SPECTRAL CLUSTERING, Lesson 1.

Quadratic placement problems, graph spectra, and multiway cuts (hypergraphs can be skipped)

Marianna Bolla, DSc. Prof. BME Math. Inst.

September 10, 2024

You are assumed to learn only the most important notions, theorems, and methods, not the proofs. Some sections (e.g., those related to hypergraphs) can be skipped, it will be noted after the section title.

1 Graph based matrices and spectra

Graph spectra are used for about 50 years to recover the structure of graphs. Different kinds of spectra are capable of finding multiway cuts corresponding to different optimization criteria. While eigenvalues give estimates for the objective functions of the discrete optimization problems, eigenvectors are used to find clusters of vertices which approximately solve the problems. These methods are reminiscent of some classical methods of Multivariate Statistical Analysis. Throughout this chapter, methods of Principal Component Analysis and Correspondence Analysis are used to solve quadratic placement tasks on weighted graphs and contingency tables. As a result, we get low rank representation of the graph's vertices or rows and columns of the contingency table by means of linear methods so that the representation somehow favors our classification criteria. Non-linearities are treated by mapping the data into a feature space (reproducing kernel Hilbert space).

Then minimum ratio- or normalized multiway cut problems are discussed together with modularity cuts. Since the optima of the corresponding objective functions are taken on partition vectors corresponding to the hidden clusters, they are related to Laplacian, normalized Laplacian, or modularity spectra, whereas the precision of the estimates depends on the distance between the subspaces spanned by the corresponding eigenvectors and partition vectors. By an Analysis of Variance argument, this distance is the sum of the inner variances of the underlying clusters, the objective function of the k-means clustering.

2 Quadratic placement problems for edge-weighted graphs

In Multivariate Statistical Analysis, Principal Component Analysis and Factor Analysis are crucial methods for reducing the dimensionality of the data via representing them by a smaller number of independent factors. The representation also gives rise to clustering the data in the factor space. Given an $n \times n$ positive definite covariance matrix $\boldsymbol{C}$ of a random vector, the principal components are determined successively, and they are the maximum variance normalized linear combinations of the components of the underlying random vector conditioned on the uncorrelatedness (in the Gaussian case, independence) of them.

Now, our data matrix corresponds to a graph. First, let $G = (V, E)$ be a simple graph on the vertex-set V and edge-set E with $|V| = n$ and $|E| \leq \binom{n}{2}$. Thus, the $|E| \times n$ data matrix $\boldsymbol{B}$ has 0-1 entries, the rows correspond to the edges, the columns to the vertices, and b_{ij} is 1 or 0 depending on whether the edge i contains the vertex j as an endpoint or not. The Gram-matrix $\boldsymbol{C} = \boldsymbol{B}^T\boldsymbol{B}$ is the non-centralized covariance matrix based on the data matrix $\boldsymbol{B}$, and is both positive definite (provided there are no multiple edges) and a Frobenius type matrix with nonnegative entries to which the Frobenius Theorem is applicable. Sometimes the matrix $\boldsymbol{C}$ is called signess Laplacian and its eigenspaces are used to compare cospectral graphs. It is easy to see that $\boldsymbol{C} = \boldsymbol{D}+\boldsymbol{A}$, where $\boldsymbol{A} = (a_{ij})$ is the usual *adjacency matrix* of G (it is symmetric and a_{ij} is 1 or 0 depending on whether vertices i and j are connected or not; $a_{ii} = 0$, $i = 1, \ldots, n$), while $\boldsymbol{D}$ is the so-called degree-matrix, i.e. diagonal matrix, containing the vertex degrees in its main diagonal. $\boldsymbol{A}$ being a Frobenius-type matrix, its maximum absolute value eigenvalue is positive, and it is at most the maximum vertex-degree, and apart from the trivial case – when there are no edges at all – it is indefinite, as the sum of its eigenvalues, i.e. the trace of $\boldsymbol{A}$, is zero.

Instead of the positive definite matrix $\boldsymbol{C}$, for optimization purposes, as will be derived below, the Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ is more suitable, which is positive semidefinite, having always a zero eigenvalue, since the row sums are zeros. This $\boldsymbol{L}$ is sometimes called combinatorial or difference Laplacian, whereas we will introduce the so-called normalized Laplacian, $\boldsymbol{L}_D$ too. If our graph is regular, then $\boldsymbol{D} = d\boldsymbol{I}$ (where d is the common degree of the vertices and $\boldsymbol{I}$ is the identity matrix) and the eigenvalues of $\boldsymbol{C}$ and $\boldsymbol{L}$ are obtained from those of $\boldsymbol{A}$ by adding d to them or subtracting them from d, respectively.

There are too many matrices around, not speaking about the modularity and normalized modularity matrices, latter one closely related to the normalized Laplacian, akin to the so-called transition probability matrix $\boldsymbol{D}^{-1}\boldsymbol{A}$, or $\boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{A}$ which is sometimes called random walk Laplacian (this matrix is not symmetric, still it has real eigenvalues). Our purpose is to clarify in which situation which of these matrices is the best applicable. The whole story simplifies if we use edge-weighted graphs, and all these matrices come into existence naturally, while solving some optimization problems.

2.1 Representation of edge-weighted graphs

From now on, we will use the more general framework of an edge-weighted graph. A simple graph is a special case of it with 0-1 weights. Let $G = (V, \boldsymbol{W})$

be a graph on n vertices, where $V = \{1, \ldots, n\}$ and the $n \times n$ symmetric matrix $\boldsymbol{W}$ has nonnegative real entries and zero diagonal. Here w_{ij} is the similarity between vertices i and j, where 0 similarity means no connection (edge) at all. If G is a simple graph, $\boldsymbol{W}$ is its adjacency matrix. Since $\boldsymbol{W}$ is symmetric, the weight of the edge between two vertices does not depend on its direction, i.e. our graph is *undirected*. In this book, we will mostly treat undirected graphs.

Let the row-sums of $\boldsymbol{W}$ be

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

which are called *generalized vertex-degrees* and collected in the main diagonal of the diagonal *degree-matrix* $\boldsymbol{D} = \mathrm{diag}(\mathbf{d})$, where $\mathbf{d} = (d_1, \ldots, d_n)^T$ is the so-called *degree-vector*.

For a given integer $1 \leq k \leq n$ we are looking for k-dimensional representatives $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^k$ of the vertices such that they minimize the objective function

$$Q_k = \sum_{i<j} w_{ij} \|\mathbf{r}_i - \mathbf{r}_j\|^2 \geq 0 \tag{1}$$

subject to

$$\sum_{i=1}^{n} \mathbf{r}_i \mathbf{r}_i^T = \boldsymbol{I}_k \tag{2}$$

where $\boldsymbol{I}_k$ is the $k \times k$ identity matrix. When minimized, the objective function Q_k favors k-dimensional placement of the vertices such that vertices connected with large-weight edges are forced to be close to each other.

Let us put both the objective function and the constraint in a more favorable form. Denote by $\boldsymbol{X}$ the $n \times k$ matrix of rows $\mathbf{r}_1^T, \ldots, \mathbf{r}_n^T$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ be the columns of $\boldsymbol{X}$, for which fact we use the notation $\boldsymbol{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$. Because of the constraint (2), the columns of $\boldsymbol{X}$ form an orthonormal system, hence, $\boldsymbol{X}$ is a suborthogonal matrix. Therefore, the constraint (2) can be formulated as $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}_k$. With this notation, the objective function (1) is rewritten in the symmetrized form

$$\begin{aligned} Q_k &= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \|\mathbf{r}_i - \mathbf{r}_j\|^2 = \sum_{i=1}^{n} d_i \|\mathbf{r}_i\|^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \mathbf{r}_i^T \mathbf{r}_j \\ &= \sum_{\ell=1}^{k} \mathbf{x}_\ell^T (\boldsymbol{D} - \boldsymbol{W}) \mathbf{x}_\ell = \mathrm{tr}[\boldsymbol{X}^T (\boldsymbol{D} - \boldsymbol{W}) \boldsymbol{X}]. \end{aligned} \tag{3}$$

Definition 1 *The matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ is called the Laplacian corresponding to the edge-weighted graph $G = (V, \boldsymbol{W})$.*

For simple graphs, we get back the usual definition of the Laplacian. About the physical meaning of it, see papers [16, 25, 43, 24].

The Laplacian is always positive semidefinite that can easily be seen from the $Q_1 \geq 0$ relation, and it always has a zero eigenvalue, since its rows sum up to zero. It can be shown, that the multiplicity of 0 as an eigenvalue of $\boldsymbol{L}$ is equal to the number of connected components of $G = (V, \boldsymbol{W})$, i.e. the maximum number of disjoint subsets of V such that there are no edges connecting vertices

of distinct subsets (where no edge means an edge with zero weight). In terms of $\boldsymbol{W}$, the number of connected components of G is the maximum number of the diagonal blocks which can be achieved by the same permutation of the rows and columns of $\boldsymbol{W}$. For simple graphs, the proof is in [61], among others, and it is analogous for the edge-weighted case. Consequently, if G is connected, then 0 is a single eigenvalue with corresponding unit-norm eigenvector $\mathbf{u}_0 = \frac{1}{\sqrt{n}}\mathbf{1}$, where $\mathbf{1}$ denotes the all 1's vector. In the sequel, we will assume that G is connected (or equivalently, $\boldsymbol{W}$ is irreducible).

