) Laplacian. Here, we use the ensemble of the first $k-1$ transformed eigenvectors of the normalized Laplacian together, the so-called Fiedler-carpet.

4 Applications

We investigated the international migrant stock by the country of origin and destination in the years 2015 and 2019. The focus is on 41 European countries plus the United States of America and Canada. The data[1] are based on the official registered migrants numbers, where columns and rows correspond to the country of origin and the country of destination, respectively.

Here, the quadratic, but not symmetric edge-weight matrix contains weights of bidirected edges (the diagonal is zero): the i, j entry is the number of persons going $j \to i$. Via SVD of the normalized table,

1 United Nations, Department of Economic and Social Affairs, Population Division, International Migrant Stock 2019 (https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp).

we can find emigration (column) and immigration (row) clusters, between which the migration is the best homogeneous (in terms of discrepancy).

Both for the 2015 and 2019 data, there was a gap after four nontrivial singular values, and therefore, the corresponding four singular vector pairs were used to find five emigration and immigration trait clusters.

- Singular values, 2015:
  1, 0.79098, 0.71857, 0.67213, 0.56862, 0.45293, 0.40896, 0.38178, 0.36325, 0.34785, 0.32648, 0.31769, 0.2996, 0.27927, 0.26566, 0.24718, 0.22638, 0.20632, 0.18349, 0.1651, 0.14384, 0.1359, 0.12721, 0.12092, 0.11816, 0.10374, 0.09545, 0.08278, 0.0738, 0.06371, 0.05673, 0.04553, 0.03488, 0.03107, 0.02967, 0.02693, 0.01557, 0.00788, 0.00584, 0.00519, 0.00191, 0.0017, 0.00099.
- Singular values, 2019:
  1, 0.77844, 0.70989, 0.65059, 0.55122, 0.43612, 0.39512, 0.36194, 0.3558, 0.33882, 0.32174, 0.30719, 0.29601, 0.28181, 0.26865, 0.259, 0.22421, 0.1917, 0.17988, 0.1516, 0.13671, 0.13243, 0.12397, 0.11542, 0.10598, 0.09216, 0.08889, 0.07958, 0.06835, 0.06154, 0.05377, 0.04412, 0.03436, 0.03124, 0.02899, 0.02745, 0.01507, 0.00814, 0.00619, 0.0051, 0.00216, 0.00129, 0.00089.

In the discrepancy-based spectral clustering, the number of clusters is one more than the number of structural singular values (excluding the trivial 1). In our case, the number of clusters was five. The five emigration and immigration trait clusters for 2015 are shown in Tables 1 and 2, whereas those for 2019 are shown in Tables 3 and 4. Figure 3(a) and (b) illustrate the immigration–emigration cluster-pairs with the countries rearranged by their cluster memberships. The frequency counts are represented with light to dark squares.

In 2015, by χ^2 test, we found the smallest discrepancy between the emigration trait cluster number 2 and immigration trait cluster number 4, i.e., the subtable formed by them was close to an independent table of rank 1. The clusters were similar in the 2 years; in 2019, the smallest discrepancy was found between the emigration trait cluster number 2 and immigration trait cluster number 5, i.e., between the Balcanian and

Table 1: Country memberships of emigration trait clusters, 2015

Cluster #	Emigration countries
1	Austria, Belgium, Bulgaria, Canada, Czechia, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland, United Kingdom, United States of America
2	Bosnia and Herzegovina, Croatia, Montenegro, Slovenia
3	Albania
4	Belarus, Estonia, Republic of Moldova, Ukraine
5	Russian Federation

Table 2: Country memberships of immigration trait clusters, 2015

Cluster #	Immigration countries
1	Albania, Austria, Belgium, Bulgaria, Canada, Czechia, Denmark, Finland, France, Germany, Hungary, Iceland, Ireland, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, Romania, Slovakia, Spain, Sweden, Switzerland, United Kingdom, United States of America
2	Greece, Italy, North Macedonia
3	Bosnia and Herzegovina, Croatia, Montenegro, Serbia, Slovenia
4	Belarus, Estonia, Latvia, Lithuania, Ukraine
5	Poland, Republic of Moldova, Russian Federation

Table 3: Country memberships of emigration trait clusters, 2019

Cluster #	Emigration countries
1	Austria, Belgium, Bulgaria, Canada, Czechia, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Monaco, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland, United Kingdom, United States of America
2	Bosnia and Herzegovina, Croatia, Montenegro, Slovenia
3	Albania
4	Belarus, Estonia, Republic of Moldova, Ukraine
5	Russian Federation

Table 4: Country memberships of immigration trait clusters, 2019

Cluster #	Immigration countries
1	Albania, Austria, Belgium, Bulgaria, Canada, Czechia, Denmark, Finland, France, Germany, Hungary, Iceland, Ireland, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, Romania, Slovakia, Spain, Sweden, Switzerland, United Kingdom, United States of America
2	Bosnia and Herzegovina, Croatia, Montenegro, Serbia, Slovenia
3	Greece, Italy, North Macedonia
4	Poland, Republic of Moldova, Russian Federation
5	Belarus, Estonia, Latvia, Lithuania, Ukraine

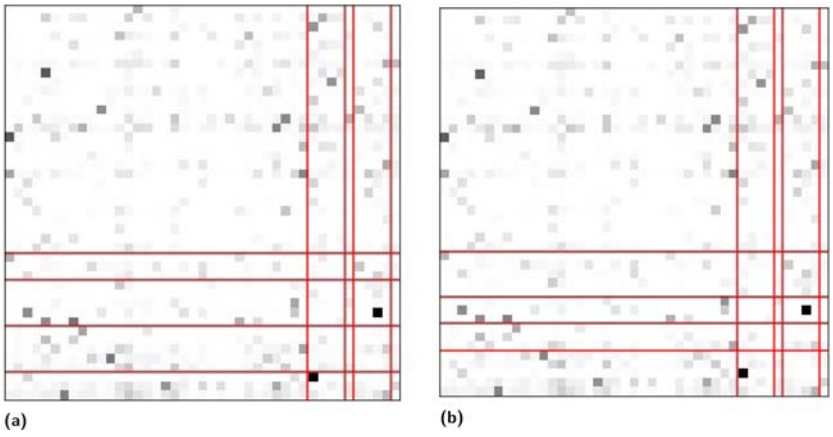

Figure 3: Immigration–emigration cluster-pairs with the countries rearranged by their cluster memberships. The frequency counts are represented with light to dark squares. (a) 2015 and (b) 2019.

Baltic countries in both years. Correspondence analysis results with cluster memberships are found in Appendix C.

Here, we wanted to find as homogeneous cluster pairs as possible, as for migration patterns vis-a-vis the emigration–immigration. Our graph was a relatively small and sparse one, so Figure 3(a) and (b) show a large and some small clusters for both emigration and immigration. However, the emigration–immigration cluster pair with the smallest discrepancy is spotted by the χ^2 test. In larger and more dense networks, e.g., in metabolic networks with bidirected edges between the vertices (enzymes), the weights representing the intensity of chemical reactions, so-called autocatalytic subnetworks could be found with our method, based on discrepancies.

