æ

ISOPERIMETRIC PROPERTIES OF WEIGHTED GRAPHS RELATED TO THE LAPLACIAN SPECTRUM AND CANONICAL CORRELATIONS

Marianna Bolla, Gábor Molnár-Sáska

Institute of Mathematics, Budapest University of Technology and Economics
1521 Budapest, P.O.Box 91. Bldg. H. V/7. Hungary,
e-mail: marib@math.bme.hu, molnar@math.bme.hu

Abstract
The relation between isoperimetric properties and Laplacian spectra of weighted graphs is investigated. The vertices are classified into k clusters with "few" inter-cluster edges of "small" weights (area) and "similar" cluster sizes (volumes). For $k = 2$ the Cheeger constant represents the minimum requirement for the area/volume ratio and it is estimated from above by $\sqrt{\lambda_1(2-\lambda_1)}$, where λ_1 is the smallest positive eigenvalue of the weighted Laplacian. For $k > 2$ we define the k-density of a weighted graph that is a generalization of the Cheeger constant and estimated from below by $\sum_{i=1}^{k-1} \lambda_i$ and from above by $c^2 \sum_{i=1}^{k-1} \lambda_i$, where $0 < \lambda_1 \le \cdots \le \lambda_{k-1}$ are the smallest Laplacian eigenvalues and the constant $c > 1$ depends on the metric classification properties of the corresponding eigenvectors. Laplacian spectra are also related to canonical correlations in a probabilistic setup.

1. Introduction

Let $G = (V, \mathbf{W})$ be a weighted graph, where $V = \{1, \dots, n\}$ is the vertex set and $\mathbf{W}$ is the $n \times n$ symmetric weight matrix of the edges with nonnegative entries and zero diagonal. We assume that $\sum_{i=1}^n \sum_{j=1}^n w_{ij} = 1$ (this restriction does not influence the eigenvalues investigated later on). We set $d_i = \sum_{j=1}^n w_{ij}$, $i = 1, \dots, n$. Let the $n \times n$ diagonal matrix $\mathbf{D}$ contain the entries $d_1, \dots, d_n$ in its main diagonal. If the graph G is connected then clearly, $0 < d_i < 1/2$ for all i.

The following optimization problems arise when we want to classify the vertices into k clusters.

(a) *Embedding problem* (discussed thoroughly in [2]): We are looking for optimal k-dimensional representatives $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^k$ $(1 < k \le n)$ of the vertices such that the minimum of the quadratic objective function

$$L(\mathbf{X}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^2 w_{ij} \tag{1.1}$$

is attained on the constraint

$$\mathbf{X}\mathbf{D}\mathbf{X}^T = \mathbf{I}_k, \tag{1.2}$$

where the $k \times n$ matrix $\mathbf{X} = (\mathbf{x}_1 \dots \mathbf{x}_n)$ contains the representatives in its columns. The solution of the problem is as follows. The matrix $\mathbf{C} = \mathbf{D} - \mathbf{W}$ is the Laplacian, while $\mathbf{C}_D = \mathbf{D}^{-1/2}\mathbf{C}\mathbf{D}^{-1/2} = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is the weighted Laplacian of G. Both are symmetric, positive semidefinite matrices, the number of their zero eigenvalues is equal to the number of connected components of G. Suppose that G is connected. Then the eigenvalues of $\mathbf{C}_D$ are

$$0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \dots \le \lambda_{n-1} \le 2$$

in increasing order with corresponding orthonormal system of eigenvectors (column vectors)

$$\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n-1}.$$

If $\lambda_{k-1} < \lambda_k$ then the minimum of (1.1) on (1.2) is attained with the choice $\mathbf{X} = (\mathbf{u}_0 \dots \mathbf{u}_{k-1})^T \mathbf{D}^{-1/2}$ and the minimum is equal to $\sum_{j=1}^{k-1} \lambda_j$. If the zero is a single eigenvalue, then the first coordinate of each representative is 1 and hence, the representatives are in fact of $k-1$ dimension.

(b) *Probabilistic setup* (conditional expectation and canonical correlations): Now $\mathbf{W}$ is regarded as a joint distribution on a finite product probability space. It will be shown that the eigenvalues of the conditional expectation operator are the numbers $1 > r_1 \ge \dots \ge r_{n-1}$, where $r_i = 1 - \lambda_i$ and they are the so-called canonical correlations of the correspondence analysis. Here r_1 maximizes the correlation between the two underlying discrete variables and it is the maximal correlation introduced by Rényi [14]. The objective functions in **(a)** and **(b)** are closely related, as well as the L_2-norm minimum and maximum obtained from them.

(c) *Isoperimetry* (k=2): The Cheeger constant introduced in [6] finds L_1-norm minimum of an analogous objective function in terms of the vertex representation in **(a)**. The minima of **(a)** and **(c)** are compared in Section 3.

(d) *Probabilistic setup* (conditional probability): With the notation of **(b)** we find the minimum of some conditional probabilities. An interesting result follows in comparison to **(b)** analogously to the results obtained as **(a)** and **(c)** are compared.

(e) *Isoperimetry* (generalization of the Cheeger constant for $k > 2$): As an extension of **(c)** we define the weighted k-density ρ_k and estimate it with the sum of the $k-1$ smallest positive Laplacian eigenvalues.

The main purpose of the paper is to summarize and compare the above frequently used optimization problems. New results are the comparisons of different types of problems and the part **(e)**.