Theorem 1 (Representation theorem for edge-weighted graphs) *Let $G = (V, \boldsymbol{W})$ be a connected edge-weighted graph with Laplacian matrix $\boldsymbol{L}$. Let $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{n-1}$ be the eigenvalues of $\boldsymbol{L}$ with corresponding unit-norm eigenvectors $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$. Let $k < n$ be a positive integer such that $\lambda_{k-1} < \lambda_k$. Then the minimum of (1) subject to (2) is*

$$\sum_{i=0}^{k-1} \lambda_i = \sum_{i=1}^{k-1} \lambda_i$$

and it is attained with the optimum representatives $\mathbf{r}_1^*, \ldots, \mathbf{r}_n^*$ the transposes of which are row vectors of $\boldsymbol{X}^* = (\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{k-1})$.

Proof 1 *Using (3), our objective function is*

$$Q_k = \mathrm{tr}[\boldsymbol{X}^T \boldsymbol{L} \boldsymbol{X}]$$

and it is to be minimized under $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}_k$. Then a linear algebra theorem guarantees that the solution is the required one.

Definition 2 *The vectors $\mathbf{r}_1^*, \ldots, \mathbf{r}_n^*$ giving the optimum in Theorem 1 are called optimal k-dimensional representatives of the vertices, while the eigenvectors $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$ of $\boldsymbol{L}$ are called vector components taking part in the optimal k-dimensional representation.*

We remark the following.

- The dimension k does not play an important role here, the vector components can be included one after the other up to a k such that $\lambda_{k-1} < \lambda_k$.
- We remark that the eigenvectors can be arbitrarily chosen in the eigenspaces corresponding to possible multiple eigenvalues, under the orthogonality conditions. Further, the representatives can as well be rotated in $\mathbb{R}^k$. Indeed, nor the objective function, neither the constraint is changed if we use $\boldsymbol{R}\mathbf{r}_i$'s instead of $\mathbf{r}_i$'s, or equivalently, $\boldsymbol{X}\boldsymbol{R}$ instead of $\boldsymbol{X}$, where $\boldsymbol{R}$ is an arbitrary $k \times k$ orthogonal matrix.
- Since the eigenvector $\mathbf{u}_0$ has equal coordinates, the same first coordinates of the vertex representatives do not play an important role in the representation, especially when the representatives are used for clustering purposes. Therefore, $\mathbf{u}_0$ can be omitted, and an optimum $(k-1)$-dimensional representation is performed based on the eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$.
- For the time being, we assumed that $\boldsymbol{W}$ has zero diagonal. We can as well see that in the presence of possible loops (some or all diagonal entries of $\boldsymbol{W}$ are positive) the objective function and the Laplacian remains the same, hence, Theorem 1 is applicable to this situation too.

2.2 Representation of hypergraphs

[This section can be skipped.]

For hypergraphs, the minimum placement problem is formulated in terms of the representatives of vertices and hyperedges, but it will turn out that we can always assign an edge-weighted graph to our hypergraph so that the two quadratic placement problems are equivalent in terms of the vertices.

Let $H = (V, E)$ be a hypergraph with vertex-set $V = \{v_1 \dots v_n\}$ and edge-set $E = \{e_1 \dots e_m\}$. H is uniquely defined by its $m \times n$ incidence matrix $\boldsymbol{B}$ (0-1 data matrix) such that $b_{ij} = \mathcal{I}(v_j \in e_i)$, where $\mathcal{I}(v_j \in e_i)$ is 1 if the vertex v_j is contained in the hyperedge e_i, and zero otherwise.

For a fixed integer k ($1 \leq k \leq n$), we are looking for k-dimensional representatives $\mathbf{r}_1, \dots, \mathbf{r}_n$ of the vertices and $\mathbf{q}_1, \dots, \mathbf{q}_m$ of the edges subject to

$$\sum_{i=1}^{n} \mathbf{r}_i \mathbf{r}_i^T = \boldsymbol{I}_k \tag{4}$$

so that the following sum of the costs of edges in this representation is minimized:

$$Q_k = \sum_{i=1}^{m} C(e_i), \tag{5}$$

where the cost of the edge e_i is

$$C(e_i) = \sum_{j=1}^{n} b_{ij} \|\mathbf{r}_j - \mathbf{q}_i\|^2.$$

The construction of the above objective function forces the representatives of the vertices to be close to those of the hyperedges they are contained in, while the constraint keeps them in a distance. As a compromise, the representatives of the vertices which are together in many hyperedges will be close to each other.

To minimize the objective function (5), let $\bar{\mathbf{r}}(e)$ denote the barycenter of the vertex-representatives contained in the hyperedge e:

$$\bar{\mathbf{r}}(e) = \frac{1}{|e|} \sum_{j=1}^{n} \mathcal{I}(v_j \in e) \mathbf{r}_j.$$

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote the $n \times k$ and $m \times k$ matrices containing the vertex- and edge-representatives as row vectors, respectively. Further, let $\boldsymbol{D}_v$ and $\boldsymbol{D}_e$ be the $n \times n$ and $m \times m$ vertex- and edge-valence matrices: they are diagonal matrices, with diagonal entries that are the column- and row-sums of the incidence matrix $\boldsymbol{B}$, respectively. Assume that $\boldsymbol{D}_e$ is not singular (there are no empty hyperedges).

With this notation, $C(e)$ is decreased by means of the Steiner's inequality:

$$C(e) \geq \sum_{j=1}^{n} \mathcal{I}(v_j \in e) \|\mathbf{r}_j - \bar{\mathbf{r}}(e)\|^2, \quad e \in E. \tag{6}$$

The right-hand side only depends on the incidence relations $\mathcal{I}(v_j \in e)$ and on the representatives of the vertices comprising the row vectors of the matrix $\boldsymbol{X}$.

Denoting the right-hand side of (6) by $Q(e, \boldsymbol{X})$, with an easy calculation we get the following formula for it:

$$Q(e, \boldsymbol{X}) = \frac{1}{2|e|} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathcal{I}(v_i \in e)\mathcal{I}(v_j \in e)\|\mathbf{r}_i - \mathbf{r}_j\|^2, \quad e \in E. \tag{7}$$

With the notation $Q(\boldsymbol{X}) = \sum_{e \in E} Q(e, \boldsymbol{X})$, the inequality $Q_k \geq Q(\boldsymbol{X})$ holds with any representation $\boldsymbol{X}$ of the vertices. But $Q(\boldsymbol{X})$ can be rewritten like we did with weighted graphs:

$$Q(\boldsymbol{X}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{1}{2} \sum_{e \in E} \mathcal{I}(v_i \in e)\mathcal{I}(v_j \in e) \frac{1}{|e|} \right] \|\mathbf{r}_i - \mathbf{r}_j\|^2 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \ell_{ij} \mathbf{r}_i^T \mathbf{r}_j,$$

where

$$\ell_{ij} = \begin{cases} -\mathcal{I}(v_i \in e)\mathcal{I}(v_j \in e)\frac{1}{|e|} & \text{if} \quad i \neq j \\ d_{vi} - \mathcal{I}(v_i \in e)\frac{1}{|e|} & \text{if} \quad i = j, \end{cases} \tag{8}$$

where d_{vi} is the ith diagonal entry of the diagonal matrix $\boldsymbol{D}_v$.

It is easy to see that the $n \times n$ matrix of the entries ℓ_{ij} is $\boldsymbol{D}_v - \boldsymbol{B}^T \boldsymbol{D}_e^{-1} \boldsymbol{B}$.

Definition 3 *The matrix $\tilde{\boldsymbol{L}} = \boldsymbol{D}_v - \boldsymbol{B}^T \boldsymbol{D}_e^{-1} \boldsymbol{B}$ is called Laplacian of the hypergraph H.*

For simple graphs, the above $\tilde{\boldsymbol{L}}$ is one-half of the usual Laplacian $\boldsymbol{L}$, as each edge has valence 2. Further, to any hypergraph $H = (V, E)$ an edge-weighted graph $G = (V, \boldsymbol{W})$ can be assigned such that between their Laplacians the relation $\tilde{\boldsymbol{L}} = \frac{1}{2}\boldsymbol{L}$ holds, in the following manner:

$$w_{ij} = \begin{cases} \sum_{e \in E} \mathcal{I}(v_i \in e)\mathcal{I}(v_j \in e)\frac{1}{|e|} & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j. \end{cases} \tag{9}$$

Therefore, $\tilde{\boldsymbol{L}}$ is positive semidefinite, and the multiplicity of 0 as an eigenvalue of H is equal to the number of its connected components. The connected components of a hypergraph are spanned by disjoint subsets of its vertices such that there are no hyperedges containing vertices from more than one component. If vertices of the distinct components are colored with different colors, there are only monocolored hyperedges. H is connected if it has only one connected component. In the sequel, only connected hypergraphs will be considered.