5 Summary and author contributions

Spectral clustering is a collection of methods for clustering data points or vertices of a graph based on combinatorial criteria with spectral relaxation. Here, we generalize spectral clustering in several ways:

- Instead of simple or edge-weighed graphs, we consider directed graphs and rectangular arrays of non-negative entries. For these, the method of correspondence analysis gives low-dimensional representation of the row and column items by means of SVD of the normalized table. This is applied to real-life data.
- Instead of the usual multiway minimum cut objective, we consider discrepancy minimization, for which a near optimal solution is given by the k-means algorithm applied to the row and column representatives, where the number of clusters (k) is concluded from the gap in the singular spectrum. There are theoretical results supporting this. Eventually, the near optimal clusters can be refined to decrease discrepancy.
- The number of clusters can be larger than the number of eigenvectors entered in the classification. The two are the same only in case of generalized random or quasirandom graphs, see [17]. In Theorem 1, an exact estimate for the sum of the inner variances of 2^{k-1} clusters is given by means of the spectral gap between the $(k-1)$st and kth smallest normalized Laplacian eigenvalues by using the Fiedler-carpet of the corresponding eigenvectors.
- When the number of clusters is the same as the number of eigenvectors entered in the classification, Davis-Kahan type subspace perturbation theorems are applicable, as in this case, the k-partition vectors span a k-dimensional subspace. However, if the number of parts is larger than that of the eigenvectors, then finer methods are applicable, like our construction in the proof, using a "witness."

Acknowledgments: The research was done under the auspices of the Budapest Semesters in Mathematics program, in the framework of an undergraduate online research course in the fall semester 2021, with the participation of US undergraduate students. Also, Fatma Abdelkhalek's work is related to a scholarship under the Stipendium Hungaricum program between Egypt and Hungary; further, Vilas Winstein's work is related to the DYNASNET (Dynamic and Structure in Networks) Project of the Rényi Institute, Budapest. The paper is dedicated to Gábor Tusnády for his 80th birthday, with whom the first author of this paper posed a conjecture in [8]. Here, the conjecture is proved in the $k > 2$ case, albeit with much more clusters than the number of the eigenvalues preceding the spectral gap.

Funding information: The authors acknowledge the contribution of Pro Mathematica and Arte and Fulbright Hungary organizations for the generous donation of funding for the open access fees.

Conflict of interest: Authors state no conflict of interest.

Data Availability Statement: The third-party dataset analyzed during the current study is based on the official registered migrants numbers of the United Nations, Department of Economic and social Affairs, Population Division, International Migrant Stock 2019. Availability of the data in https://www.un.org/en/development/desa/population/migration/data/estimates2/estimates19.asp.

References

[1] A. J. Hoffman, *Eigenvalues and partitionings of the edges of a graph*, Linear Algebra Appl. **5** (1972), 137–146.
[2] M. Fiedler, *Algebraic connectivity of graphs*, Czech. Math. J. **23** (1973), 298–305.
[3] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs*, Academic Press, New York, 1979.
[4] F. Chung, *Spectral graph theory*, In: CBMS Regional Conference Series in Mathematics, vol. 92, American Mathematical Society, Providence RI, 1997.

[5] C. J. Alpert and S.-Z. Yao, *Spectral partitioning: the more eigenvectors, the better*, In: B. T. Preas, P. G. Karger, B. S. Nobandegani, M. Pedram, Eds., Proceedings on 32nd ACM/IEEE International Conference on Design Automation, 1995, Association of Computer Machinery, New York, pp. 195–200.
[6] B. Mohar, *Isoperimetric inequalities, growth and the spectrum of graphs*, Linear Algebra Appl. **103** (1988), 119–131.
[7] S. Hoory, N. Linial, and A. Widgerson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N. S.) **43** (2006), 439–561.
[8] M. Bolla and G. Tusnády, *Spectra and optimal partitions of weighted graphs*, Discret. Math. **128** (1994), 1–20.
[9] M. Bolla, K. Friedl, and A. Krámli, *Singular value decomposition of large random matrices* (*for two-way classification of microarrays*), J. Multivariate Anal. **101** (2010), 434–446.
[10] M. Bolla, *Relating multiway discrepancy and singular values of nonnegative rectangular matrices*, Discrete Appl. Math. **203** (2016), 26–34.
[11] B. Bollobás and V. Nikiforov, *Hermitian matrices and graphs: singular values and discrepancy*, Discrete Math. **285** (2004), 17–32.
[12] Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein, *Spectral biclustering of microarray data: coclustering genes and conditions*, Genome Res. **13** (2003), 703–716.
[13] M. Bolla, *Spectral Clustering and Biclusterig*, Wiley, Chichester, UK, 2013.
[14] F. Chung and R. Graham, *Quasi-random graphs with given degree sequences*, Random Struct. Algorithms **12** (2008), 1–19.
[15] M. E. J. Newman, *Networks, An Introduction*, Oxford University Press, Oxford, United Kingdom, 2010.
[16] 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, CNRS, Paris, 1976, pp. 399–401.
[17] M. Bolla, *Generalized quasirandom properties of expanding graph sequences*, Nonlinearity **33** (2020), no. 4, 1405–1424.
[18] N. Alon, A. Coja-Oghlan, H. Han, M. Kang, V. Rödl, and M. Schacht, *Quasi-randomness and algorithmic regularity for graphs with general degree distributions*, Siam J. Comput. **39** (2010), no. 6, 2336–2362.
[19] R. Ostrovsky, Y. Rabani, L. J. Schulman, and C. Swamy, *The effectiveness of Lloyd-type methods for the k-means problem*, J. ACM **59** (2012), no. 6, Article 28.
[20] U. VonLuxburg, *A tutorial on spectral clustering*, Stat. Comput. **17** (2006), 395–416.
[21] A. Y. Ng, M. I. Jordan, and Y. Weiss, *On spectral clustering: analysis and an algorithm*, In: T. G. Dietterich, S. Becker, Z. Ghahramani, Eds., Proceedings on 14th Neural Information Processing Systems Conference (NIPS 2001), MIT Press, Cambridge, USA, 2010.
[22] J. Shi and J. Malik, *Normalized cuts and image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000), 888–905.
[23] K. Faust, *Using correspondence analysis for joint displays of affiliation networks*, In: P. J. Carrington, J. Scott, Eds, Models and Methods in Social Network Analysis, Vol. 7, Cambridge University Press, Cambridge, 2005, pp. 117–147.
[24] M. Bolla, *Penalized versions of the Newman-Girvan modularity and their relation to multi-way cuts and k-means clustering*, Phys. Rev. E **84** (2011), 016108.

Appendix A

In the $f:\mathbb{R}\to\mathbb{R}$ case: $f=f_1$, $a=a_1$ and with the notation $A=\min_i x_{1i}$, $B=\max_i x_{1i}$ for

$$f(a)=\sum_{i=1}^{n} d_i x_{1i}|x_{1i}-a|$$

we have that $f(A)=1$ and $f(B)=-1$. As f is continuous, it must have a root in (A,B), by the Bolzano theorem. Also note that this root of f is around the median of the coordinates of $\mathbf{x}_1$ with respect to the discrete measure $d_1,\dots,d_n$.

Then consider the $f=(f_1,f_2):\mathbb{R}^2\to\mathbb{R}^2$ case. The coordinate functions are stepwise linear, continuous functions. The system of equations is expressed as follows:

$$\begin{aligned} f_1(a_1,a_2)&=\sum_{i=1}^{n} d_i x_{1i}|x_{1i}-a_1|+\sum_{i=1}^{n} d_i x_{1i}|x_{2i}-a_2|=0,\\ f_2(a_1,a_2)&=\sum_{i=1}^{n} d_i x_{2i}|x_{1i}-a_1|+\sum_{i=1}^{n} d_i x_{2i}|x_{2i}-a_2|=0. \end{aligned} \tag{A1}$$

With the notation $A=\min_i x_{1i}$, $B=\max_i x_{1i}$, $C=\min_i x_{2i}$, $D=\max_i x_{2i}$, where $A<0$, $B>0$, $C<0$, $D>0$,

$$f(A,C)=(1,1),\quad f(B,D)=(-1,-1),\quad f(A,D)=(1,-1),\quad f(B,C)=(-1,1).$$

Furthermore, $f(x,y)=(1,1)$ if $x\le A$, $y\le C$; $f(x,y)=(-1,-1)$ if $x\ge B$, $y\ge D$; $f(x,y)=(1,-1)$ if $x\le A$, $y\ge D$; and $f(B,C)=(-1,1)$ if $x\ge B$, $y\le C$.