2. Conditional expectation and canonical correlations

Let $\mathbf{W}$ denote the joint probability distribution of two discrete random variables taking on at most n different values. (Tipically, they are categorical variables, such as eye-color, hair-color, that have no preassigned values.) Suppose that the $n \times n$ matrix $\mathbf{W}$ is symmetric and it has zero diagonal. Therefore, the two marginal distributions are the same: $d_1, \dots, d_n$, the diagonal of the matrix $\mathbf{D}$ of Section 1. Sometimes, D also refers to the marginal. With this notations, the matrix form of the operator taking conditional expectation between the Hilbert spaces $L_2(V, \mathcal{A}, D)$ and $L_2(V, \mathcal{A}', D)$ with respect to the joint distribution $\mathbf{W}$ is just $\mathbf{I}_n - \mathbf{C}_D$. (Here $L_2(V, \mathcal{A}, D)$ is the set of random variables with finite variance on the probability space $(V, \mathcal{A}, D)$, where both $\mathcal{A}$ and $\mathcal{A}'$ are the induced σ-algebras by V, but we are going to use different notation for the two copies belonging to the two marginals.) To show this, let X be a random variable on $L_2(V, \mathcal{A}, D)$ taking on values $x_1, \dots, x_n$ with respective probabilities $d_1, \dots, d_n$ and satisfying the conditions:

$$\mathrm{E}(X) = \sum_{i=1}^n x_i d_i = 0, \qquad \mathrm{Var}\,(X) = \sum_{i=1}^n x_i^2 d_i = 1. \tag{2.1}$$

(E and Var stand for the expectation and variance of a random variable, respectively.) Put $Y = \mathrm{E}_W(X|\mathcal{A}')$ and suppose that Y takes on values $y_1, \dots, y_n$. Then

$$y_i = \sum_{j=1}^n \frac{w_{ij}}{d_i} x_j = \sum_{j=1}^n \frac{w_{ij}}{\sqrt{d_i}\sqrt{d_j}} x_j \frac{\sqrt{d_j}}{\sqrt{d_i}},$$

and hence,

$$\sqrt{d_i} y_i = \sum_{j=1}^n \frac{w_{ij}}{\sqrt{d_i}\sqrt{d_j}} (\sqrt{d_j} x_j).$$

With the notations $\mathbf{x} = (x_1, \dots, x_n)^T$, $\mathbf{u} = \mathbf{D}^{1/2}\mathbf{x}$, $\mathbf{y} = (y_1, \dots, y_n)^T$, and $\mathbf{v} = \mathbf{D}^{1/2}\mathbf{y}$ we have

$$\mathbf{v} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}\mathbf{u},$$

where $\|\mathbf{u}\|^2 = \sum_{i=1}^n (\sqrt{d_i} x_i)^2 = 1$ and obviously, $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} = \mathbf{I}_n - \mathbf{C}_D$.

The largest eigenvalue of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ is 1 with eigenvector $(\sqrt{d_1}, \dots, \sqrt{d_n})^T$ (its Euclidean norm is 1), while the other eigenvalues are less then 1 and the corresponding eigenvectors are orthogonal to $(\sqrt{d_1}, \dots, \sqrt{d_n})^T$. So, denoting by $\mathbf{u}$ such an eigenvector, the coordinates of $\mathbf{x} = D^{-1/2}\mathbf{u}$ will satisfy the conditions (2.1), in other words, $\mathbf{x}$ is a harmonic eigenvector.

Let

$$1 = r_0 > r_1 \geq r_2 \geq \cdots \geq r_{n-1} \geq -1$$

denote the eigenvalues of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ in decreasing order.

It is easy to see that $\lambda_i = 1 - r_i$ $(i = 0, \ldots, n-1)$. Note that r_1 is the maximum correlation. The corresponding eigenvectors are the same as in Section 2: $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ in this order. In [3], it was shown that r_i-s are correlation-like quantities. For them, $|r_i| \leq 1$ holds, therefore $0 \leq \lambda_i \leq 2$, $i = 0, \ldots, n-1$. In case of 0-1 weights $\lambda_{n-1} = 2$ is attained if and only if G is bipartite.

The second largest eigenvalue r_1 and the corresponding eigenvector $\mathbf{u}_1$ can be found by the spectral decomposition of the matrix $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$. Finally, the harmonic eigenvector is obtained by the transformation $\mathbf{D}^{-1/2}\mathbf{u}_1$. This eigenvector of the conditional expectation operator can be also determined by the so-called ACE (Alternating Conditional Expectation) algorithm of [4]. The algorithm developed by Breiman and Friedman starts with an X satisfying (2.1), it calculates $Y = \mathrm{E}_W(X|\mathcal{A}')$, and statndardizes it. Let Y' denote the standardized Y, take $X' = \mathrm{E}_W(Y'|\mathcal{A})$, and so on ... If $\mathbf{W}$ is symmetric then the series of the so constructed consecutive standardized Xs and Ys both converge to the harmonic eigenvector belonging to r_1.

Now, let us investigate the objective functions the maximum/minimum of which is found by the algorithm. In [3], it was proved that the following two tasks are equivalent:

$$\sup_{\substack{X,Y \in L_2(V,\mathcal{A},D) \text{ i.d.} \\ \mathrm{E}(X)=0 \\ \mathrm{Var}\,(X)=1}} \mathrm{E}_W XY = r_1, \tag{2.2}$$

where i.d. means identically distributed random variables. The supremum is attained for the X, Y pair both taking on values $x_1, \ldots, x_n$ (the coordinates of the harmonic eigenvector of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$) with probabilities $d_1, \ldots, d_n$, and the joint distribution of them is $\mathbf{W}$, while r_1 is the symmetric maximal correlation.

$$\begin{aligned} &\inf_{\substack{X,Y \in L_2(V,\mathcal{A},D) \text{ i.d.} \\ \mathrm{E}(X)=0}} \frac{\mathrm{E}_W\|X - X'\|^2}{\mathrm{Var}\,(X)} = \inf_{\substack{X,Y \in L_2(V,\mathcal{A},D) \text{ i.d.} \\ \mathrm{E}(X)=0 \\ \mathrm{Var}\,(X)=1}} \mathrm{E}_W\|X - X'\|^2 = \\ &\inf_{\substack{X,Y \in L_2(V,\mathcal{A},D) \text{ i.d.} \\ \mathrm{E}(X)=0 \\ \mathrm{Var}\,(X)=1}} \sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{x}_i - \mathbf{x}_j\|^2 w_{ij} = 2\lambda_1, \end{aligned} \tag{2.3}$$

where the infimum is attained for the same X as the supremum in (2.2), and $\lambda_1 = 1 - r_1$ is the smallest positive eigenvalue of $\mathbf{C}_D$. Here both X and Y take on values $x_1, \ldots, x_n$, as in (2.2), their joint distribution being $\mathbf{W}$. λ_1 also gives the minimum of $L(\mathbf{X})$ in **(a)**.