Theorem 2 (Representation theorem for hypergraphs) *Let $H = (V, E)$ be a connected hypergraph with Laplacian matrix $\tilde{\boldsymbol{L}}$. Let $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{n-1}$ be the eigenvalues of $\tilde{\boldsymbol{L}}$ with corresponding unit-norm eigenvectors $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$. Let $k < n$ be a positive integer such that $\lambda_{k-1} < \lambda_k$. Then the minimum of the cost function (5) subject to (4) is*

$$\sum_{j=0}^{k-1} \lambda_j = \sum_{j=1}^{k-1} \lambda_j$$

and it is attained with the optimum representatives $\mathbf{r}_1^*, \ldots, \mathbf{r}_n^*$ of the vertices, the transposes of which are row vectors of $\boldsymbol{X}^* = (\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{k-1})$. Further, the optimum representatives of the edges are row vectors of the matrix

$$\boldsymbol{Y}^* = \boldsymbol{D}_e^{-1} \boldsymbol{B} \boldsymbol{X}^*.$$

Proof 2 *It is easy to see that*

$$Q(\boldsymbol{X}) = \mathrm{tr}[\boldsymbol{X}^T \tilde{\boldsymbol{L}} \boldsymbol{X}]$$

and it is to be minimized subject to $\boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{I}_k$*. We again use the linear algebra fact of* **??** *and the following easy observation: the relation (6) implies that the optimum representative of an edge* e *is the barycenter of the representatives of its vertices.*

From the defining formula (8) of the Laplacian, it can easily be seen that the loop-edges ($|e| = 1$) do not give any contribution to it. Akin to weighted graphs, $(k-1)$-dimensional representatives will as well do after eliminating the first (trivial) coordinate.

We first defined the Laplacian of hypergraphs in a binary clustering problem, see [17]. We remark that [27] treats a similar problem in the framework of bipartite graphs, also related to contingency tables.

2.3 Examples for spectra and representation of simple graphs

[For your information, but you need not know these spectra by heart.]

Here we give some examples for the smallest Laplacian eigenvalues and the representation based on the corresponding eigenvectors in a straightforward dimension of basic simple graphs. Their adjacency matrix $\boldsymbol{A}$ can as well be considered as a 0-1 weight matrix of an edge-weighted graph, and they are also hypergraphs with all edge-valences two. The reader can find many examples for adjacency spectra in [16, 25, 35] and [39]. In most of our examples the underlying graph is nearly regular, therefore there is an asymptotic relation between their Laplacian and adjacency spectra, at least for large n.

Note that neither the adjacency nor the Laplacian spectrum is affected by the labeling of the vertices, in other words, isomorphic graphs have the same spectrum. However, the converse is not true: there are graphs with the same spectrum, though they are not isomorphic (e.g. [25] discussed such cospectral graphs in details). This is not surprising, since it is the SD (spectrum and eigenvectors or eigenspaces together) which uniquely characterizes a symmetric matrix, and not the spectrum itself. However, [37] found a class of graphs that is characterized by the Laplacian spectrum. For example, it can be shown that almost all trees are cospectral, see also [44].

(a) The adjacency matrix $\boldsymbol{A}$ of the *complete graph* C_n on n vertices is comprised of entries $a_{ij} = 1$ for $i \neq j$ and $a_{ii} = 0$ for $i = 1, \ldots, n$. Consequently, its corresponding Laplacian is $\boldsymbol{L} = (n-1)\boldsymbol{I}_n - \boldsymbol{A}$ and $\mathrm{tr}(\boldsymbol{L}) = n(n-1)$. For symmetry reasons, the spectrum of $\boldsymbol{L}$ is nothing else but

$$\lambda_0 = 0, \quad \lambda_1 = \cdots = \lambda_{n-1} = \frac{n(n-1)}{n-1} = n.$$

The unique eigendirection corresponding to the eigenvalue zero is designated by the vector $\mathbf{1}$, while the eigenspace corresponding to the multiple eigenvalue n is $\mathbf{1}^{\perp}$, the orthogonal complementary subspace of $\mathbf{1}$ in $\mathbb{R}^n$. Since there is only one positive eigenvalue (with multiplicity $n-1$), here

only the trivial one- or the n-dimensional representation of C_n makes sense. The one-dimensional representatives of the vertices are the same points, whereas the n-dimensional representatives form a simplex on n vertices in the $(n-1)$-dimensional hyperplane $\mathbf{1}^\perp$ of $\mathbb{R}^n$. These are, in fact, zero- and $(n-1)$-dimensional representations, because of the trivial first eigenvector $\mathbf{1}$.

(b) The adjacency matrix $\boldsymbol{A}$ of the *path graph* P_n on n vertices is a tridiagonal matrix with zero diagonal and all 1's above and below the diagonal entries. Here the vertices are labeled in their natural succession. It is proved in [25, 39] that the adjacency spectrum of $\boldsymbol{A}$ consists of the numbers

$$2\cos\frac{i\pi}{n+1}, \quad i=1,\ldots,n.$$

The Laplacian spectrum of P_n (see circulant matrices) consists of the eigenvalues

$$\lambda_i = 4\sin^2\frac{i\pi}{2n} = 2(1-\cos\frac{i\pi}{n}), \quad i=0,1,\ldots,n-1.$$

Hence, the smallest positive Laplacian eigencalue of P_n is $\lambda_1 = 2(1-\cos\frac{\pi}{n})$. For odd n (say, $n=2\ell+1$), disregarding of the trivial dimension, the one-dimensional representatives of the vertices, i.e. the coordinates of $\mathbf{u}_1$, are the numbers

$$x_j = \sqrt{\frac{2}{n}}\sin j\frac{\pi}{n}, \quad j=-\ell,\ldots,-1,0,1,\ldots,\ell \tag{10}$$

forming a path, where the distances between representatives of neighboring vertices follow the sine rhythm of (10).

(c) The 2-dimensional $m\times n$ grid $G_{m,n}$ is the Cartesian product (in other words, direct sum) of P_m and P_n, hence its adjacency eigenvalues (see [25]) are the numbers

$$\alpha_{i,j} = 2\cos\frac{i\pi}{m+1} + 2\cos\frac{j\pi}{n+1}, \quad i=1,\ldots,m; \quad j=1,\ldots,n.$$

With the considerations of [30], a similar result holds for he Laplacian eigenvalues of the Cartesian product of the simple graphs G_1 and G_2: they are equal to all possible sums of eigenvalues of the two factors. Since $G_{m,n}$ is the Cartesian product of P_m and P_n, the Laplacian eigenvalues of the $m\times n$ grid are

$$\lambda_{i,j} = 4\sin^2\frac{i\pi}{2m} + 4\sin^2\frac{i\pi}{2n} = 2(1-\cos\frac{i\pi}{m}) + 2(1-\cos\frac{i\pi}{n}),$$
$$i=0,1,\ldots,m-1; \quad j=0,\ldots,n-1.$$

More generally, denote by $Grid_{d,\ell}$ the d-dimensional cubic grid ($d\geq 2$ is an integer) with $n=(2\ell+1)^d$ vertices, where the vertices are characterized by d-tuples of integers $-\ell,\ldots,-1,0,1,\ldots,\ell$ such that two vertices are adjacent if and only if their d-tuples differ in exactly one coordinate. $Grid_{2,\ell} = G_{2\ell+1,2\ell+1}$ and $Grid_{d,\ell}$ is the Cartesian product of d copies

of $P_{2\ell+1}$. Using these facts, the adjacency eigenvalues of $Grid_{d,\ell}$ are the numbers

$$2\sum_{j=1}^{d}\cos\frac{i_j\pi}{2\ell+2}, \quad i_1,\ldots,i_d = 1,\ldots,2\ell+1.$$

In [39], the adjacency spectrum of the d-dimensional hypercube is also derived via Cartesian products. With similar considerations, the Laplacian eigenvalues of $Grid_{d,\ell}$ are

$$\lambda_{i_1,\ldots,i_d} = 4\sum_{j=1}^{d}\sin^2\frac{i_j\pi}{2(2\ell+1)} = 2\sum_{j=1}^{d}(1-\cos\frac{i_j\pi}{2\ell+1}) \qquad (11)$$
$$i_1,\ldots,i_d = 0,\ldots,2\ell.$$

The smallest positive Laplacian eigenvalue of the d-dimensional cubic grid on $n = (2\ell+1)^d$ vertices is a λ with all but one subscripts 0, and the non-zero subscript is 1 in (11). As there are d choices for the non-zero index, the smallest positive eigenvalue is

$$4\sin^2\frac{\pi}{2(2\ell+1)} = 2(1-\cos\frac{\pi}{2\ell+1})$$

with multiplicity d. The d-dimensional representatives of the vertices (after leaving out the trivial dimension) form a grid in a d-dimensional hyperplane of $\mathbb{R}^n$, its center of gravity being the origin, whereas the distances between the representatives of adjacent vertices follow the sine rhythm of (10).

(d) Let $K_{n_1,\ldots,n_k}$ be the *complete k-partite graph*, where $n = \sum_{i=1}^{k} n_i$ is the number of its vertices. Let $V_1,\ldots,V_k$ denote the non-empty, disjoint, independent sets of the vertices (called clusters), where $|V_i| = n_i$ $(i = 1,\ldots,k)$.