We want to show that f has a root within the rectangle $[A,B]\times[C,D]$. By the multivariate version of the Bolzano theorem, the f-map of this rectangle is a connected region in $\mathbb{R}^2$ that contains the points $(1,1)$, $(-1,-1)$, $(1,-1)$, $(-1,1)$ as "corners." We will show that it contains $(0,0)$ too.

Note that together with $\mathbf{u}_j$, the vector $-\mathbf{u}_j$ is also a unit-norm eigenvector, so instead of $\mathbf{x}_j$, we can as well use $-\mathbf{x}_j$ for $j=1,2$, which gives four possible domains of f: next to the rectangle $[A,B]\times[C,D]$, the rectangles $[-B,-A]\times[C,D]$, $[A,B]\times[-D,-C]$, and $[-B,-A]\times[-D,-C]$ are also closed, bounded regions, and the f-images of them show symmetry with respect to the coordinate axes. Therefore, it suffices to prove that the map of the union of them contains the origin. In Section 2, we saw that neither the objective function (Q_k), nor the clustering is affected by the orientation of the eigenvectors, so the orientation is not denoted in the sequel. Also notice that with counter-orienting $\mathbf{u}_1$ or/and $\mathbf{u}_2$: if (a_1,a_2) is a root of f, then $-a_1$ instead of a_1 and/or $-a_2$ instead of a_2 will result in a root of f too.

The images are closed, bounded regions (usually not rectangles), but we will show that the opposite sides of them are parallel curves and sandwich the f_1 and f_2 axes, respectively. As the f-values sweep the region between these boundaries, the total range should contain the origin. Now the above, below, right, and left boundaries are investigated.

- **Above:** Consider the boundary curve between $(-1,1)$ and $(1,1)$. Along that, $a_2=C$ and $A<a_1<B$. Let $H:=\{i: x_{1i}>a_1\}$. Then, $H\neq\varnothing$ and $\bar{H}\neq\varnothing$; further,

$$\begin{aligned} f_2(a_1,C)&=\sum_{i=1}^{n} d_i x_{2i}|x_{1i}-a_1|+\sum_{i=1}^{n} d_i x_{2i}(x_{2i}-C)\\ &=\sum_{i\in H} d_i x_{2i}(x_{1i}-a_1)-\sum_{i\in\bar{H}} d_i x_{2i}(x_{1i}-a_1)+1\\ &=2\sum_{i\in H} d_i x_{2i}(x_{1i}-a_1)+1\\ &=2\sum_{i\in\bar{H}} d_i x_{2i}(a_1-x_{1i})+1, \end{aligned} \tag{A2}$$

  where we intensively used conditions (10).
- **Below:** Consider the boundary curve between $(-1,-1)$ and $(1,-1)$. Along that, $a_2=D$ and $A<a_1<B$. Then,

$$\begin{aligned} f_2(a_1, D) &= \sum_{i=1}^{n} d_i x_{2i} |x_{1i} - a_1| - \sum_{i=1}^{n} d_i x_{2i}(x_{2i} - D) \\ &= \sum_{i\in H} d_i x_{2i}(x_{1i} - a_1) - \sum_{i\in \bar{H}} d_i x_{2i}(x_{1i} - a_1) - 1 \\ &= 2\sum_{i\in H} d_i x_{2i}(x_{1i} - a_1) - 1 \\ &= 2\sum_{i\in \bar{H}} d_i x_{2i}(a_1 - x_{1i}) - 1. \end{aligned} \tag{A3}$$

– **Between (horizontally):** Consider the case when $a_2 = u \in (C, D)$ fixed and $A < a_1 < B$. Then

$$\begin{aligned} f_2(a_1, u) &= \sum_{i=1}^{n} d_i x_{2i} |x_{1i} - a_1| + \sum_{i=1}^{n} d_i x_{2i} |x_{2i} - u| \\ &= \sum_{i\in H} d_i x_{2i}(x_{1i} - a_1) - \sum_{i\in \bar{H}} d_i x_{2i}(x_{1i} - a_1) + \sum_{i:x_{2i}>u} d_i x_{2i}(x_{2i} - u) + \sum_{i:x_{2i}\le u} d_i x_{2i}(u - x_{2i}) \\ &= 2\sum_{i\in H} d_i x_{2i}(x_{1i} - a_1) - 1 + 2\sum_{i:x_{2i}>u} d_i x_{2i}^2 - 2u \sum_{i:x_{2i}>u} d_i x_{2i}. \end{aligned}$$

So the $f_2(a_1, u)$ arcs are all parallel to the boundary curves $f_2(a_1, C)$ and $f_2(a_1, D)$ and to each other. In particular,

$$f_2(a_1, 0) = 2\sum_{i\in H} d_i x_{2i}(x_{1i} - a_1) - 1 + 2\sum_{i:x_{2i}>0} d_i x_{2i}^2.$$

This arc is either closer to the above or the below curve (which are in distance 2 from each other), depending on whether $\sum_{i:x_{2i}>0} d_i x_{2i}^2$ is less or greater than $\frac{1}{2}$, but is strictly positive by condition (10). Hence, if u and u' are "close" to 0, then the $f_2(a_1, u)$ and $f_2(a_1, u')$ arcs are not identical, otherwise it can happen that for some $u \neq u'$:

$$\sum_{i:x_{2i}>u} d_i x_{2i}^2 - u \sum_{i:x_{2i}>u} d_i x_{2i} = \sum_{i:x_{2i}>u'} d_i x_{2i}^2 - u' \sum_{i:x_{2i}>u'} d_i x_{2i}. \tag{A4}$$

The same is true vertically.

– **Right:** Consider the boundary curve between (1, –1) and (1, 1). Along that, $a_1 = A$ and $C < a_2 < D$. Let $F := \{i : \ x_{2i} > a_2\}$. Then, $F \neq \emptyset$ and $\bar{F} \neq \emptyset$; further,

$$\begin{aligned} f_1(A, a_2) &= \sum_{i=1}^{n} d_i x_{1i}(x_{1i} - A) - \sum_{i=1}^{n} d_i x_{1i} |x_{2i} - a_2| \\ &= 1 + \sum_{i\in F} d_i x_{1i}(x_{2i} - a_2) - \sum_{i\in \bar{F}} d_i x_{1i}(x_{2i} - a_2) \\ &= 1 + 2\sum_{i\in F} d_i x_{1i}(x_{2i} - a_2) \\ &= 1 + 2\sum_{i\in \bar{F}} d_i x_{1i}(a_2 - x_{2i}). \end{aligned} \tag{A5}$$

– **Left:** As from the left, consider the boundary curve between (–1, –1) and (–1, 1). Along that, $a_1 = B$ and $C < a_2 < D$. Then,

$$f_1(B, a_2) = -1 + 2\sum_{i\in F} d_i x_{1i}(x_{2i} - a_2) = -1 + 2\sum_{i\in \bar{F}} d_i x_{1i}(a_2 - x_{2i}). \tag{A6}$$