In fact, the equivalence of the two statements can be proven in a more general form for an appropriate joint distribution (the matrix $\mathbf{W}$ is neither symmetric nor square, so the two marginals are different). The maximum possible correlation of the zero-expectation, one-variance factor pairs is also called maximal correlation. Rényi [14] uses the notion for continuous distributions, too. More exactly, he looks for maximum correlated functions of two random variables with a given joint distribution.

The process can be continued. On the one hand, after finding the maximum correlated pair, we look for the next pair, that are uncorrelated to the previous ones, and have the possible largest correlation, etc. In this way, we can choose as many pairs as the rank of the matrix containing the joint probabilities (including the trivial pair with 1 correlation). This process is the so-called correspondence analysis used in multivariate statistics. The maximum correlated pairs are called correspondence factors. With formulas

$$\sup_{\substack{X,Y\in L_2(V,\mathcal{A},D)\text{ i.d.}\\ \mathrm{E}(X)=0\\ \mathrm{Var}\,(X)=1\\ \mathrm{Cov}\,(X,X_i)=0,\,(i=1,\dots,k-1)\\ \mathrm{Cov}\,(Y,Y_i)=0,\,(i=1,\dots,k-1)}} \mathrm{E}_W XY = r_k \qquad (k=1,\dots,\mathrm{rank}\,(\mathbf{W})), \tag{2.4}$$

where the pair (X_i, Y_i) gives the supremum in the ith step (X_0 and Y_0 are constantly 1). In the kth step the supremum is attained for the pair (X_k, Y_k) taking on the same values with probabilities $d_1, \dots, d_n$, and the joint distribution of them is $\mathbf{W}$, while r_k is the kth canonical correlation.

The corresponding infimum problem is just (1.1) based on the k-dimensional representatives $\mathbf{x}_i$s.

On the other hand, the values taken on by the correspondence factor pairs also give rise to an Euclidean representation of the variable categories. If the first k factors – with the largest possible correlations – are used for the representation then we obtain a k-dimensional representation of the variable categories that is equivalent to the k-dimensional representation of the weighted graph G in part **(a)** of Section 1. Numerically, the transformed (by $\mathbf{D}^{-1/2}$) eigenvectors corresponding to the $k-1$ smallest positive eigenvalues of the weighted Laplacian $\mathbf{C}_D$ are used for the representation that are identical to the harmonic eigenvectors of the conditional expectation operator belonging to its $k-1$ largest (excluding the number 1) eigenvalues. This representation is widely used in the correspondence analysis for the visual illustration of the variable categories.

3. The isoperimetric number and λ_1.

Now, let us turn to L_1-based minima. For graphs, there are combinatorial measures (e.g., edge density in [12], minimal weighted cut in [3], isoperimetric number and Cheeger constant in [1], [5], [6], [7], [9], [10]) that indicate the two-clustering properties of the graph. The so-called Cheeger constant defined for the weighted graph $G = (V, \mathbf{W})$ can be best compared to the smallest positive eigenvalue of the weighted Laplacian $\mathbf{C}_D$:

$$h = \min_{\substack{U \subset V \\ \mathrm{Vol}\,(U) \le 1/2}} \frac{\sum_{i \in U} \sum_{j \in \bar{U}} w_{ij}}{\mathrm{Vol}\,(U)} \tag{3.1}$$

where $\mathrm{Vol}\,(U) = \sum_{i \in U} d_i$ is the weight sum of edges with at least one endpoint in U, while the numerator contains the weight sum of edges with one endpoint in U and other endpoint in $\bar{U}$. Therefore, $h \le 1$ is trivial. Note that h can be "small" if "low-weight" edges connect together two disjoint vertex-sets with "not significantly" differing volumes; therefore, h reflects the clustering ability of the graph.

In the framework of joint distributions, h can be formulated by

$$h = \min_{\substack{B \subset \mathbb{R} \text{ Borel-set} \\ X, Y \in L_2(V, \mathcal{A}, D) \text{ i.d.} \\ \mathrm{P}_D(X \in B) \le 1/2}} \mathrm{P}_W(Y \in \bar{B} | X \in B) = \min_{\substack{U \subset V \\ \mathrm{Vol}\,(U) \le 1/2}} \mathrm{E}_W(\chi_{\bar{U}} | \chi_U),$$

where X and Y are identically distributed (their values are immaterial here), and their joint distribution is given by W, while χ_U is the indicator random variable belonging to the set U.

In [7] it is proved that with our notations

$$\begin{aligned} h &= \min_{\substack{X \in L_2(V, \mathcal{A}, D) \\ X \text{ is not constant}}} \max_{c \in \mathbb{R}} \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |x_i - x_j| w_{ij}}{\sum_{i=1}^{n} |x_i - c| d_i} \\ &= \frac{1}{2} \min_{\substack{X, Y \in L_2(V, \mathcal{A}, D) \text{ i.d.} \\ X \text{ is not constant}}} \max_{c \in \mathbb{R}} \frac{\mathrm{E}_W |X - Y|}{\mathrm{E}_D |X - c|}. \end{aligned}$$

If we take advantage of the fact that the median of X minimizes the expectation $\mathrm{E}_D|X - c|$ in c, we can simplify the above formula to

$$h = \frac{1}{2} \min_{\substack{X, Y \in L_2(V, \mathcal{A}, D) \text{ i.d.} \\ X \text{ is not constant}}} \frac{\mathrm{E}_W |X - Y|}{\mathrm{E}_D |X - \mathrm{med}_X|}, \tag{3.2}$$

the denominator containing the absolute deviation of X from its median, med_X. Thus (3.2) is the L_1-norm analog of the L_2-norm based minimum problem in (2.3). Utilizing this analog, the following proposition can easily be proved.

Proposition 3.1. Let $G = (V, \mathbf{W})$ be a weighted graph. The diagonal matrix $\mathbf{D}$ and the weighted Laplacian $\mathbf{C}_D$ are defined as in Section 2. Let λ_1 denote the smallest positive eigenvalue of $\mathbf{C}_D$ with harmonic eigenvector $\mathbf{x}^*$. Let $X^* \in L_2(V, \mathcal{A}, D)$ be a random variable with range vector $\mathbf{x}^*$ taking on values with the diagonal entries of $\mathbf{D}$. Then the Cheeger constant h of G satisfies

$$h \le \frac{\sqrt{\lambda_1}}{\sqrt{2}\mathrm{E}_D|X^* - \mathrm{med}_{X^*}|}$$

with med_{X^*} being the median of X^*.