Proposition 1 *The Laplacian spectrum of $K_{n_1,\ldots,n_k}$ consists of a single 0, the numbers $n-n_i$ with multiplicity n_i-1 $(i=1,\ldots,k)$ and the number n with multiplicity $k-1$. Further, in the $(k-1)$-dimensional representation of the vertices via any orthonormal set of $k-1$ eigenvectors corresponding to the largest eigenvalue n, the representatives of vertices of the same cluster coincide.*

Proof 3 *Choose a labeling of the vertices such that the first n_1 vertices are contained in V_1, the next n_2 ones in V_2, etc. Since the complete multipartite graph is connected, 0 is a single eigenvalue with eigendirection* $\mathbf{1}$. *The adjacency matrix $\boldsymbol{A}$ of $K_{n_1,\ldots,n_k}$ is a symmetric block matrix of $k\times k$ blocks, where the diagonal blocks are all zeros, and the off-diagonal ones have all 1 entries. The diagonal degree-matrix is $\boldsymbol{D} = \boldsymbol{D}_1\oplus\cdots\oplus\boldsymbol{D}_k$, where $\boldsymbol{D}_i = (n-n_i)\boldsymbol{I}_{n_i}$. The eigenvalue–eigenvector equation for the Laplacian $\boldsymbol{L} = \boldsymbol{D}-\boldsymbol{A}$ of $K_{n_1,\ldots,n_k}$ yields the system of equations*

$$(n-n_i)x_j - \sum_{\ell\notin V_i} x_\ell = \lambda x_j, \quad j\in V_i;\ i=1,\ldots,k \qquad (12)$$

where λ is an eigenvalue of $\boldsymbol{L}$ with corresponding eigenvector $\mathbf{x} = (x_1, \ldots, x_n)^T$. It is easy to see that with $\lambda = n$, the system (12) simplifies to

$$n_i x_j + \sum_{\ell \notin V_i} x_\ell = 0, \quad j \in V_i,\, i = 1, \ldots, k$$

which is solved by any piecewise constant vector $\mathbf{x}$ on the partition $(V_1, \ldots, V_k)$ such that it is orthogonal to the $\mathbf{1} \in \mathbb{R}^n$ vector at the same time. Indeed, let $x_j = y_i$ for $j \in V_i$ and because of the orthogonality condition, the numbers $y_1, \ldots, y_k$ satisfy $\sum_{i=1}^k n_i y_i = 0$. The subspace of $\mathbb{R}^n$ spanned by these $\mathbf{x}$'s is of dimension $k-1$. Therefore, the multiplicity of the eigenvalue n is $k-1$, and the $(k-1)$-dimensional representatives of the vertices – with vector components that form an orthonormal system within this eigenspace – yield k distinct points such that vertices of the same cluster are represented with the same point.

One can also verify that for any $i \in \{1, \ldots, k\}$, substituting $n - n_i$ for λ, the system (12) becomes:

$$\sum_{\ell \notin V_i} x_\ell = 0,$$

which is solved by any $\mathbf{x}$ such that $x_j = 0$ whenever $j \notin V_i$. For $j \in V_i$, we select the coordinates of $\mathbf{x}$ such that $\mathbf{x}$ be orthogonal to the $\mathbf{1}$ vector. This condition results in the restriction $\sum_{j \in V_i} x_j = 0$ for the non-zero coordinates of $\mathbf{x}$. This restriction also ensures the orthogonality to the piecewise constant vectors in the eigenspace corresponding to the eigenvalue n. Trivially, the subspace of such $\mathbf{x}$'s is of dimension $n - n_i$, and hence, the multiplicity of the eigenvalue $n - n_i$ is $n_i - 1$, for $i = 1, \ldots, k$.

(e) Let S_d denote the *star graph* on $n = d + 1$ vertices. In fact, the star graph is a complete 2-partite graph, namely, $S_d = K_{1,d}$. Applying the result of (d), its spectrum consists of a single 0, the number $n - d = 1$ (with multiplicity $d - 1$) and the single eigenvalue n. Making a d-dimensional representation, based on the d smallest Laplacian eigenvalues, the representatives of the d endpoints form the vertices of a simplex in the $(d-1)$-dimensional eigenspace corresponding to the eigenvalue 1, but they are not connected to each other, they are merely connected to the single vertex of degree d, the representative of which is the origin (the center of gravity of the simplex). This kind of representation is reminiscent of graph drawing and imitates the physical picture of the graph. However, a 1-dimensional representation is also possible based on the eigenvector corresponding to the largest single eigenvalue. In this structure-revealing representation, the representatives form two points on the real number line: one corresponds to the endpoints, and the other to the middle point.

(f) Let $S_{d,\ell}$ denote the subdivision graph of S_d, where each of the edges of S_d is divided into ℓ parts. We call $S_{d,\ell}$ *spider graph* with d feet of ℓ sections. The number of its vertices is $n = d\ell + 1$. The Laplacian spectrum inherits features of that of the path and the star. Namely, the smallest positive Laplacian eigenvalue of $S_{d,\ell}$ is of multiplicity $d-1$ and equal to $1 - \cos\frac{\pi}{2\ell+1}$. The optimal d-dimensional (in fact, $(d-1)$-dimensional) representation of

the spider is that of (e), where the feet of the spider are divided according to the sine rhythm of (10).

We remark that the star graph is the only tree with smallest positive eigenvalue 1. Indeed, [41] and [42] independently proved that for a tree T, $\lambda_1(T) \leq 1$, with equality if and only if T is a star.

Proposition 1 illustrates that in case of empty clusters, the representation, based on eigenvectors corresponding to eigenvalues in the top of the Laplacian spectrum, is able to reveal this so-called anti-community cluster structure. This is also true for clusters (subsets) of vertices with sparse intra- and dense inter-cluster edge-densities (we will define these notions more precisely later. For example, hub authorities are of this type. Another example is a game of strategic substitutes, where an increase in other players' actions leads to relatively lower payoffs under higher actions of a given player. In strategic interaction games the agents are vertices of a graph and only agents connected with an edge influence each other's actions.

Analogously to the derivation of the Representation Theorem (Theorem 1), now the maximum of the quadratic form $Q_k = \boldsymbol{X}^T \boldsymbol{L} \boldsymbol{X}$ subject to $\boldsymbol{X}^T \boldsymbol{X} = \boldsymbol{I}_k$ is looked for, and hence, representatives of vertices connected with few, low-weight edges are stressed to be close to each other. As an easy linear algebra fact, the maximum is the sum of the k largest Laplacian eigenvalues and it is attained with an $\boldsymbol{X}^*$ containing the corresponding unit-norm eigenvectors in its columns.

Later we will study more general cluster structures such that clusters are not necessarily dense or sparse subsets of the network, but rather homogeneous ones, as far as the intra- and inter-cluster relations of the vertices are concerned. In other words, vertices of the same cluster behave similarly with respect to each other and to vertices of any other cluster, like synopses of the brain. These types of clusters can be recovered via representation based on eigenvectors corresponding to some structural eigenvalues of the normalized modularity matrix to be introduced in the next lesson. The so-called structure-revealing representation of Example (e) is also of this flavor. Bottom Laplacian eigenvalues and eigenvectors are mainly advisable for graph drawing purposes, for a detailed description see [40]. Even in case of Laplacian spectra of chemical graphs we may select not necessarily consecutive eigenvalues together with eigenvectors for spacial representation of molecules, see [50] for details.

3 Estimating multiway cuts via spectral relaxation

Clusters (in other words, modules or communities) of graphs are typical (strongly or loosely connected) subsets of vertices that can be identified, for example, with social groups or interacting enzymes in social or metabolic networks, respectively; they form special partition classes of the vertices. To measure the performance of a clustering, different kinds of multiway cuts are introduced and estimated by means of Laplacian spectra. The key motif of these estimations is that minima and maxima of the quadratic placement problems of Section 2 are attained on some appropriate eigenspaces of the Laplacian, while optimal multiway cuts are special values of the same quadratic objective function realized by piecewise constant vectors. Hence, the optimization problem, formu-

lated in terms of the Laplacian eigenvectors, is the continuous relaxation of the underlying maximum or minimum multiway cut problem. The objective functions defined on the partitions of the vertices are sometimes called modularities, see [45, 47].

3.1 Maximum, minimum, and ratio cuts of edge-weighted graphs

In the sequel, we will use the general framework of an edge-weighted graph introduced in Chapter 1. Let $G = (V, \boldsymbol{W})$ be an edge-weighted graph on n vertices with weight matrix $\boldsymbol{W}$ and generalized degrees $d_1, \ldots, d_n$. For a fixed integer $1 \leq k \leq n$, let $P_k = (V_1, \ldots, V_k)$ be a k-partition of the vertices, where the disjoint, non-empty vertex subsets $V_1, \ldots, V_k$ will be referred to as clusters or modules. The number of k-partitions is k^n, or the Stirling's partition number $\left\{ {n \atop k} \right\}$ if there are no empty clusters, see [39]. Let $\mathcal{P}_k$ denote the set of all k-partitions. Optimization over $\mathcal{P}_k$ is usually NP-complete, except some special classes of graphs.

To illustrate the relaxation technique, first we perform an easy estimation in the simplest case of $k = 2$. To this end, we introduce some definitions.

Definition 4 *The weighted cut between the non-empty vertex-subsets $U, T \subset V$ of the edge-weighted graph $G = (V, \boldsymbol{W})$ is*

$$w(U, T) = \sum_{i \in U} \sum_{j \in T} w_{ij}.$$

Note that $w(U, T)$ is the sum of the weights of edges connecting vertices of U and T. For now, U and T are not necessary disjoint subsets, though, in the sequel we will mainly use this notion for disjoint cluster pairs of a partition of V.

The so-called *maximum cut problem* looks for the maximum of the above weighted cut over 2-partitions of the vertices.

Definition 5 *The maximum cut of the edge-weighted graph $G = (V, \boldsymbol{W})$ is*

$$\mathrm{maxcut}(G) = \max_{U \subset V} w(U, \overline{U}), \tag{13}$$

where $\overline{U}$ is the complement of U in V.

The following statement is due to [**?**].