– **Between (vertically):** Consider the case when $a_1 = v \in (A, B)$ fixed and $C < a_2 < D$. Then,

$$f_1(v, a_2) = 1 + 2\sum_{i\in F} d_i x_{1i}(x_{2i} - a_2) - 1 + 2\sum_{i:x_{1i}>v} d_i x_{1i}^2 - 2v \sum_{i:x_{1i}>v} d_i x_{1i}.$$

So the $f_1(v, a_2)$ arcs are all parallel to the boundary curves $f_1(A, a_2)$ and $f_1(B, a_2)$ and to each other. In particular,

$$f_1(0, a_2) = 2\sum_{i\in F} d_i x_{1i}(x_{2i} - a_2) - 1 + 2\sum_{i:x_{1i}>0} d_i x_{1i}^2.$$

This arc is either closer to the right or the left curve (which are in distance 2 from each other), depending on whether $\sum_{i:x_{1i}>0} d_i x_{1i}^2$ is less or greater than $\frac{1}{2}$, but is strictly positive by condition (10). For this reason, if v and v' are "close" to 0, then the $f_1(v, a_2)$ and $f_1(v', a_2)$ arcs are not identical, otherwise it can happen that for some $v \neq v'$:

$$\sum_{i:x_{1i}>v} d_i x_{1i}^2 - v\sum_{i:x_{1i}>v} d_i x_{1i} = \sum_{i:x_{1i}>v'} d_i x_{1i}^2 - v'\sum_{i:x_{1i}>v'} d_i x_{1i}. \tag{A7}$$

Therefore, any grid on the rectangle of the domain (its horizontal and vertical lines parallel to the a_1 and a_2 axes) is mapped by f onto a lattice with horizontal and vertical, parallel arcs. This proves that f is one-to-one whenever these arcs are not identical. The possible inconvenient phenomenon, when both equations (A4) and (A7) hold for some $u \neq u'$ and $v \neq v'$ pairs, is experienced at the dark parts of Figure A1 near the boundaries. However, f is injective in the neighborhood of the origin that does not contain any of the finitely many eigenvector coordinates (because there it has purely linear coordinate functions). This also depends on the underlying graph: if it shows symmetries, then its weighted Laplacian has multiple eigenvalues and/or multiple coordinates of the eigenvectors that may cause complications.

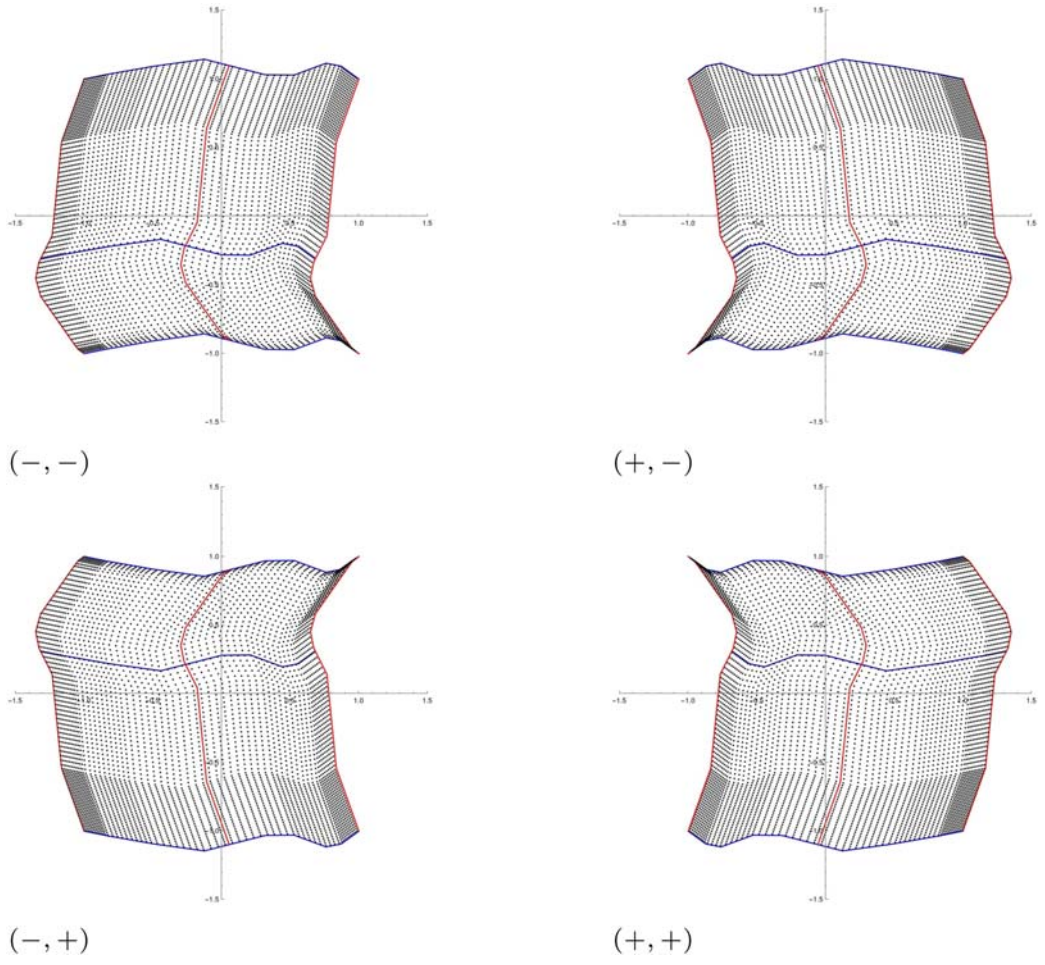

Figure A1: Images of the four possible orientations of the three-dimensional Fiedler-carpet of the graph in Figure A2.

To prove that the above–below boundaries sandwich the f_1 axis and the right–left boundaries sandwich the f_2 axis, the following investigations are made. Because the investigations are of similar vein, only the first of them will be discussed in details. We distinguish between eight cases (denoted by an acronym), depending on, where half of boundary is considered. The estimates are supported by specific orientations of the unit norm eigenvectors. If $\mathbf{u}_1$ is oriented so that for the coordinates of $\mathbf{x}_1 = \boldsymbol{D}^{-1/2}\mathbf{u}_1$

$$\sum_{i:x_{1i}>0} d_i x_{1i}^2 < \frac{1}{2}$$

holds, then it is called *positive orientation*, whereas the opposite is negative. Likewise, the orientation of $\mathbf{u}_2$ is positive if for the coordinates of of $\mathbf{x}_2 = \boldsymbol{D}^{-1/2}\mathbf{u}_2$

$$\sum_{i:x_{2i}>0} d_i x_{2i}^2 < \frac{1}{2} \tag{A8}$$

holds, otherwise it is negative.