A direct relation between h and λ_1 are stated in the following theorem. Similar statements are proved in [7] and [12] for unweighted graphs. To be self-contained, we include the proof of this theorem.

Theorem 3.2. *Let $G = (V, \mathbf{W})$ be a weighted graph. If h denotes its Cheeger constant and λ_1 is the smallest positive eigenvalue of its weighted Laplacian $\mathbf{C}_D$ then*

$$\frac{\lambda_1}{2} \le h \le \min\{1, \sqrt{2\lambda_1}\}$$

holds true. If $\lambda_1 \le 1$ then the upper estimate can be improved to

$$h \le \sqrt{\lambda_1(2-\lambda_1)}.$$

Proof. With the notations used throughout the paper, λ_1 is the smallest positive eigenvalue of $\mathbf{C}_D$, and $\mathbf{x}^*$ is a harmonic eigenvector belonging to it.

Lower bound. It follows easily by the optimum property of λ_1 and $\mathbf{x}^*$. Let U^* denote a vertex-subset at which the minimum of (3.1) is attained. Thus $\mathrm{Vol}\,(U^*) \le 1/2$ holds. Let us define the following representation of the vertices:

$$x_i := \begin{cases} 1/\mathrm{Vol}\,(U^*), & \text{if} \quad i \in U^* \\ -1/\mathrm{Vol}\,(\bar{U}^*), & \text{if} \quad i \in \bar{U}^* \end{cases}.$$

Then (2.3) gives that

$$\lambda_1 \le \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i-x_j)^2 w_{ij}}{\sum_{i=1}^{n} x_i^2 d_i} = \frac{\sum_{i\in U^*}\sum_{j\in\bar{U}^*}\left(\frac{1}{\mathrm{Vol}\,(U^*)}+\frac{1}{\mathrm{Vol}\,(\bar{U}^*)}\right)^2 w_{ij}}{\frac{1}{\mathrm{Vol}\,(U^*)}+\frac{1}{\mathrm{Vol}\,(\bar{U}^*)}}$$
$$\le \frac{\mathrm{Vol}\,(U^*)+\mathrm{Vol}\,(\bar{U}^*)}{\mathrm{Vol}\,(U^*)\mathrm{Vol}\,(\bar{U}^*)}\sum_{i\in U^*}\sum_{j\in\bar{U}^*} w_{ij} \le 2\frac{\sum_{i\in U^*}\sum_{j\in\bar{U}^*} w_{ij}}{\mathrm{Vol}\,(U^*)} = 2h,$$

which implies $\lambda_1/2 \le h$.

Upper bound. Let $\mathbf{x}^*$ be a harmonic eigenvector of $\mathbf{C}_D$ belonging to λ_1 directed such that

$$\sum_{i:\, x_i^*<0} d_i \ge \sum_{i:\, x_i^*\ge 0} d_i.$$

To simplify notation we drop * from $\mathbf{x}^*$ from now on. We rearrange the coordinates of $\mathbf{x}$ in increasing order:

$$x_1 \le \cdots \le x_{r-1} < 0 \le x_r \le \cdots \le x_n.$$

Actually, we took advantage of the fact that there are both negative and positive numbers among the coordinates. Say, the number of strictly negative coordinates is $r-1$, $r \geq 2$. The vertex set $V = \{1, \dots, n\}$ is rearranged, accordingly. Put $V_- := \{1, \dots, r-1\}$ and $V_+ := \{r, \dots, n\}$. By the above assumption, for the coordinates of $\mathbf{x}$ we have that

$$\sum_{i=1}^{r-1} d_i \geq \sum_{i=r}^{n} d_i. \tag{3.3}$$

Set $\mathbf{y} := \mathbf{x}_+$, that is the coordinates of the vector $\mathbf{y}$ are

$$y_i = \begin{cases} x_i, & \text{if} \quad x_i \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

We shall choose special two-partitions of the rearranged vertex-set induced by the subsets $U_k = \{k, \dots, n\}$ and put

$$c_k = \sum_{i \in U_k} \sum_{j \in \bar{U}_k} w_{ij}, \qquad (k = 2, \dots, n). \tag{3.4}$$

Obviously,

$$h \leq c = \min_{2 \leq k \leq n} \frac{c_k}{\min\{\mathrm{Vol}\,(U_k), \mathrm{Vol}\,(\bar{U}_k)\}}. \tag{3.5}$$

We remark that in view of (3.3), the relation

$$\min\{\mathrm{Vol}\,(U_k), \mathrm{Vol}\,(\bar{U}_k)\} = \mathrm{Vol}\,(U_k) = \sum_{i=k}^{n} d_i \qquad \text{for} \qquad k = r, \dots, n \tag{3.6}$$

is valid.

As $\mathbf{x}$ is a harmonic eigenvector of $\mathbf{C}_D = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ with eigenvalue λ_1,

$$\lambda_1 \mathbf{D}\mathbf{x} = \mathbf{D}\mathbf{x} - \mathbf{W}\mathbf{x}$$

holds, equivalently for the coordinates

$$\lambda_1 d_i x_i = d_i x_i - \sum_{j=1}^{n} w_{ij} x_j = \sum_{j=1}^{n} w_{ij}(x_i - x_j), \qquad (i = 1 \dots, n). \tag{3.7}$$

Multiplying both sides of (3.7) by x_i and summing up for indices $i \in V_+$ we get that

$$\lambda_1 \sum_{i \in V_+} d_i x_i^2 = \sum_{i \in V_+} x_i \sum_{j=1}^{n} w_{ij}(x_i - x_j),$$

or equivalently,

$$\lambda_1 = \frac{\sum_{i \in V_+} x_i \sum_{j=1}^{n} w_{ij}(x_i - x_j)}{\sum_{i \in V_+} d_i x_i^2} = \frac{A}{\sum_{i=1}^{n} d_i y_i^2}. \tag{3.8}$$