Proposition 2 *Let $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{n-1}$ denote the eigenvalues of the Laplacian $\boldsymbol{L}$ of the connected edge-weighted graph $G = (V, \boldsymbol{W})$ on n vertices. Then*

$$\mathrm{maxcut}(G) \leq \frac{n}{4} \lambda_{n-1}.$$

Proof 4 *Simple linear algebra guarantees that*

$$\lambda_{n-1} = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T \boldsymbol{L} \mathbf{x}. \tag{14}$$

Due to Equation (3), it follows that

$$\mathbf{x}^T \boldsymbol{L}\mathbf{x} = Q_1 = \sum_{i<j} w_{ij}(r_i - r_j)^2 \tag{15}$$

where the coordinates of the vector $\mathbf{x} = (r_1, \ldots, r_n)^T$ *are considered as one-dimensional representatives of the vertices.*

Let U^* *be the subset of* V *giving the maximum in (13). With it, let us define the following 1-dimensional representation of the vertices:*

$$r_i = \begin{cases} n - |U^*| & \text{if} \quad i \in U^* \\ -|U^*| & \text{if} \quad i \in \overline{U^*}. \end{cases} \tag{16}$$

Let us normalize the vector $\mathbf{x}$ *of coordinates* r_i*'s. Because of*

$$\|\mathbf{x}\|^2 = \sum_{i=1}^{n} r_i^2 = |U^*|(n - |U^*|)^2 + (n - |U^*|)|U^*|^2 = |U^*|(n - |U^*|)n,$$

the so obtained unit-norm vector $\tilde{\mathbf{x}}$ *will have coordinates* $\tilde{r}_i = \frac{r_i}{\sqrt{|U^*|(n-|U^*|)n}}$*. With this special vector, we will go below the maximum in (14), that is*

$$\lambda_{n-1} \geq \tilde{\mathbf{x}}^T \boldsymbol{L}\tilde{\mathbf{x}} = \sum_{i<j} w_{ij}(\tilde{r}_i - \tilde{r}_j)^2 = w(U^*, \overline{U^*}) \frac{(n - |U^*| + |U^*|)^2}{|U^*|(n - |U^*|)n},$$

where we also used (15). Therefore, utilizing that $w(U^*, \overline{U^*}) = \mathrm{maxcut}(G)$*,*

$$\mathrm{maxcut}(G) \leq \frac{|U^*|(n - |U^*|)}{n} \lambda_{n-1} \leq \frac{n}{4} \lambda_{n-1}.$$

We remark the following.

- Observe that the role of U^* and $\overline{U^*}$ is symmetric in the proof. The last estimation suggests that the upper bound for the maximum cut can be sharp for balanced partition of V, i.e. when U^* and $\overline{U^*}$ nearly have the same cardinality. If one looks for the maximum in (13) under the condition that the clusters of the 2-partition are of equal sizes, we get the maximum bisection problem which makes sense, of course, for even n.

- The proof also gives us a hint how to find the optimal U^*: the eigenvector $\mathbf{u}_{n-1}$ should be close to a piecewise constant vector over an appropriate 2-partition of the vertices. Later on, in this section, we will prove that it can be achieved by the 2-partition which is the output of the k-means algorithm with 2 clusters applied for the coordinates of the vector $\mathbf{u}_{n-1}$. With this spectral relaxation, the maximum cut can be approached in polynomial time for finding the largest eigenvalue and corresponding eigenvector of the Laplacian matrix. However, the exact solution of the maximum cut problem is NP-complete, even for simple graphs, called the maximum bipartite subgraph problem. Though, for some special classes of graphs, it is polynomially solvable (for example, for planar graphs).

The minimum cut of an edge-weighted graph is defined analogously, and for simple graphs, it is the edge-connectivity of [30]. The solution is often given

by an uneven 2-partition, for example, if there is an almost isolated vertex connected to few other vertices, it may form a cluster itself. To prevent this situation and rather find real-life loosely connected clusters, we require some balancing for the cluster sizes. For this purpose, we define the minimum cut and ratio cut for not only 2 clusters, but for any k-partition of the vertices and minimize it over the set of k-partitions. Roughly speaking, the minimum k-way cut minimizes the sum of the weights of intersecting edges (between the clusters), whereas the ratio cut (see e.g. [6, 17, 34]) in addition, penalizes partitions with very unequal cluster sizes.

Definition 6 *Let $G = (V, \boldsymbol{W})$ be an edge-weighted graph and $P_k = (V_1, \ldots, V_k)$ a k-partition of its vertices. The k-way cut of G corresponding to the k-partition P_k is*

$$\mathrm{cut}(P_k, G) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} w(V_a, V_b)$$

and the minimum k-way cut of G is

$$\mathrm{mincut}_k(G) = \min_{P_k \in \mathcal{P}_k} \mathrm{cut}(P_k, G). \tag{17}$$

Definition 7 *Let $G = (V, \boldsymbol{W})$ be an edge-weighted graph and $P_k = (V_1, \ldots, V_k)$ a k-partition of its vertices. The k-way ratio cut of G corresponding to the k-partition P_k is*

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

and the minimum k-way ratio cut of G is

$$g_k(G) = \min_{P_k \in \mathcal{P}_k} g(P_k, G).$$

The equivalent form in the formula of $g(P_k, G)$ is obtained by an easy calculation.

Proposition 3 *Let $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{n-1}$ denote the eigenvalues of the Laplacian $\boldsymbol{L}$ of the connected edge-weighted graph $G = (V, \boldsymbol{W})$ on n vertices. Then*

$$\mathrm{mincut}_2(G) \geq \frac{n-1}{n} \lambda_1. \tag{18}$$

Proof 5 *We will use a similar relaxation technique as in the Proof of Proposition 2. Here simple linear algebra guarantees that*

$$\lambda_1 = \min_{\substack{\|\mathbf{x}\|=1 \\ \mathbf{x}^T \mathbf{1}=0}} \mathbf{x}^T \boldsymbol{L} \mathbf{x}. \tag{19}$$

Let us recall that due to Equation (3), the relation

$$\mathbf{x}^T \boldsymbol{L} \mathbf{x} = Q_1 = \sum_{i<j} w_{ij} (r_i - r_j)^2$$

holds with the coordinates of the vector $\mathbf{x} = (r_1, \ldots, r_n)^T$ which are considered as one-dimensional representatives of the vertices.

Let U^* be the subset of V giving the minimum in (17) for $k = 2$. With it, we define the same representation (16) of the vertices as in the Proof of Proposition 2 and we use the same normalization, so that the vector $\tilde{\mathbf{x}} = (\tilde{r}_1, \ldots, \tilde{r}_n)$ satisfies $\|\tilde{\mathbf{x}}\| = 1$ and $\tilde{\mathbf{x}}^T \mathbf{1} = 0$, thanks to $\sum_{i=1}^n \tilde{r}_i = 0$. With this special vector, we will go above the minimum in (19), that is

$$\lambda_1 \leq \tilde{\mathbf{x}}^T \boldsymbol{L} \tilde{\mathbf{x}} = \sum_{i<j} w_{ij} (\tilde{r}_i - \tilde{r}_j)^2 = w(U^*, \overline{U^*}) \frac{(n - |U^*| + |U^*|)^2}{|U^*|(n - |U^*|)n}.$$

Therefore, utilizing that $w(U^*, \overline{U^*}) = \mathrm{mincut}_2(G)$,

$$\mathrm{mincut}_2(G) \geq \frac{|U^*|(n - |U^*|)}{n} \lambda_1 \geq \frac{n-1}{n} \lambda_1.$$

A number of comments are in order.

- For a simple graph G, [30] called the quantity $\mathrm{mincut}_2(G)$ the edge-connectivity of G, because it is equal to the minimum number of edges that should be removed to make G disconnected. He used the notation $e(G)$ for the edge-connectivity of the simple graph G, and $v(G)$ for its vertex-connectivity (minimum how many vertices should be removed to make G disconnected). In his papers [29, 30], the author proved that $v(G) \leq e(G)$, and for any graph G on n vertices, that differs from the complete graph C_n, the relation

  $$\lambda_1 \leq v(G) \leq e(G) \tag{20}$$

  holds, which gives a sharper estimate for λ_1 by $e(G)$ than Inequality (18). However, for the complete graph C_n, equality holds in (18). Indeed, on the one hand, Example (a) of Section 2.3 shows that $\lambda_1(C_n) = n$, and on the other hand, $e(C_n) = n - 1$, since the minimum cut is realized by a 2-partition consisting of a single vertex and all the other vertices.

- [30] also provided two lower estimates for λ_1 by $e(G)$:

  $$\lambda_1 \geq 2e(G)(1 - \cos\frac{\pi}{n}) \tag{21}$$

  and

  $$\lambda_1 \geq C_1 e(G) - C_2 d_{\max}, \tag{22}$$

  where $C_1 = 2(\cos\frac{\pi}{n} - \cos\frac{2\pi}{n})$, $C_2 = 2\cos\frac{\pi}{n}(1 - \cos\frac{\pi}{n})$, and $d_{\max} = \max_i d_i$ is the maximum vertex degree. Compared to (20), this estimation makes sense in the $n \geq 3$ case. The bound of (22) is better than that of (21) if and only if $e(G) \geq \frac{1}{2} d_{\max}$. The two estimates are equal and sharp for the path graph P_n with $e(G) = 1$ and $\lambda_1 = 2(1 - \cos\frac{\pi}{n})$, see Example (b) of Section 2.3. The path graph can be split into two clusters by removing any of its edges, however, we would not state that it has two underlying clusters. The ratio cut of P_n is minimized by removing the middle edge (for even n) or one of the middle edges (for odd n), thus, providing balanced clusters. Note that P_n is a tree, and hence, a sparse graph, that is not expected to have a remarkable cluster structure, unlike the dense graphs, incarnating a statistical sample.