AL (Above boundary, left half), when $a_1 > 0$: using the last but one line of equation (A2),

$$\begin{aligned} f_2(a_1, C) &= 2\sum_{i\in H} d_i x_{2i}(x_{1i} - a_1) + 1 \\ &= 2\sum_{i\in H, x_{2i}>0} d_i x_{2i}(x_{1i} - a_1) + 2\sum_{i\in H, x_{2i}\le 0} d_i x_{2i}(x_{1i} - a_1) + 1 \\ &\ge 2\sum_{i\in H, x_{2i}\le 0} d_i x_{2i}(x_{1i} - a_1) + 1. \end{aligned}$$

To prove that $f_2(a_1, C) \ge 0$, it suffices to prove that

$$\sum_{i\in H, x_{2i}\le 0} d_i x_{2i}(x_{1i} - a_1) \ge -\frac{1}{2},$$

which is equivalent to

$$\sum_{i\in H, x_{2i}\le 0} d_i(-x_{2i})(x_{1i} - a_1) \le \frac{1}{2}.$$

We use the Cauchy-Schwarz inequality by keeping in mind that $x_{1i} - a_1 > 0$ ($i \in H$) and because of $a_1 > 0$, $x_{1i} - a_1 < x_{1i}$. Therefore,

$$\left[\sum_{i\in H x_{2i}\le 0} (\sqrt{d_i}(-x_{2i}))(\sqrt{d_i}(x_{1i} - a_1))\right]^2 \le \left[\sum_{i\in H\ x_{2i}\le 0} d_i x_{2i}^2\right]\left[\sum_{i\in H\ x_{2i}\le 0} d_i x_{1i}^2\right] \le \frac{1}{2}\frac{1}{2}$$

holds true if $\mathbf{u}_1$ is positively and $\mathbf{u}_2$ is negatively oriented.

AR (Above boundary, right half), when $a_1 \le 0$: using the last line of equation (A2), to prove that $f_2(a_1, C) \ge 0$ it suffices to prove that

$$\sum_{i\in \bar{H}, x_{2i}\le 0} d_i x_{2i}(a_1 - x_{1i}) \ge -\frac{1}{2},$$

which is equivalent to

$$\sum_{i\in \bar{H}, x_{2i}\le 0} d_i(-x_{2i})(a_1 - x_{1i}) \le \frac{1}{2}.$$

We again use the Cauchy-Schwarz inequality by keeping in mind that $a_1 - x_{1i} \ge 0$ ($i \in \bar{H}$) and $a_1 - x_{1i} = -x_{i1} - (-a_1) \le -x_{1i}$ as now $-a_1 \ge 0$ and $-x_{1i} > -a_1$. Therefore,

$$\left[\sum_{i\in \bar{H}, x_{2i}\le 0} (\sqrt{d_i}(-x_{2i}))(\sqrt{d_i}(a_1 - x_{1i}))\right]^2 \le \left[\sum_{i\in \bar{H}, x_{2i}\le 0} d_i x_{2i}^2\right]\left[\sum_{i\in \bar{H}, x_{2i}\le 0} d_i(-x_{1i})^2\right] \le \frac{1}{2}\frac{1}{2}$$

holds true with negatively orienting $\mathbf{u}_1$ and negatively $\mathbf{u}_2$.

BL (Below boundary, left half), when $a_1 > 0$: using the last but one line of equation (A3),

$$\begin{aligned} f_2(a_1, D) &= 2 \sum_{i\in H} d_i x_{2i}(x_{1i} - a_1) - 1 \\ &= 2 \sum_{i\in H, x_{2i}<0} d_i x_{2i}(x_{1i} - a_1) + 2 \sum_{i\in H, x_{2i}\geq 0} d_i x_{2i}(x_{1i} - a_1) - 1 \\ &\leq 2 \sum_{i\in H, x_{2i}\geq 0} d_i x_{2i}(x_{1i} - a_1) - 1. \end{aligned}$$

To prove that $f_2(a_1, D) \leq 0$, it suffices to prove that

$$\sum_{i\in H, x_{2i}\geq 0} d_i x_{2i}(x_{1i} - a_1) \leq \frac{1}{2}.$$

We use the Cauchy-Schwarz inequality by keeping in mind that $x_{1i} - a_1 > 0$ ($i \in H$) and because of $a_1 > 0$, $x_{1i} - a_1 < x_{1i}$. Therefore,

$$\left[\sum_{i\in H\ x_{2i}\geq 0} (\sqrt{d_i} x_{2i})(\sqrt{d_i}(x_{1i} - a_1))\right]^2 \leq \left[\sum_{i\in H\ x_{2i}\geq 0} d_i x_{2i}^2\right]\left[\sum_{i\in H\ x_{2i}\geq 0} d_i x_{1i}^2\right] \leq \frac{1}{2}\frac{1}{2}$$

holds true if $\mathbf{u}_1$ is positively and $\mathbf{u}_2$ is positively oriented.

BR (Below boundary, right half), when $a_1 \leq 0$: using the last line of equation (A3), to prove that $f_2(a_1, D) \leq 0$, it suffices to prove that

$$\sum_{i\in \bar{H}, x_{2i}\geq 0} d_i x_{2i}(a_1 - x_{1i}) \leq \frac{1}{2}.$$

We again use the Cauchy-Schwarz inequality by keeping in mind that $a_1 - x_{1i} \geq 0 (i \in \bar{H})$ and $a_1 - x_{1i} = -x_{i1} - (-a_1) \leq -x_{1i}$ as now $-a_1 \geq 0$ and $-x_{1i} > -a_1$. Therefore,

$$\left[\sum_{i\in \bar{H}, x_{2i}\geq 0} (\sqrt{d_i} x_{2i})(\sqrt{d_i}(a_1 - x_{1i}))\right]^2 \leq \left[\sum_{i\in \bar{H}, x_{2i}\geq 0} d_i x_{2i}^2\right]\left[\sum_{i\in \bar{H}, x_{2i}\geq 0} d_i(-x_{1i})^2\right] \leq \frac{1}{2}\frac{1}{2}$$

holds true with negatively orienting $\mathbf{u}_1$ and positively $\mathbf{u}_2$.

RB (Right boundary, below half), when $a_2 > 0$: using the last but one line of equation (A5),

$$f_1(A, a_2) \geq 2 \sum_{i\in F, x_{1i}\leq 0} d_i x_{1i}(x_{2i} - a_2) + 1.$$

To prove that $f_1(A, a_2) \geq 0$, it suffices to prove that

$$\sum_{i\in F, x_{1i}\leq 0} d_i x_{1i}(x_{2i} - a_2) \geq -\frac{1}{2},$$

which is equivalent to

$$\sum_{i\in F, x_{1i}\leq 0} d_i(-x_{1i})(x_{2i} - a_2) \leq \frac{1}{2}.$$

By the Cauchy-Schwarz inequality,

$$\left[\sum_{i\in F\ x_{1i}\leq 0} (\sqrt{d_i}(-x_{1i}))(\sqrt{d_i}(x_{2i} - a_2))\right]^2 \leq \left[\sum_{i\in F\ x_{1i}\leq 0} d_i x_{1i}^2\right]\left[\sum_{i\in F\ x_{1i}\leq 0} d_i x_{2i}^2\right] \leq \frac{1}{2}\frac{1}{2}$$

holds true if $\mathbf{u}_1$ is negatively and $\mathbf{u}_2$ is positively oriented.

RA (Right boundary, above half), when $a_2 \leq 0$: using the last line of equation (A5), to prove that $f_1(A, a_2) \geq 0$, it suffices to prove that

$$\sum_{i\in\bar{F},x_{1i}\le 0} d_i x_{1i}(a_2 - x_{2i}) \ge -\frac{1}{2},$$

which is equivalent to

$$\sum_{i\in\bar{F},x_{1i}\le 0} d_i(-x_{1i})(a_2 - x_{2i}) \le \frac{1}{2}.$$

By the Cauchy-Schwarz inequality,

$$\left[\sum_{i\in\bar{F},x_{1i}\le 0} (\sqrt{d_i}(-x_{1i}))(\sqrt{d_i}(a_2 - x_{2i}))\right]^2 \le \left[\sum_{i\in\bar{F},x_{1i}\le 0} d_i x_{1i}^2\right]\left[\sum_{i\in\bar{F},x_{1i}\le 0} d_i(-x_{2i})^2\right] \le \frac{1}{2}\frac{1}{2}$$

holds true with negatively orienting $\mathbf{u}_1$ and negatively $\mathbf{u}_2$.