We shall estimate the numerator (A) from below as follows:

$$\begin{aligned}
A &= \sum_{i\in V_+}\sum_{j\in V_+} w_{ij}x_i(x_i-x_j) + \sum_{i\in V_+}\sum_{j\in V_-} w_{ij}x_i(x_i-x_j)\\
&= \sum_{\substack{i\in V_+,\, j\in V_+\\ i>j}} [w_{ij}x_i(x_i-x_j)+w_{ji}x_j(x_j-x_i)]+\\
&+\sum_{i\in V_+}\sum_{j\in V_-} w_{ij}x_i^2 - \sum_{i\in V_+}\sum_{j\in V_-} w_{ij}x_ix_j\\
&\overset{(1)}{=} \sum_{\substack{i\in V_+,\, j\in V_+\\ i>j}} w_{ij}(x_i-x_j)^2 + \sum_{i\in V_+}\sum_{j\in V_-} w_{ij}y_i^2 - \sum_{i\in V_+}\sum_{j\in V_-} w_{ij}x_ix_j\\
&\overset{(2)}{\geq} \sum_{\substack{i\in V_+,\, j\in V_+\\ i>j}} w_{ij}(y_i-y_j)^2 + \sum_{i\in V_+}\sum_{j\in V_-} w_{ij}(y_i-y_j)^2\\
&\overset{(3)}{=} \sum_{i\in V_+}\sum_{j<i} w_{ij}(y_i-y_j)^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-y_j)^2.
\end{aligned}$$

In the steps (1) and (2) we used the fact that y_i is equal to x_i on V_+ and 0 on V_-. We decreased the expression between the two steps by $-\sum_{i\in V_+}\sum_{j\in V_-} w_{ij}x_ix_j$ that is a nonnegative quantity due to the different signs of x_i and x_j for indices $i \in V_+$ and $j \in V_-$. In the step (3) we utilized that for such indices $i > j$ automatically holds true. We also used the symmetry of $\mathbf{W}$ several times.

Now, let us go back to (3.8). Using the lower estimate for A we get that

$$\lambda_1 \geq \frac{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-y_j)^2}{\sum_{i=1}^{n} d_i y_i^2} = Q. \tag{3.9}$$

The quantity Q defined above will be important later when we improve the estimate.

Q will be further decreased as follows.

$$
\begin{aligned}
Q &= \frac{\frac{1}{2}\left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i - y_j)^2\right] \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i + y_j)^2\right]}{\sum_{i=1}^n d_i y_i^2 \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i + y_j)^2\right]} \\
&= \frac{1}{2} \frac{\left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}|y_i - y_j|^2\right] \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}|y_i + y_j|^2\right]}{\sum_{i=1}^n d_i y_i^2 \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i + y_j)^2\right]} \\
&\geq \frac{1}{2} \frac{\left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}|y_i - y_j| \cdot |y_i + y_j|\right]^2}{\sum_{i=1}^n d_i y_i^2 \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i + y_j)^2\right]} \\
&= \frac{1}{2} \frac{\left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}|y_i^2 - y_j^2|\right]^2}{\sum_{i=1}^n d_i y_i^2 \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i + y_j)^2\right]^2} \\
&= \frac{1}{2} \frac{\left[2\sum_{i>j} w_{ij}|y_i^2 - y_j^2|\right]^2}{\sum_{i=1}^n d_i y_i^2 \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i + y_j)^2\right]}
\end{aligned}
$$

(3.10)

$$
= 2 \frac{\left[\sum_{i>j} w_{ij}(y_i^2 - y_j^2)\right]^2}{\sum_{i=1}^n d_i y_i^2 \cdot \left[\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i + y_j)^2\right]} = 2\frac{A_1^2}{B}.
$$

In the third line we used the Cauchy–Schwarz inequality for the expectation of random variables $|Y - Y'|$ and $|Y + Y'|$ with joint distribution $\mathbf{W}$ (Y and Y' are identically distributed with range vector $\mathbf{y}$).

To estimate A_1 from below, we shall use the fact that $y_i \geq y_j$ for $i > j$ and write the members $y_i^2 - y_j^2$ as a telescopic sum:

$$
y_i^2 - y_j^2 = (y_i^2 - y_{i-1}^2) + \cdots + (y_{j+1}^2 - y_j^2) \qquad \text{for } i > j.
$$

By this,

$$
\begin{aligned}
A_1 &= \sum_{i>j} w_{ij}(y_i^2 - y_j^2) = \sum_{k=2}^n (y_k^2 - y_{k-1}^2) \sum_{i \geq k > j} w_{ij} \overset{(4)}{=} \sum_{k=2}^n (y_k^2 - y_{k-1}^2) c_k \\
&= \sum_{k=r}^n (y_k^2 - y_{k-1}^2) c_k \overset{(5)}{\geq} \sum_{k=r}^n (y_k^2 - y_{k-1}^2) c \sum_{i=k}^n d_i \geq \sum_{k=r}^n (y_k^2 - y_{k-1}^2) h \sum_{i=k}^n d_i \\
&= h \sum_{k=r}^n (y_k^2 - y_{k-1}^2) \sum_{i=k}^n d_i \overset{(6)}{=} h \sum_{k=r}^n y_k^2 d_k
\end{aligned}
$$

where in (4) we used the definition of c_k, in (5) the relations (3.5) and (3.6) were exploited, while in (6) a partial summation was performed.

The denominator B is estimated from above:

$$B = \sum_{i=1}^{n} d_i y_i^2 \cdot \left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i+y_j)^2\right] \le \sum_{i=1}^{n} d_i y_i^2 \cdot \left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(2y_i^2+2y_j^2)\right]$$
$$= \sum_{i=1}^{n} d_i y_i^2 \cdot 4\sum_{i=1}^{n} y_i^2 d_i = 4\left(\sum_{i=1}^{n} y_i^2 d_i\right)^2.$$

There remains to collect the terms together:

$$\lambda_1 \ge \frac{2A_1^2}{B} \ge \frac{2h^2\left(\sum_{k=1}^{n} y_k^2 d_k\right)^2}{4\left(\sum_{i=1}^{n} y_i^2 d_i\right)^2} = \frac{h^2}{2},$$

and so, the upper estimate $h \le \sqrt{2\lambda_1}$ follows.