- Because of this two-sided relation between λ_1 and $e(G)$, the smallest positive Laplacian eigenvalue of a connected graph is able to detect the strength of its connectivity; therefore, Fiedler called λ_1 the *algebraic connectivity* of G, and denoted it by $a(G)$. This relation between $e(G)$ and $a(G)$ was also recovered by [35]. Graphs with 'large' algebraic connectivity play an important role in communication networks, since the information goes through them very quickly. The so-called concentrators and expanders are graphs with high connectivity properties, see [3, 5, 12, 24].

- The proof of Proposition 3 gives us the following hint how to find the optimal U^*: the eigenvector $\mathbf{u}_1$ should be close to a piecewise constant vector over an appropriate 2-partition of the vertices. Note that because of its orthogonality to the vector $\mathbf{1}$, the vector $\mathbf{u}_1$ consists of both positive and negative coordinates, and [36] separated the two clusters according to the signs. In the sequel, we will use the k-means algorithm for this purpose, in a more general setup. Note that the vector $\mathbf{u}_1$ is frequently called *Fiedler-vector*.

To find the minima of the above multiway cuts over k-partitions is NP-complete. However, spectral techniques working in polynomial time are at our disposal. How accurately these minima can be approximated by means of spectral clustering depends on, how close the partition vectors can get to the eigenvectors corresponding to the k smallest Laplacian or normalized Laplacian eigenvalues. The measure of the closeness of the involved subspaces is the k-variance and it is minimized by the k-means algorithm. More precisely, we will apply the k-means algorithm for the optimal representatives of the vertices. For this purpose, instead of partitions, we will use partition vectors.

The k-partition P_k is uniquely determined by the $n \times k$ balanced partition matrix $\boldsymbol{Z}_k = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$, where the a-th *balanced k-partition vector* $\mathbf{z}_a = (z_{1a}, \ldots, z_{na})^T$ is the following:

$$z_{ia} = \begin{cases} \frac{1}{\sqrt{|V_a|}} & \text{if} \quad i \in V_a \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

The matrix $\boldsymbol{Z}_k$ is trivially suborthogonal, and the set of balanced k-partition matrices is denoted by $\mathcal{Z}_k^B$. With the special representation in which the representatives $\tilde{\mathbf{r}}_1, \ldots, \tilde{\mathbf{r}}_n \in \mathbb{R}^k$ are row vectors of $\mathcal{Z}_k^B$, the ratio cut of $G = (V, \boldsymbol{W})$ corresponding to the k-partition P_k (see Definition 7) can be rewritten as

$$g(P_k, G) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \|\tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_j\|^2 = \sum_{a=1}^{k} \mathbf{z}_a^T \boldsymbol{L} \mathbf{z}_a = \mathrm{tr}(\boldsymbol{Z}_k^T \boldsymbol{L} \boldsymbol{Z}_k). \tag{24}$$

We want to minimize it over balanced k-partition matrices $\boldsymbol{Z}_k \in \mathcal{Z}_k^B$.

Assume that G is connected. Let $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{n-1}$ denote the eigenvalues of its Laplacian matrix $\boldsymbol{L}$ with corresponding unit-norm, pairwise orthogonal eigenvectors $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$. Namely, $\mathbf{u}_0 = \frac{1}{\sqrt{n}}\mathbf{1}$.

Theorem 3 *For the minimum k-way ratio cut of the edge-wighted graph $G = (V, \boldsymbol{W})$ the lower estimate*

$$g_k(G) \geq \sum_{i=1}^{k-1} \lambda_i$$

holds.

Proof 6 *The discrete problem is relaxed to a continuous one. Let* $\mathbf{r}_1, \ldots, \mathbf{r}_n$ *denote* k*-dimensional representatives of the vertices. Let* $\boldsymbol{X}$ *be the* $n \times k$ *matrix with these representatives as row-vectors. The Representation theorem for edge-weighted graphs (see Theorem 1) states that*

$$\min_{\boldsymbol{X}^T\boldsymbol{X}=\boldsymbol{I}_k} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij}\|\mathbf{r}_i - \mathbf{r}_j\|^2 = \min_{\boldsymbol{X}^T\boldsymbol{X}=\boldsymbol{I}_k} \operatorname{tr}(\boldsymbol{X}^T\boldsymbol{L}\boldsymbol{X}) = \sum_{i=0}^{k-1} \lambda_i,$$

and equality is attained with $\boldsymbol{X} = (\mathbf{u}_0, \ldots, \mathbf{u}_{k-1})$.

As a balanced k*-partition matrix is a special suborthogonal matrix,*

$$g_k(G) = \min_{\boldsymbol{Z}_k \in \mathcal{Z}_k^B} \operatorname{tr}(\boldsymbol{Z}_k^T\boldsymbol{L}\boldsymbol{Z}_k) \geq \sum_{i=0}^{k-1} \lambda_i \tag{25}$$

and equality can be attained only in the $k = 1$ *trivial case, otherwise the eigenvectors* $\mathbf{u}_i$ $(i = 1, \ldots, k-1)$ *cannot be partition vectors, since their coordinates sum to 0 because of the orthogonality to the* $\mathbf{u}_0$ *vector.*

In the case of $k = 2$, in view of Theorem 3, $g_2(G)$ is bounded from below by λ_1, akin to the edge-connectivity of [30]. The proof also suggests that the quality of the above estimation depends on, how close the k bottom eigenvectors are to partition vectors. Concerning this issue and the optimum choice of k, we have the following considerations.

Let us recall the method of the **k-means clustering**. Given the points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ and an integer $1 < k < n$, we are looking for the k-partition of the index set $\{1, \ldots, n\}$ (or equivalently, the clustering of the points into k disjoint non-empty subsets) which minimizes the following k*-variance* of the points over all possible k-partitions $P_k = (C_1, \ldots, C_k)$:

$$S_k^2(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \min_{P_k} S_k^2(P_k; \mathbf{x}_1, \ldots, \mathbf{x}_n) = \min_{P_k} \sum_{a=1}^{k} \sum_{j \in C_a} \|\mathbf{x}_j - \mathbf{c}_a\|^2 \tag{26}$$

where $\mathbf{c}_a = \frac{1}{|C_a|} \sum_{j \in C_a} \mathbf{x}_j$ is the center of cluster a $(a = 1, \ldots, k)$.

In general, $d \leq k$, and they are much less than n. To find the global minimum is NP-complete, but the iteration of the k-means algorithm is capable to find a local minimum in polynomial time. Moreover, if S_{k+1}^2 is 'much less' than S_k^2, then even the global minimum is obtained in polynomial time with a PTAS (Polynomial Time Approximating Scheme). The vectors $\mathbf{c}_1, \ldots, \mathbf{c}_k$ are usually referred to as the *centroids* of the clusters, and in a more abstract formulation of the above optimization task, they are also looked for. Roughly speaking, starting with an initial clustering, the iteration of the simple k-means algorithm consists of the following two alternating steps. In the first step, fixing the clustering of the points, it finds the cluster centers (they will be the barycenters by the Steiner's theorem). In the second one, the algorithm relocates the points in such a way that it assigns a point to the cluster, the center of which is the closest to it (in case of ambiguity the algorithm chooses the smallest index such cluster). If there exists a well-separated k-clustering of the points (even the largest intra-cluster distance is smaller than the smallest inter-cluster one) the convergence

of the algorithm to the global minimum is proved, with a convenient starting. The algorithm runs faster if the separation between the clusters increases and an overall running time of $\mathcal{O}(kn)$ can be guaranteed.

Sometimes the points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are endowed with the positive weights $d_1, \ldots, d_n$, where without loss of generality $\sum_{i=1}^n d_i = 1$ can be assumed. In such cases the *weighted k-variance* of the points

$$\tilde{S}_k^2(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \min_{P_k} \tilde{S}_k^2(P_k; \mathbf{x}_1, \ldots, \mathbf{x}_n) = \min_{P_k} \sum_{a=1}^{k} \sum_{j \in C_a} d_j \|\mathbf{x}_j - \mathbf{c}_a\|^2 \qquad (27)$$

is considered, where $\mathbf{c}_a = \frac{1}{\sum_{j \in C_a} d_j} \sum_{j \in C_a} d_j \mathbf{x}_j$ is the weighted center of cluster a $(a = 1, \ldots, k)$. The above algorithm can be easily adapted to this situation. Note that $\tilde{S}_k^2(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ corresponds to the k-variance with respect to the distribution $d_1, \ldots, d_n$. In this contexts, $S_k^2(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is the special case when this law is uniform. Likewise, instead of L^2-distances, other kind of distance functions in the objective function may be used.

A well-known drawback of the k-means clustering is that the clusters need to be convex in order to achieve satisfactory results. That is, the k-means algorithm forms spherical clusters whether or not the underlying data distribution obeys this form. Otherwise, our data can be mapped into a feature space and we apply k-means clustering for the mapped data which already have this spherical structure (see reproducing kernel Hilbert spaces in the next lesson).