LB (Left boundary, below half), when $a_2 > 0$: using the last but one line of equation (A6),

$$f_1(B, a_2) = 2\sum_{i\in F} d_i x_{1i}(x_{2i} - a_2) - 1 \le 2\sum_{i\in F,x_{1i}\ge 0} d_i x_{1i}(x_{2i} - a_2) - 1.$$

To prove that $f_1(B, a_2) \le 0$ it suffices to prove that

$$\sum_{i\in F,x_{1i}\ge 0} d_i x_{1i}(x_{2i} - a_2) \le \frac{1}{2},$$

By the Cauchy-Schwarz inequality,

$$\left[\sum_{i\in F\ x_{1i}\ge 0} (\sqrt{d_i}x_{1i})(\sqrt{d_i}(x_{2i} - a_2))\right]^2 \le \left[\sum_{i\in F\ x_{1i}\ge 0} d_i x_{1i}^2\right]\left[\sum_{i\in F\ x_{1i}\ge 0} d_i x_{2i}^2\right] \le \frac{1}{2}\frac{1}{2}$$

holds true if $\mathbf{u}_1$ is positively and $\mathbf{u}_2$ is positively oriented.

LA (Left boundary, above half), when $a_2 \le 0$: using the last line of equation (A6), to prove that $f_1(B, a_2) \le 0$ it suffices to prove that

$$\sum_{i\in\bar{F},x_{1i}\ge 0} d_i x_{1i}(a_2 - x_{2i}) \le \frac{1}{2}.$$

By the Cauchy-Schwarz inequality,

$$\left[\sum_{i\in\bar{F},x_{1i}\ge 0} (\sqrt{d_i}x_{1i})(\sqrt{d_i}(a_2 - x_{2i}))\right]^2 \le \left[\sum_{i\in\bar{F},x_{1i}\ge 0} d_i x_{1i}^2\right]\left[\sum_{i\in\bar{F},x_{1i}\ge 0} d_i(-x_{2i})^2\right] \le \frac{1}{2}\frac{1}{2}$$

holds true with positively orienting $\mathbf{u}_1$ and negatively $\mathbf{u}_2$.

Consequently, the convenient orientation of the **AL** scenario matches that of the **LA** one. Similarly, the **AR**-**RA**, **BL**-**LB**, and **BR**-**RB** scenarios can be realized with the same orientation of $\mathbf{u}_1$, $\mathbf{u}_2$. So the ranges under the four different orientations are connected regions (by the multivariate analogue of the Bolzano theorem) and all contain the "corners" (1, 1), (−1, −1), (1, −1), (−1, 1). Therefore, the union is also a connected region in $\mathbb{R}^2$. As it is bounded from above, from below, from the right, and from the left with curves that enclose the origin, it should contain the origin too. Moreover, in each orientation, there should be a quadrant that contains the origin.

Indeed, in the **AL** case, $f_2(a_1, C)$ is greater than or less than 1, depending on whether the absolute value of $\sum_{i\in H,x_{2i}>0} d_i x_{2i}(x_{1i} - a_1)$ or that of $\sum_{i\in H,x_{2i}\le 0} d_i x_{2i}(x_{1i} - a_1)$ is larger. By the Cauchy-Schwarz inequality, this happens with the positive orientation of $\mathbf{u}_1$ and either with the positive or negative orientation of $\mathbf{u}_2$. Simultaneously, in the **BL** case, $f_2(a_1, D)$ greater than or less than −1, depending on whether the absolute value of $\sum_{i\in H,x_{2i}>0} d_i x_{2i}(x_{1i} - a_1)$ or that of $\sum_{i\in H,x_{2i}\le 0} d_i x_{2i}(x_{1i} - a_1)$ is larger. So with the positive orientation of $\mathbf{u}_1$ and either with the positive or negative orientation of $\mathbf{u}_2$, $f_2(a_1, C)$, and $f_2(a_1, D)$ are between −2 and 2, they

are parallel, in distance 2 from each other and sandwich the f_1 axis. Moreover, depending on the orientation of $\mathbf{u}_2$, the f_1 axis is either between the arcs $f_2(a_1, C)$ and $f_2(a_1, 0)$ or between the arcs $f_2(a_1, D)$ and $f_2(a_1, 0)$.

Likewise, the **AR**, **BR** estimates imply that with the negative orientation of $\mathbf{u}_1$ and either with the positive or negative orientation of $\mathbf{u}_2$, $f_2(a_1, C)$, and $f_2(a_1, D)$ are between –2 and 2, they are parallel, in distance 2 from each other and sandwich the f_1 axis. However, changing the orientation of $\mathbf{u}_1$ just means that $-a_1$ instead of a_1 can be the first coordinate of the root of f. Consequently, with the negative orientation of $\mathbf{u}_1$ and with the positive orientation of $\mathbf{u}_2$, the f_2 axis is sandwiched by the $f_2(a_1, D)$ and $f_2(a_1, 0)$ arcs, whereas the f_1 axis is sandwiched by the $f_1(0, a_2)$ and $f_1(A, a_2)$ arcs. This complies with the vertical **RB** situation.

In summary, depending on the orientation of $\mathbf{u}_1$ and $\mathbf{u}_2$, the root is in a definite quadrant of the domain (with coordinate axes a_1, a_2), and its f-map is in a definite quadrant of the range (separated by curves with $a_1 = 0$ or $a_2 = 0$). For example, if $\mathbf{u}_1$ is negatively and $\mathbf{u}_2$ is positively oriented, then a root with $a_1 < 0$ and $a_2 > 0$ is expected in the intersection of the **BR** and **RB** regions; if both $\mathbf{u}_1$ and $\mathbf{u}_2$ are negatively oriented, then a root with $a_1 < 0$ and $a_2 < 0$ is expected in the intersection of the **AR** and **RA** regions. Moreover, the signs of the coordinates of the root are compatible with the orientations of the two eigenvectors.

More simply, the orientation of $\mathbf{u}_1$ can be chosen arbitrarily. If it is negative, then $a_1 < 0$ can be expected for the sign of the first coordinate of the root; otherwise, with $-a_1$ another root is expected with the positive orientation of $\mathbf{u}_1$. The same holds vertically, depending on the orientation of $\mathbf{u}_2$. From equation (A1), it is obvious that counterorienting $\mathbf{u}_1$ and/or $\mathbf{u}_2$ will result in the same equations multiplied by –1, with $-a_1$ and/or $-a_2$ instead of a_1 and/or a_2 in the equations.

In Figure A1, we show the images of four different possible orientations of the Fiedler-carpet associated with the graph in Figure A2. For example, in the (+,+) orientation, we equidistantly subdivided the intervals $[A, B]$ and $[C, D]$ and then mapped the so-obtained grid over the $[A, B] \times [C, D]$ rectangle by the f function. The maps show the f-images of the grid points.