We can improve this upper bound by using the exact value of B and going back to (3.9) that implies

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-y_j)^2 = 2Q\sum_{i=1}^{n} d_i y_i^2.$$

An equivalent form of B is

$$B = \sum_{i=1}^{n} d_i y_i^2 \cdot \left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i+y_j)^2\right] = \sum_{i=1}^{n} d_i y_i^2 \cdot \left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(2y_i^2+2y_j^2-(y_i-y_j)^2)\right]$$
$$= \sum_{i=1}^{n} d_i y_i^2 \cdot \left[4\sum_{i=1}^{n} y_i^2 d_i - \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-y_j)^2\right] = 2\left(\sum_{i=1}^{n} d_i y_i^2\right)^2(2-Q).$$

Starting the estimation of Q at (3.10) and continuing with the B above, yields

$$Q \ge 2\frac{A_1^2}{B} \ge 2\frac{h^2(\sum_{k=1}^{n} y_k^2 d_k)^2}{2(\sum_{i=1}^{n} d_i y_i^2)^2(2-Q)} = \frac{h^2}{2-Q}.$$

By (3.9), Q is non-negative implying

$$Q \ge \frac{h^2}{2-Q} \qquad \text{or equivalently,} \qquad 1-\sqrt{1-h^2} \le Q \le 1+\sqrt{1-h^2}.$$

Summarizing, we derive that

$$\lambda_1 \ge Q \ge 1-\sqrt{1-h^2} \qquad \text{or equivalently,} \qquad \sqrt{1-h^2} \ge 1-\lambda_1.$$

For $\lambda_1 > 1$ this is a trivial statement. For $\lambda_1 < 1$ it implies that $h \le \sqrt{\lambda_1(2-\lambda_1)} < 1$ while for $\lambda_1 = 1$ we get the trivial bound $h \le 1$. This finishes the proof. □

In terms of the symmetric maximal correlation the result of Theorem 3.2 can be written in an equivalent form as follows.

Corollary 3.3. Let $\mathbf{W}$ be a symmetric joint distribution of two discrete random variables taking on at most n different values. If the symmetric maximal correlation (r_1 of Section 3) is non-negative then the estimation

$$\frac{1-r_1}{2} \le \min_{\substack{B\subset\mathbb{R}\text{ Borel-set}\\ X,Y\in L_2(V,\mathcal{A},D)\text{ i.d.}\\ \mathrm{P}_D(X\in B)\le 1/2}} \mathrm{P}_W(X' \in \bar{B}|X\in B) \le \sqrt{1-r_1^2} \tag{3.11}$$

holds.

Proof. Since $\lambda_1 = 1 - r_1$, the lower bound trivially follows. $r_1 \ge 0$ implies that $\lambda_1 \le 1$, so the improved upper bound of Theorem 4.2 becomes $\sqrt{(1-r_1)(1+r_1)}$. This finishes the proof. □

Consequently, the symmetric maximal correlation somehow regulates the minimum conditional probability that provided a random variable takes values in a category set (with probability less than 1/2) then another one with the same distribution (their joint distribution is given by $\mathbf{W}$) will take values in the complement category set. In particular, if r_1 is the eigenvalue of $\mathbf{I}-\mathbf{C}_D$ with the largest absolute value (apart from 1), then r_1 is the usual maximal correlation, and in this case inequality (3.11) also holds for it.

4. k-density, a generalization of the isoperimetric number

We can generalize the notion of the Cheeger constant for $k \ge 2$. Let $P_k = (V_1,\dots,V_k)$ be a given k-partition. The k-density of P_k is defined as

$$\rho(P_k) = \sum_{l=1}^{k-1}\sum_{m=l+1}^{k}\left(\frac{1}{Vol(V_l)}+\frac{1}{Vol(V_m)}\right)w(V_l,V_m),$$

where for $l \ne m$ we set $w(V_l,V_m) = \sum_{i\in V_l}\sum_{j\in V_m} w_{ij}$, and let

$$\rho_k = \min_{P_k\in\mathcal{P}_k} \rho(P_k)$$

be the *k-density* of G, where $\mathcal{P}_k$ denotes the set of all possible k-partitions into disjoint, non-empty subsets of a set of cardinality n. (The cardinality of $\mathcal{P}_k$ is the Stirling number of second order.)

It is easy to see that ρ_k punishes k-partitions with "many" inter-cluster edges of "large" weights and with "strongly" differing volumes. Further, $\rho_2 \le 2h$. The quantity ρ_2 was also introduced in Mohar [12] for ordinary graphs (with 0-1 weights).

Theorem 4.1. *Suppose that $G = (V,\mathbf{W})$ is connected. With the notations of the previous sections*

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

and in the case when the optimal k-dimensional representatives can be classified into k well-separated clusters in such a way that the maximum cluster diameter ε satisfies *the relation* $\varepsilon \le \min\{1/\sqrt{2k}, \sqrt{2}\min_i \sqrt{p_i}\}$ *– where* $p_i = Vol(V_i)$*,* $i = 1, \dots, k$*, with* k*-partiton* $(V_1, \dots, V_k)$ *induced by the clusters above – then*

$$\rho_k \le c^2 \sum_{i=1}^{k-1} \lambda_i,$$

where $c = 1 + \varepsilon c'/(\sqrt{2} - \varepsilon c')$ *and* $c' = 1/\min_i \sqrt{p_i}$*.*

Proof. Lower bound. Let $P_k^* = (V_1^*, \dots, V_k^*)$ be a k-partition with $\rho(P_k^*) = \rho_k$ and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ be the following representation:

$$x_{ij} = \begin{cases} 1/\sqrt{Vol(V_i^*)}, & \text{if} \quad j \in V_i^*, \\ 0, & \text{otherwise}, \end{cases}$$

where x_{ij}s are the entries of $\mathbf{X}$. In this representation the objective function (1.1) becomes $L(\mathbf{X}) = \rho_k(P_k^*)$, but this is greater than or equal to $\sum_{i=1}^{k-1} \lambda_i$, latter one being the minimum of $L(\mathbf{X})$ that finishes the proof of the first part.

Upper bound. To prove the reversed statement, let $P_k = (V_1, \dots, V_k)$ be a k-partition obtained by k-means classification of the optimal k-dimensional Euclidean representatives, $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$ (their first coordinates are identically 1). As they fall into k well-separated clusters (the maximum cluster diameter being less than or equal to the minimum distance of $\mathbf{x}_j^*$s of different clusters), the MacQueen method [11] converges to a unique solution. According to our assumption

$$\varepsilon = \max_{i \sim j} \|\mathbf{x}_i^* - \mathbf{x}_j^*\| \le \min\{\frac{1}{\sqrt{2k}}, \sqrt{2}\min_i \sqrt{p_i}\},$$

where the relation $i \sim j$ denotes that the vertices i and j belong to the same cluster. The representatives satisfy the condition: $\sum_{j=1}^n d_j \mathbf{x}_j^* \mathbf{x}_j^{*T} = \mathbf{X}^* \mathbf{D} \mathbf{X}^{*T} = \mathbf{I}_k$.