Now let us denote by $S_k^2(P_k; \boldsymbol{X}^*)$ the k-variance of the optimal vertex representatives comprising row vectors of $\boldsymbol{X}^* = (\mathbf{u}_0, \ldots, \mathbf{u}_{k-1})$ in the clustering P_k of the vertices. Let $\mathbf{y}_1, \ldots \mathbf{y}_k$ denote an orthonormal set of eigenvectors corresponding to the k smallest eigenvalues of the Laplacian $\boldsymbol{L}$. Consider the squared distances between $\mathbf{y}_i$'s and F, where F is the subspace spanned by partition vectors of (23). Then

$$S_k^2(P_k; \boldsymbol{X}^*) = \sum_{i=1}^{k} \mathrm{dist}^2(\mathbf{y}_i, F).$$

Indeed, $\mathrm{dist}^2(\mathbf{y}_i, F)$ is the minimum squared distance between $\mathbf{y}_i$ and F. In view of the Steiner's theorem, the minimum is attained with the piecewise constant vector of coordinates having at most k different values $c_{1i}, \ldots, c_{ki}$ over the sets $V_1, \ldots, V_k$ of the underlying k-partition. Namely, if $i \in V_a$, then the ith coordinate of the distance-minimizing piecewise constant vector is $c_{ai} = \frac{1}{|V_a|} \sum_{j \in V_a} y_{ji}$, yielding

$$\mathrm{dist}^2(\mathbf{y}_i, F) = \sum_{a=1}^{k} \sum_{j \in V_a} (y_{ji} - c_{ai})^2,$$

where y_{ji} is the jth coordinate of the vector $\mathbf{y}_i$. By summing it for $i = 1, \ldots, k$ and rearranging the finite summation, $\sum_{i=1}^k \mathrm{dist}^2(\mathbf{y}_i, F)$ equals $S_k^2(P_k; \boldsymbol{X}^*)$ with cluster centers $\mathbf{c}_a = (c_{a1}, \ldots, c_{ak})$, $a = 1, \ldots, k$.

Therefore, the k-variance also minimizes the distance between the eigensubspaces spanned by the eigenvectors corresponding to the k smallest Laplacian eigenvalues and that of the partition vectors. Here the number of clusters is the same as the dimension of the representatives.

3.2 Multiway cuts of hypergraphs

[It can be skipped.]

For hypergraphs, different kinds of multiway cuts can be defined. By Equation (9), to any hypergraph $H = (V, E)$ an edge-weighted graph $G = (V, \boldsymbol{W})$ can be assigned such that the Laplacian of H is half of the Laplacian of G. Accordingly, minimum multiway cuts and ratio cuts are defined for H such that they can be directly related to the same quantities of G. Since the formula of (9) takes into consideration the cardinalities of the hyperedges, the minimum multiway and ratio cuts of a hypergraph depend on these cardinalities. Therefore, a cardinality-free version of them, as well as a version which takes into consideration how many vertices of the individual clusters a cut-hyperedge contains, will also be introduced. We will discuss relations between these cuts and the Laplacian spectrum of H.

Definition 8 *Let $H = (V, E)$ be a hypergraph and $P_k = (V_1, \ldots, V_k)$ a k-partition of its vertices. The k-way cut of H corresponding to the k-partition P_k is*

$$\mathrm{cut}(P_k, H) = \sum_{e \in E} \frac{1}{|e|} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} |e \cap V_a| \cdot |e \cap V_b|$$

and the minimum k-way cut of H is

$$\mathrm{mincut}_k(H) = \min_{P_k \in \mathcal{P}_k} \mathrm{cut}(P_k, H). \tag{28}$$

Definition 9 *Let $H = (V, E)$ be a hypergraph and $P_k = (V_1, \ldots, V_k)$ a k-partition of its vertices. The k-way ratio cut of H corresponding to the k-partition P_k is*

$$g(P_k, H) = \sum_{e \in E} \frac{1}{|e|} \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left(\frac{1}{|V_a|} + \frac{1}{|V_b|} \right) |e \cap V_a| \cdot |e \cap V_b|$$

and the minimum k-way ratio cut of H is

$$g_k(H) = \min_{P_k \in \mathcal{P}_k} g(P_k, H).$$

Definition 10 *Let $H = (V, E)$ be a hypergraph and $P_k = (V_1, \ldots, V_k)$ a k-partition of its vertices. The k-sector of H corresponding to the k-partition P_k is the following set of its hyperedges:*

$$\mathrm{sector}(P_k, H) = \{e \in E : \textit{there exist} i \neq j \textit{ s. t. } e \cap V_i \neq \emptyset \textit{ and } e \cap V_j \neq \emptyset\}.$$

The cardinality of the minimum k-sector of H is

$$\theta_k(H) = \min_{P_k \in \mathcal{P}_k} |\mathrm{sector}(P_k, H)|. \tag{29}$$

Note that the k-partition P_k defines the coloring c of the vertices with k different colors: $c(v) = i$ if $v \in V_i$. In terms of this coloring, $\mathrm{sector}(P_k, H)$ consists of the multicolored edges of H (i.e. hyperedges of H having at least two vertices of different colors).

Remark 1 *The above quantities are trivially monotonous in the sense that*

$$\begin{aligned}
&\mathrm{mincut}_2(H) \leq \mathrm{mincut}_3(H) \leq \cdots \leq \mathrm{mincut}_n(H),\\
&g_2(H) \leq g_3(H) \leq \cdots \leq g_n(H),\\
&\theta_2(H) \leq \theta_3(H) \leq \cdots \leq \theta_n(H) = |E|.
\end{aligned}$$

Let $\tilde{L}$ be the Laplacian of the hypergraph H defined in Definition 3. For the bottom of its spectrum, in [17], the following two-sided estimation was proved.

Theorem 4 *Let $0 = \tilde{\lambda}_0 < \tilde{\lambda}_1 \leq \cdots \leq \tilde{\lambda}_{n-1}$ be the Laplacian eigenvalues of the connected hypergraph H, and k be a fixed integer $(2 \leq k \leq n)$. Then*

$$c_n \theta_k(H) \leq \sum_{i=1}^{k-1} \tilde{\lambda}_i \leq g_k(H)$$

holds with $c_n = \frac{6}{n(n^2-1)}$.

In the case of $k = 2$, a more precise estimation for λ_1 can be given.

Proposition 4 *For the smallest positive Laplacian eigenvalue of the connected hypergraph H the following lower estimate holds:*

$$\tilde{\lambda}_1 \geq \begin{cases} 2(1-\cos\frac{\pi}{n})\mathrm{mincut}_2(H) & \text{if} \quad 0 \leq \mathrm{mincut}_2(H) \leq \frac{1}{2} d_{v\max} \\ C_1 \mathrm{mincut}_2(H) - C_2 d_{v\max} & \text{if} \quad \frac{1}{2} d_{v\max} < \mathrm{mincut}_2(H) \end{cases}$$

where $d_{v\max} = \max_i d_{vi}$ is the maximum vertex valence of H, whereas the constants C_1 and C_2 are the same as in Equation (22).

The proof, to be found in [17], is analogous to that of the proof of Equation (22) of [30].

The upper bound in Theorem 4 shows that the existence of k relatively small eigenvalues is a necessary condition for the existence of a good classification of H's vertices (with a small minimal k-way partition cut). Thus, the spectrum gives us some idea about the choice of the number k of the clusters for which good coloring may exist. But the spectrum itself does not divulge anything about the optimal k-partition, moreover it does not give a sufficient condition for the existence of a good clustering. The lower bound in Theorem 4 depends on the constant c_n, and there are graphs for which the lower bound is attained in order of magnitude, e.g. for lattices and spiders (see Examples (d) and (e) in Section 2.3), which can not be classified into k clusters in a meaningful way.

Now we want to recognize optimal k-partitions by means of classification of k-dimensional representatives of the vertices in an optimal k-dimensional Euclidean representation of the hypergraph. The classification is performed by the k-means algorithm. We will be confined to the case, when a 'very' well-separated k-partition of the above k-dimensional points exists.

Definition 11 *A k-partition $P_k = (V_1, \ldots, V_k)$ is called well-separated k-partition of the vertex-set V in the k-dimensional Euclidean representation $\mathbf{r}_1, \ldots, \mathbf{r}_n$ of the vertices if for the coloring c, corresponding ing to P_k, the relation*

$$\frac{\min_{c(v_i) \neq c(v_j)} \|\mathbf{x}_i - \mathbf{x}_j\|}{\max_{c(v_i) = c(v_j)} \|\mathbf{x}_i - \mathbf{x}_j\|} \geq 1$$

holds.

Theorem 5 *Assume that for some integer $1 < k < n$ there exists a well-separated k-partition of the optimal representatives of the vertices of $H = (V, E)$, for the clusters of which the diameters are at most ε, where $\varepsilon < \frac{1}{2\sqrt{n}}$. Then*

$$g_k(H) \leq q^2 \sum_{i=1}^{k-1} \tilde{\lambda}_i$$

where $q = 1 + \frac{2\varepsilon}{1-\varepsilon\sqrt{n}}$.

For the proof see [17]. Comparing the results of Theorems 4 and 5, under the constraints of the latter one, we obtain that

$$\sum_{i=1}^{k-1} \tilde{\lambda}_i \leq g_k(H) \leq q^2 \sum_{i=1}^{k-1} \tilde{\lambda}_i, \quad \text{where} \quad 1 < q < 2.$$

This means that, provided ε is less than $\frac{1}{2\sqrt{n}}$, then q is at most 2, and hence, $g_k(H)$ and $\sum_{i=1}^{k-1} \tilde{\lambda}_i$ differ at most by a factor of 4.