By equation (A1) and denoting the orientations in the superscripts, we have that

$$f_1^{-+}(a_1, a_2) = -f_1^{++}(-a_1, a_2), \quad f_2^{-+}(a_1, a_2) = f_2^{++}(-a_1, a_2),$$
$$f_1^{+-}(a_1, a_2) = f_1^{++}(a_1, -a_2), \quad f_2^{+-}(a_1, a_2) = -f_2^{++}(a_1, -a_2),$$
$$f_1^{--}(a_1, a_2) = -f_1^{++}(-a_1, -a_2), \quad f_2^{--}(a_1, a_2) = -f_2^{++}(-a_1, -a_2).$$

Therefore, the (−+) panel is the reflection of the (++) one through the f_2 axis, the (+−) panel is the reflection of the (++) one through the f_1 axis, and the (−−) panel is the reflection of the (++) one through the origin. Because of this symmetry, either all of them or none of them contain the origin, i.e., (0,0) of the f_1, f_2 plane. But we saw that in each orientation, the image of a definite quadrant sandwiches both the f_1 and f_2 axes, and the signs of the coordinates of the root are exactly the same as the orientations of the corresponding eigenvectors.

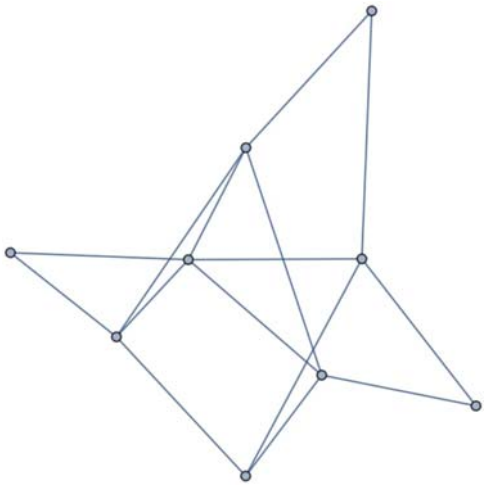

Figure A2: Graph with `graph6` string `HFRJIOY`.

We can also plot the map of the three-dimensional Fiedler-carpet of this graph. In Figure A3, different viewpoints are shown. The blue, orange, and green lines are the coordinate axes in $\mathbb{R}^3$. As can be seen from the images, they intersect at the origin *inside* of the three-dimensional body. Every plot we have created shows this intersection lying inside the map of the three-dimensional Fiedler-carpet.

If $k > 2$, then $f : \mathbb{R}^k \to \mathbb{R}^k$ maps the k-dimensional hyperrectangle with vertices of jth coordinate $\min_i x_{ji}$ or $\max_i x_{ji}$ into the k-dimensional region with vertices of jth coordinate ± 1.

Along the one-dimensional faces of this k-dimensional range, all but one a_i is fixed at its minimum/maximum. Without loss of generality, assume that $A < a_1 < B$. Akin to the $k = 2$ case, we are able to show that for each $A < a_1 < B$: $f_j(a_1, M_{-1}) \geq 0$ and $f_j(a_1, \tilde{M}_{-1}) \leq 0 (j = 2, \ldots, k)$, where $M_{-1} = (\min_m x_{2m}, \ldots, \min_m x_{km})$ is the $(k-1)$-tuple of the values of $a_2, \ldots, a_k$ all fixed at their minimum and $\tilde{M}_{-1} = (\max_m x_{2m}, \ldots, \max_m x_{km})$ is the $(k-1)$-tuple of the values of $a_2, \ldots, a_k$ all fixed at their maximum values, respectively. Indeed, for $j = 2, \ldots, k$:

$$\begin{aligned} f_j(a_1, M_{-1}) &= \sum_{i=1}^{n} d_i x_{ji} |x_{1i} - a_1| + \sum_{l=2}^{k} \sum_{i=1}^{n} d_i x_{ji} (x_{li} - \min_m x_{lm}) \\ &= \sum_{i=1}^{n} d_i x_{ji} |x_{1i} - a_1| + 1 + 0 \\ &= 2 \sum_{i \in H} d_i x_{ji} (x_{1i} - a_1) + 1 \\ &= 2 \sum_{i \in \bar{H}} d_i x_{ji} (a_1 - x_{1i}) + 1, \end{aligned}$$

and as in the second double summation, only the term for $l = j$ is 1, the others are zeros. Likewise,

$$\begin{aligned} f_j(a_1, \tilde{M}_{-1}) &= \sum_{i=1}^{n} d_i x_{ji} |x_{1i} - a_1| + \sum_{l=2}^{k} \sum_{i=1}^{n} d_i x_{ji} (x_{li} - \max_m x_{lm}) \\ &= \sum_{i=1}^{n} d_i x_{ji} |x_{1i} - a_1| - 1 + 0 \\ &= 2 \sum_{i \in H} d_i x_{ji} (x_{1i} - a_1) - 1 \\ &= 2 \sum_{i \in \bar{H}} d_i x_{ji} (a_1 - x_{1i}) - 1. \end{aligned}$$

as in the second double summation, only the term for $l = j$ is -1, the others are zeros. Note that counter-orienting $\mathbf{u}_1$ just results in $-a_1$ instead of a_1 in the root of f.

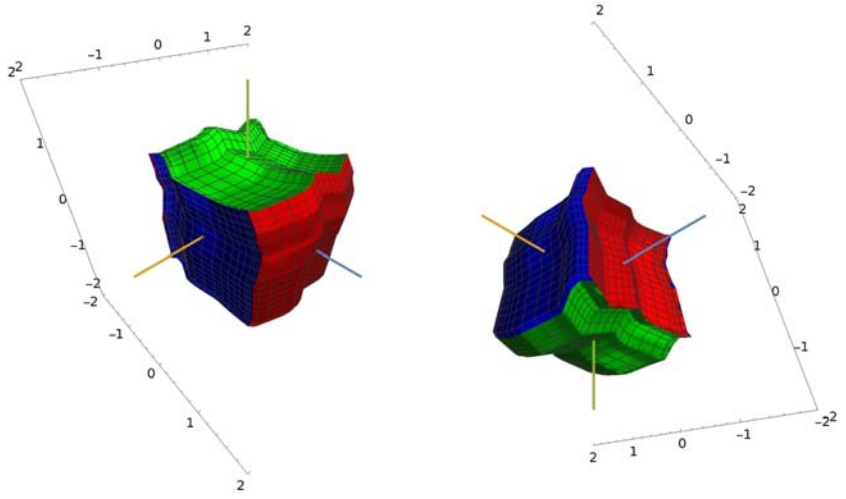

Figure A3: Image of the three-dimensional Fiedler-carpet, in two different viewpoints, of the graph in Figure A2.

The corresponding "corners" are: $f(A, M_{-1}) = (1, 1, \ldots, 1)$, $f(B, M_{-1}) = (-1, 1, \ldots, 1)$, $f(A, \tilde{M}_{-1}) = (1, -1, \ldots, -1)$, and $f(B, \tilde{M}_{-1}) = (-1, -1, \ldots, -1)$. These points are in one hyperplane, in which the pieces of the parallel arcs $f_j(a_1, M_{-1})$ and $f_j(a_1, \tilde{M}_{-1})$ are in distance 2 from each other and sandwich the f_1 axis ($j = 2, \ldots, k$). Depending on the orientations of the eigenvectors, one is confined to an orthant, and this will favor the sandwiching of the f_l axis, when another a_l moves along a one-dimensional face and the others are fixed at their minima/maxima, for $l = 2, \ldots, k$. Therefore, the connected regions between these parallel curves sandwich the corresponding coordinate axes. Consequently, their intersection, which is subset of the whole connected region, contains the origin too. Note that if $(a_1, \ldots, a_j, \ldots, a_k)$ is a root of f in some orientation of the eigenvectors, then by counter orienting only $\mathbf{u}_j$, $(a_1, \ldots, -a_j, \ldots, a_k)$ will be a root of the so-obtained f-function. Moreover, the signs of the coordinates of the root are compatible with the orientations of the eigenvectors (2^k possibilities). Also, in any combination of the orientations, the root is in a definite orthant (one of the 2^k possibilities) of the k-dimensional range (divided by surfaces, along which one of the a_js is zero).