Let $\bar{\mathbf{x}}^{(i)}$ denote the center of the ith cluster:

$$\bar{\mathbf{x}}^{(i)} = \frac{1}{p_i} \sum_{j \in V_i} d_j \mathbf{x}_j, \quad i = 1, \dots, k.$$

Furthet, let $\mathbf{y}_i$ denote the k-dimensional vector with coordinates

$$y_{ij} = \begin{cases} 1/\sqrt{p_i}, & \text{if} \quad j \in V_i, \\ 0, & \text{otherwise}, \end{cases} \tag{4.1}$$

and $\mathbf{Y} =: (\mathbf{y}_1, \dots, \mathbf{y}_k)$. In fact, with $\mathbf{P} = \operatorname{diag}(p_1, \dots, p_k)$ the relation $\mathbf{Y} = \mathbf{P}^{-1/2}$ holds. Let $\mathbf{R}$ be a $k \times k$ orthogonal matrix. With the notations $\mathbf{y}_i' = \mathbf{R}\mathbf{y}_i$ and $\mathbf{Y}' = \mathbf{R}\mathbf{Y}$ we are looking for a system $\mathbf{Y}'$ such that $\mathbf{y}_i'$ is "close" to the cluster center $\bar{\mathbf{x}}^{(i)}$ for $i = 1, \dots, k$. To this end, we use the so-called MANOVA (Multivariate ANalysis Of VAriance) decomposition of the $k \times k$ covariance matrix of $\mathbf{x}_j^*$s into within-clusters and between-clusters covariances (the mean of the components of $\mathbf{x}_j^*$s is zero except the first one that is identically 1, but it will not contribute to the variances):

$$\sum_{j=1}^n d_j \mathbf{x}_j^* \mathbf{x}_j^{*T} = \sum_{i=1}^k \sum_{j \in V_i} d_j (\mathbf{x}_j^* - \bar{\mathbf{x}}^{(i)})(\mathbf{x}_j^* - \bar{\mathbf{x}}^{(i)})^T + \sum_{i=1}^k p_i \bar{\mathbf{x}}^{(i)} \bar{\mathbf{x}}^{(i)T}, \tag{4.2}$$

or briefly,

$$\mathbf{I}_k = \sum_{i=1}^{k} \mathbf{A}_i + \mathbf{B} = \mathbf{A} + \mathbf{B}, \tag{4.3}$$

where $\mathbf{A}_i = \sum_{j \in V_i} d_j (\mathbf{x}_j^* - \bar{\mathbf{x}}^{(i)})(\mathbf{x}_j^* - \bar{\mathbf{x}}^{(i)})^T$, $i = 1, \dots, k$. Here $\operatorname{tr} \mathbf{A}_i$ is the k-variance of representatives in cluster i, therefore $\operatorname{tr} \mathbf{A}_i \le \sum_{c(j)=i} d_j \varepsilon^2 = p_i \varepsilon^2$, and $\operatorname{tr} \mathbf{A} = \sum_{i=1}^k \operatorname{tr} \mathbf{A}_i \le \varepsilon^2$. As $\mathbf{A}$ is symmetric, positive semidefinite, its maximum eigenvalue is at most ε^2. So, $\mathbf{A}$ will be regarded as a perturbation on $\mathbf{B}$. The matrix $\mathbf{B} = \mathbf{I}_k - \mathbf{A}$ is also positive semidefinite and by the Weyl's perturbation theory it follows that denoting by $\lambda_1, \dots \lambda_k$ its eigenvalues, for them the relation

$$0 \le 1 - \lambda_i \le \varepsilon^2, \qquad i = 1, \dots, k \tag{4.4}$$

holds. With the notation $\bar{\mathbf{X}} = (\bar{\mathbf{x}}^{(1)}, \dots, \bar{\mathbf{x}}^{(k)})$ our matrix $\mathbf{B}$ is equal to $\bar{\mathbf{X}}\mathbf{P}\bar{\mathbf{X}}^T$.

Now, let us find such an $\mathbf{R}$ that with $(\mathbf{y}_1', \dots, \mathbf{y}_k') = \mathbf{R}\mathbf{Y}$ the sum $\sum_{i=1}^k p_i \|\bar{\mathbf{x}}^{(i)} - \mathbf{y}_i'\|^2$ be the least possible.

$$\begin{aligned}\sum_{i=1}^{k} p_i \|\bar{\mathbf{x}}^{(i)} - \mathbf{y}_i'\|^2 &= \operatorname{tr}(\bar{\mathbf{X}} - \mathbf{R}\mathbf{Y})\mathbf{P}(\bar{\mathbf{X}} - \mathbf{R}\mathbf{Y})^T \\ &= \operatorname{tr} \bar{\mathbf{X}}\mathbf{P}\bar{\mathbf{X}}^T + \operatorname{tr} \mathbf{R}\mathbf{Y}\mathbf{P}\mathbf{Y}^T\mathbf{R}^T - 2\operatorname{tr} \bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T\mathbf{R} \ge \sum_{i=1}^{k} \lambda_i + k - 2\sum_{i=1}^{k} s_i,\end{aligned} \tag{4.5}$$

where $s_1, \dots s_k$ are the singular values of the matrix $\bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T$. I.e., the first term is $\operatorname{tr} \mathbf{B}$, the second is $\operatorname{tr} \mathbf{I}_k$, while to the third one the following theorem is applicable. With our notations, $\operatorname{tr} \bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T\mathbf{R}$ is maximum (with respect to $\mathbf{R}$) if the matrix $\bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T\mathbf{R}$ is symmetric and the maximum is equal to the sum of the singular values of $\bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T$. By choosing such an $\mathbf{R}$, equality can be attained. Taking into account that