References

[1] Achlioptas D and McSherry F 2007 Fast computation of low-rank matrix approximations. *J. ACM* **54** (2), Article 9.

[2] Aldous DJ 1981 Reresentations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** (4), 581–598.

[3] Alon N 1986 Eigenvalues and expanders. *Combinatorica* **6** (2), 83–96.

[4] Alon N Milman VD 1985 λ_1, isoperimetric inequalities for graphs and superconcentrators. *J. Comb. Theory Ser. B* **38**, 73–88.

[5] Alon N Milman VD 1987 Better expanders and superconcentrators. *J. Algorithms* **8**, 337–347.

[6] Alpert CJ and Yao S.-Z 1995 Spectral partitioning: the more eigenvectors, the better. In *Proc. 32nd ACM/IEEE International Conference on Design Automation* (Preas BT, Karger PG, Nobandegani BS and Pedram M eds), pp. 195–200. Association of Computer Machinery, New York.

[7] Anderson WN and Morley TD 1985 Eigenvalues of the Laplacian of a graph. *Linear Multilinear Algebra* **18**, 141-145.

[8] Ando T 1987 *RKHS and quadratic inequalities.* Lecture Notes, Hokkaido University, Sapporo, Japan.

[9] Aronszajn N 1950 Theory of Reproducing Kernels. *Trans. Am. Math. Soc.* **68**, 337–404.

[10] Austin T 2008 On exchangeable random variables and the statistics of large graphs and hypergraphs. *Prob. Surv.* **5**, 80–145.

[11] Azran A and Ghahramani Z 2006 Spectral methods for automatic multiscale data clustering. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), New York NY* (Fitzgibbon A, Taylor CJ and Lecun Y eds), pp. 190–197. IEEE Computer Society, Los Alamitos, California.

[12] Babai L and Szegedy M 1992 Local expansion of symmetric graphs. *Comb. Probab. Comput.* **1**, 1–11.

[13] Bach FR and Jordan MI 2002 Kernel Independent Component Analysis. *J. Mach. Learn. Res.* **3**, 1–48.

[14] Baker C 1977 *The numerical treatment of integral equations.* Clarendon Press.

[15] Bengio Y, Delalleau O, Le Roux N, Paiement J-F, Vincent P and Ouimet M 2004 Learning eigenfunctions links spectral embedding and kernel PCA. *Neural Comput.* **16**, 2197–2219.

[16] Biggs NL 1974 *Algebraic Graph Theory.* Cambridge Univ. Press, Cambridge.

[17] Bolla M 1993 Spectra and Euclidean representation of hypergraphs. *Discret. Math.* **117**, 19–39.

[18] Bolla M and Tusnády G 1994 Spectra and Optimal Partitions of Weighted Graphs. *Discret. Math.* **128**, 1–20.

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

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

[21] Bolla M 2011 Penalized versions of the Newman–Girvan modularity and their relation to multi-way cuts and k-means clustering. *Phys. Rev. E* **84**, 016108.

[22] Borgs C Chayes JT Lovász L T.-Sós V and Vesztergombi K 2008 Convergent Sequences of Dense Graphs I: Subgraph Frequences, Metric Properties and Testing. *Advances in Math.* **219**, 1801–1851.

[23] Breiman L and Friedman JH 1985 Estimating optimal transformations for multiple regression and correlation. *J. Am. Stat. Assoc.* **80**, 580–619.

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

[25] Cvetković DM, Doob M and Sachs H 1979 *Spectra of Graphs.* Academic Press, New York.

[26] Cvetković DM, Rowlinson P and Simić S 1997 *Eigenspaces of Graphs*, Encyclopedia of Mathematics and Its Applications Vol. 66. Lavoisier Libraire, S.A.S. France.

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

[28] Diaconis P and Janson S 2008 Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (VII. Ser.) **28**, 33–61.

[29] Fiedler M 1972 Bounds for eigenvalues of doubly stochastic matrices. *Linear Algebra Appl.* **5** (98), 299–310.

[30] Fiedler M 1973 Algebraic connectivity of graphs. *Czech. Math. J.* **23** (98), 298–305.

[31] Frieze A, Kannan R and Vempala S 1998 Fast Monte-Carlo Algorithms for finding low-rank approximations. In *Proc. 39th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1998), Palo Alto, California*, pp. 370–386. IEEE Computer Society, Los Alamitos.

[32] Gebelein H 1941 Das statistische Problem der Korrelation als Variations und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung. *Z. Angew. Math. Mech.* **21**, 364–379.

[33] Girosi F, Jones M and Poggio T 1995 Regularization theory and neural networks architectures. *Neural Comput.* **7**, 219–269.

[34] Hagen L and Kahng AB 1992 New spectral methods for ratio cut partitioning and clustering. *IEEE Trans. Comput. Aided Des.* **11**, 1074–1085.

[35] Hoffman AL 1969 The change in the least eigenvalue of the adjacency matrix of a graph under imbedding. *SIAM J. Appl. Math.* **17**, 664–671.

[36] Juhász F and Mályusz K 1980 Problems of cluster analysis from the viewpoint of numerical analysis. In *Numerical Methods, Coll. Math. Soc. J. Bolyai* (ed. Rózsa P), Vol 22, pp. 405–415. North-Holland, Amsterdam.

[37] Kelmans AK 1967 Properties of the charasteristic polynomial of a graph. *Cibernetics in the Science of Communication* **4**, Energija, Moskva–Leningrad, 27–41 (in Russian).

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

[39] Lovász L 1993 *Combinatorial Problems and Exercises*. Akadémiai Kiadó–North Holland, Budapest–Amsterdam.

[40] Liotta G ed. 2004 *Graph drawing*. Lecture Notes in Computer Science **2912**. Springer.

[41] Maas C 1987 Transportation in graphs and the admittance spectrum. *Discret. Appl. Math.* **16**, 31–49.

[42] Merris R 1987 Characteristic vertices of trees. *Linear Multilinear Algebra* **22**, 115–131.

[43] Mohar B 1988 Isoperimetric inequalities, growth and the spectrum of graphs. *Linear Algebra Appl.* **103**, 119–131.

[44] Mohar B 1991 The Laplacian spectrum of graphs. In *Graph Theory, Combinatorics, and Applications* Vol 2 (Alavi Y, Chartrand G, Oellermann OR and Schwenk AJ eds), pp. 871–898. Wiley.

[45] Newman MEJ and Girvan M 2004 Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113.

[46] Newman MEJ 2004 Detecting community structure in networks. *Eur. Phys. J. B* **38**, 321–330.

[47] Newman MEJ 2010 *Networks, An Introduction.* Oxford University Press.

[48] Ng AY Jordan MI and Weiss Y 2001 On spectral clustering: analysis and an algorithm. In *Proceedings of the 14th Neural Information Processing Systems Conference, NIPS 2001* (Dietterich TG, Becker S and Ghahramani Z eds), pp. 849–856. MIT Press, Cambridge, USA.

[49] Parzen E 1963 Probability density functionals and reproducing kernel Hilbert spaces. In *Proceedings of the Symposium on Time Series Analysis* (ed. Rosenblatt M), pp. 155–169. Brown University, Providence, RI, USA.

[50] Pisanski T and Shawe-Taylor J 2000 Characterizing graph drawing with eigenvectors. *J. Chem. Inf. Comput. Sci.* **40**, 567–571.

[51] Reisen K and Bunke H 2010 *Graph Classification and Clustering Based on Vector Space Embedding.* World Scientific.

[52] Rényi A 1959 On measures of dependence. *Acta Math. Acad. Sci. Hung.* **10**, 441–451.

[53] Rényi A 1959 New version of the probabilistic generalization of the large sieve. *Acta Math. Acad. Sci. Hung.* **10**, 218–226.

[54] Riesz F 1907 Sur une espéce de géométrie analytique des systémes de fonctions sommables. *C. R. Acad. Sci. Paris* **144**, 1409–1411.

[55] Riesz F 1909 Sur les opérations fonctionnelles linéaires. *C. R. Acad. Sci. Paris* **149**, 974–977.

[56] Riesz F and Sz.-Nagy B 1952 *Leçons d'analyse fonctionnelle.* Academic Publishing House, Budapest.

[57] Shawe-Taylor J and Cristianini N 2004 *Kernel Methods for Pattern Analysis.* Cambridge Univ. Press, Cambridge.

[58] Shi J and Malik J 2000 Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** (8), 888–905.

[59] Schölkopf B Smola AJ and Müller K-R 1998 Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10**, 1299–1319.

[60] Schölkopf B and Smola AJ 2002 *Learning with Kernels.* MIT Press, Cambridge, USA.

[61] Von Luxburg U 2006 A tutorial on spectral clustering. *Stat. Comput.* **17**, 395–416.

[62] Von Luxburg U Belkin M and Bousquet O 2008 Consistency of spectral clustering. *Ann. Stat.* **36**, 555–586.

[63] White S and Smyth P 2005 A spectral clustering approach to find communities in graphs. In *Proc. SIAM International Conference on Data Mining* (Kargupta H, Srivastava J, Kamath Ch and Goodman A), pp. 76–84. SIAM, Newport Beach.

[64] Yan S, Xu D, Zhang B, Zhang H-J, Yang Q and Lin S 2007 Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **29** (1), 40–48.