Appendix B

Some pseudocodes follow. First for rectangular arrays of nonnegative entries.

Algorithm: finding *regular biclustering* of a contingency table

Input: $m \times n$ nondegenerate contingency table $\boldsymbol{C}$ and the number of clusters k.
1. Compute the row- and column-sums, $d_{\text{row},1}, \ldots, d_{\text{row},m}$ and $d_{\text{col},1}, \ldots, d_{\text{col},n}$;
form the diagonal matrices $\boldsymbol{D}_{\text{row}} = \text{diag}(d_{\text{row},1}, \ldots, d_{\text{row},m})$ and $\boldsymbol{D}_{\text{col}} = \text{diag}(d_{\text{col},1}, \ldots, d_{\text{col},n})$;
form the normalized contingency table $\boldsymbol{C_D} = \boldsymbol{D}_{\text{row}}^{-1/2}\boldsymbol{C}\boldsymbol{D}_{\text{col}}^{-1/2}$.
2. Compute the $k-1$ largest singular values (disregarding the trivial 1)
and the corresponding left and right singular vectors of $\boldsymbol{C_D}$: $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$;
3. Find representatives $\mathbf{r}_1, \ldots, \mathbf{r}_m$ of the rows as row vectors of the matrix $(\boldsymbol{D}_{\text{row}}^{-1/2}\mathbf{v}_1, \ldots, \boldsymbol{D}_{\text{row}}^{-1/2}\mathbf{v}_{k-1})$;
Find representatives $\mathbf{q}_1, \ldots, \mathbf{q}_n$ of the columns as row vectors of the matrix $(\boldsymbol{D}_{\text{col}}^{-1/2}\mathbf{u}_1, \ldots, \boldsymbol{D}_{\text{col}}^{-1/2}\mathbf{u}_{k-1})$.
4. Cluster the points $\mathbf{r}_1, \ldots, \mathbf{r}_m$ by the weighted k-means algorithm with weights $d_{\text{row},1}, \ldots, d_{\text{row},m}$ into k clusters;
Cluster the points $\mathbf{q}_1, \ldots, \mathbf{q}_n$ by the weighted k-means algorithm with weights $d_{\text{col},1}, \ldots, d_{\text{col},n}$ into k clusters.
Output: Clusters $R_1, \ldots, R_k$ of the row-set $\{1, \ldots, m\}$ and clusters $C_1, \ldots, C_k$ of the column-set $\{1, \ldots, n\}$.

The algorithm for edge-weighted graphs with more clusters than eigenvectors.

Algorithm: finding the *minimum cut objective* of an edge-weighted graph

Input: $n \times n$ edge-weight matrix $\boldsymbol{W}$ and the number k of the smallest separated normalized Laplacian eigenvalues of the edge-weighted graph.
1. Calculate the degree-vector $\mathbf{d}$, degree-matrix $\boldsymbol{D}$, and the normalized Laplacian
matrix $\boldsymbol{L_D} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$.
2. Compute the $k-1$ smallest positive eigenvalues (disregarding the trivial 0) and the corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$ of $\boldsymbol{L_D}$.
3. Find representatives $\mathbf{r}_1, \ldots, \mathbf{r}_n$ of the vertices as row vectors of the matrix $(\boldsymbol{D}^{-1/2}\mathbf{u}_1, \ldots, \boldsymbol{D}^{-1/2}\mathbf{u}_{k-1})$.
4. Cluster the points $\mathbf{r}_1, \ldots, \mathbf{r}_n$ by the weighted k-means algorithm with the components of $\mathbf{d}$ as weights into 2^{k-1} clusters.
Output: Clusters $V_1, \ldots, V_{2^{k-1}}$ of the vertices.

The weighted k-means algorithm follows. If the weights are the same (all equal to 1), then this is the usual k-means algorithm.

Algorithm: weighted k-means clustering

Input: finite dimensional points $\mathbf{r}_1, \ldots, \mathbf{r}_n$ with weights $d_1, \ldots, d_n$ and the number of clusters k.

1. Initialize: $V_1^{(0)}, \ldots, V_k^{(0)}$, the clusters of $\{1, \ldots, n\}$.

2. Iterate: *for* $t = 1, 2, \ldots$

a. calculate the cluster centers $\mathbf{c}_a^{(t)} = \frac{1}{\sum_{j \in V_a^{(t-1)}} d_j} \sum_{j \in V_a^{(t-1)}} d_j \mathbf{r}_j$ $(a = 1, \ldots, k)$;

b. relocate the points: $\mathbf{r}_j$ is assigned to cluster $V_a^{(t)}$ for which $\|\mathbf{r}_j - \mathbf{c}_a^{(t)}\|$ is minimum,

until convergence.

Output: Clusters $V_1, \ldots, V_k$ of $\{1, \ldots, n\}$.

Appendix C

Pairwise plots of the correspondence analysis results based on the first three coordinate axes (coordinates of the left singular vectors for immigration and right singular vectors for emigration data) are shown in Figure A4(a) and (b) for 2 years. The cluster memberships obtained in Section 4 are illustrated with different colors.

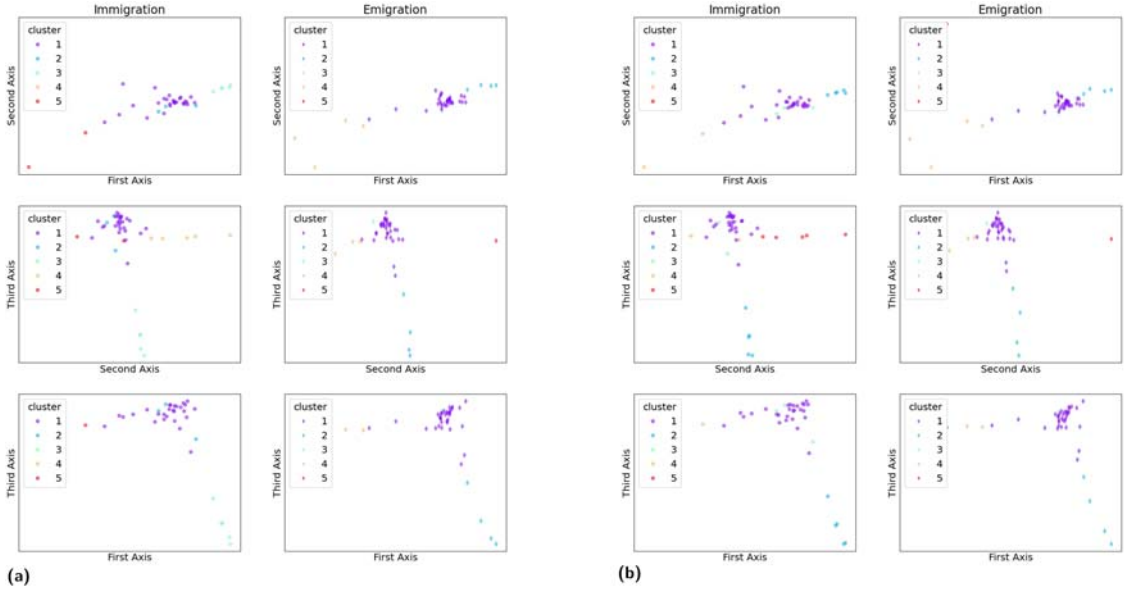

Figure A4: Pairwise plots of the correspondence analysis results based on the first three singular vector pairs, enhanced with the cluster memberships, illustrated by different colors. (a) 2015 and (b) 2019.