$$(\bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T)(\bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T)^T = \bar{\mathbf{X}}\mathbf{P}\mathbf{Y}^T\mathbf{Y}\mathbf{P}\bar{\mathbf{X}}^T = \bar{\mathbf{X}}\mathbf{P}\mathbf{P}^{-1}\mathbf{P}\bar{\mathbf{X}}^T = \bar{\mathbf{X}}\mathbf{P}\bar{\mathbf{X}}^T = \mathbf{B},$$

the eigenvalues of $\mathbf{B}$ can be enumerated in such an order that $\lambda_i = s_i^2$, $i = 1, \dots, k$. But we saw that s_i^2 is of order $1 - \varepsilon^2$, therefore via Taylor's expansion $1 - s_i \approx \frac{\varepsilon^2}{2} + \frac{\varepsilon^4}{4}$ is a good approximation. Hence, with the choice of $\mathbf{R}$ giving equality in (4.5) we have that

$$\sum_{i=1}^{k} p_i \|\bar{\mathbf{x}}^{(i)} - \mathbf{y}_i'\|^2 = \sum_{i=1}^{k} s_i^2 - k + 2k - 2\sum_{i=1}^{k} s_i = \sum_{i=1}^{k} (s_i^2 - 1) + 2\sum_{i=1}^{k} (1 - s_i) \approx 2k\varepsilon^4$$

that is less than ε^2 provided that $\varepsilon \le 1/\sqrt{2k}$ holds. Consequently, $p_i \|\bar{\mathbf{x}}^{(i)} - \mathbf{y}_i'\|^2 \le \varepsilon^2$ and $\|\bar{\mathbf{x}}^{(i)} - \mathbf{y}'\| \le \varepsilon c'$.

Let the $\mathbf{y}_i'$ nearest to $\bar{\mathbf{x}}^{(i)}$ be denoted by $\mathbf{y}(\mathbf{x}_j^*)$ for every js in V_i (thus $\mathbf{y}(\mathbf{x}_j^*) = \mathbf{y}_i'$, $\forall\, j \in V_i$). Let δ denote the minimum distance between the different $\mathbf{y}_i'$s, that is

$$\delta = \min_{l \ne m} \|\mathbf{y}_l' - \mathbf{y}_m'\| = \min_{l \ne m} \|\mathbf{y}_l - \mathbf{y}_m\| = \min_{l \ne m} \sqrt{\frac{1}{p_l} + \frac{1}{p_m}} \ge \sqrt{2}.$$

With them,

$$\begin{aligned}
\rho_k &\le \rho(V_1,\dots,V_k) = L(\mathbf{y}(\mathbf{x}_1^*),\dots,\mathbf{y}(\mathbf{x}_n^*)) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}\|\mathbf{y}(\mathbf{x}_i^*)-\mathbf{y}(\mathbf{x}_j^*)\|^2 \\
&\le \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}(c\|x_i^*-x_j^*\|)^2 \le c^2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}\|x_i^*-x_j^*\|^2 \\
&= c^2 L(x_1^*,\dots,x_n^*) = c^2\sum_{i=1}^{k-1}\lambda_i,
\end{aligned}$$

where

$$c = \frac{\delta}{\delta-\varepsilon c'} = 1+\frac{\varepsilon c'}{\delta-\varepsilon c'} \le 1+\frac{\varepsilon c'}{\sqrt{2}-\varepsilon c'},$$

that implies our statement. □

The first part of the theorem gives $\lambda_1 \le \rho_2 \le 2h$ for $k = 2$; therefore, it also implies the lower estimate of the Cheeger constant. The constant c of the second part is greater than 1, and it is the closer to 1, the smaller ε is. The latter requirement is satisfied if there exists a "very" well-separated k-partition of the k-dimensional Euclidean representatives. From the above theorem, we can also conclude that the gap in the spectrum is a necessary but not a sufficient condition of a good classification. In addition, the Euclidean representatives should be well classified in the appropriate dimension.

Acknowledgements

We would like to thank Profs. Gábor Tusnády and Bojan Mohar for their valuable comments.

References

[1] Alon, N., Milman, V.D., λ_1, Isoperimetric inequalities for graphs and superconcentrators, J. Comb. Theory B 38 (1985), 73-88.

[2] Bolla, M., Spectra, Euclidean representations and clusterings of hypergraphs, Discrete Mathematics 117 (1993), 19-39.

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

[4] Breiman, I., Friedman, H.J., Estimating optimal transformations for multiple regression and correlation, J. Amer. Stat. Assoc. 80 (1985), 580-619.

[5] Buser, P., On the bipartition of graphs, Discrete Appl. Math. 9 (1984), 105-109.

[6] Cheeger, J., A lower bound for the smallest eigenvalue of the Laplacian, Problems in Analysis (Gunnig, R. C. ed.), Princeton Univ. Press (1970), 195-199.

[7] Chung, F.R.K., Spectral graph theory, CBMS Regional Conference Series in Mathematics, No. 92 (1994), AMS Publications.

[8] Cvetković, D.M., Doob, M., Sachs, H., Spectra of graphs (1979), Academic Press, New York–San Francisco–London.

[9] Diaconis, P., Stroock, D.W., Geometric bounds for eigenvalues of Markov chains, Annals Applied Prob. 1 (1991), 36-61.

[10] Lubotzky, A., Discrete groups, expanding graphs, and invariant measures. Appendix by Rogawski, J.D. Progress in Mathematics 125 (1994), Birkhäuser, Basel–Boston–Berlin.

[11] MacQueen, J., Some methods for classification and analysis of multivariate observations. Proc. 5th Berkeley Symp. on Math. Stat. and Prob., Univ. of California Press, Berkeley (1967). Vol.1. 281-297.

[12] Mohar, B., Isoperimetric numbers of graphs, J. Comb. Theory B 47 (1989), 274-291.

[13] Nikiforov, V., Some inequalities for the largest eigenvalue of a graph, Combinatorics, Probability and Computing 11 (2002), 179-189.

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

[15] Sinclair, A., Jerrum, M., Approximate counting, uniform generation and rapidly mixing Markov chains, Information and Computation 82 (1989), 93-133.

Adress of authors
Institute of Mathematics, Budapest University of Technology and Economics
1521 Budapest, P.O.Box 91. Bldg. H. V/7. Hungary
e-mail: marib@math.bme.hu, molnar@math.bme.hu