Note

Discrepancy minimizing spectral clustering

Marianna Bolla *, Ahmed Elbanna

Institute of Mathematics, Technical University of Budapest, Hungary

ARTICLE INFO

ABSTRACT

In this short note we strengthen a former result of Bolla (2011), where in a multipartition (clustering) of a graph's vertices we estimated the pairwise discrepancies of the clusters with the normalized adjacency spectra. There we used the definition of Alon et al. (2010) for the volume-regularity of the cluster pairs. Since then, in Bolla (2016) we defined the so-called k-way discrepancy of a k-clustering and estimated the kth largest (in absolute value) normalized adjacency eigenvalue with an increasing function of it. In the present paper, we estimate the new discrepancy measure with this eigenvalue. Putting these together, we are able to establish a relation between the large spectral gap (as for the $(k-1)$th and kth non-trivial normalized adjacency eigenvalues) and the sudden decrease between the $k-1$ and k-way discrepancies. It makes rise to a new paradigm of spectral clustering, which minimizes the multiway discrepancy.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Our purpose is to relate discrepancy and spectra in a multiway classification of a graph's vertices. To do so, we develop an estimate for the multiway discrepancy by spectral tools. To simplify notation, we use unweighted graphs, though the result can as well be extended to edge-weighted ones.

Let $G = (V, \boldsymbol{A})$ be an undirected, unweighted graph on the n-element vertex-set V with the $n \times n$ adjacency matrix $\boldsymbol{A} = (a_{ij})$. We will use the normalized adjacency matrix $\boldsymbol{A_D} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ of G, where $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_n)$ is the diagonal *degree-matrix* and d_v is the degree of vertex v. Assume that G is *connected*, i.e., $\boldsymbol{A}$ is *irreducible*. Then the eigenvalues of $\boldsymbol{A_D}$, enumerated in decreasing absolute values are $1 = \mu_0 \geq |\mu_1| \geq |\mu_2| \geq \cdots \geq |\mu_{n-1}|$, and μ_0 is a single eigenvalue. (Note that $\mu_1 = -1$ can be if G is bipartite.)

Let $1 < k < n$ be a fixed integer. We look for the proper k-partition (clustering) $V_1, \ldots, V_k$ of the vertices such that the within- and between-cluster discrepancies are minimized. Let $a(X, Y) = \sum_{u \in X} \sum_{v \in Y} a_{ij}$ be the edge-cut between $X, Y \subset V$, and $\text{Vol}(X) = \sum_{v \in X} d_v$ be the volume of the vertex-subset X. Further, let $\rho(X, Y) := \frac{a(X,Y)}{\text{Vol}(X)\text{Vol}(Y)}$ be the volume–density between X and Y. The multiway discrepancy [6] of $G = (V, \boldsymbol{A})$ in the clustering $V_1, \ldots, V_k$ of its vertices is

$$\text{md}(G; V_1, \ldots, V_k) = \max_{1 \leq i \leq j \leq k} \max_{X \subset V_i, Y \subset V_j} \frac{|a(X, Y) - \rho(V_i, V_j)\text{Vol}(X)\text{Vol}(Y)|}{\sqrt{\text{Vol}(X)\text{Vol}(Y)}}.$$

The *minimum k-way discrepancy* of G is $\text{md}_k(G) = \min_{(V_1, \ldots, V_k) \in \mathcal{P}_k} \text{md}(G; V_1, \ldots, V_k)$, where $\mathcal{P}_k$ denotes the set of proper k-partitions of V. This generalizes the notion of the volume-regular cluster pairs of [2] in the following way. If $\alpha = \text{md}(G; V_1, \ldots, V_k)$, then α is the smallest positive constant such that for every V_i, V_j pair and for every $X \subset V_i, Y \subset V_j$,

$$|a(X, Y) - \rho(V_i, V_j)\text{Vol}(X)\text{Vol}(Y)| \leq \alpha\sqrt{\text{Vol}(X)\text{Vol}(Y)}$$

---

* Corresponding author.
E-mail addresses: marib@math.bme.hu (M. Bolla), ahmed@math.bme.hu (A. Elbanna).

holds. In [2], the α volume-regularity of the cluster pairs requires only

$$|a(X, Y) - \rho(V_i, V_j)\text{Vol}(X)\text{Vol}(Y)| \leq \alpha \sqrt{\text{Vol}(V_i)\text{Vol}(V_j)}, \tag{1}$$

a condition on which the estimate of [4] is based.

In Section 2, we give a strengthened discrepancy estimate by means of spectral tools. We use the following measure for the 'goodness' of a k-clustering. Assume that $|\mu_{k-1}| > |\mu_k|$, and denote by $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$ the unit-norm, pairwise orthogonal eigenvectors, corresponding to the leading (non-trivial) eigenvalues $\mu_1, \ldots, \mu_{k-1}$ of $\boldsymbol{A_D}$. Let $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^{k-1}$ be the row vectors of the $n \times (k-1)$ matrix of column vectors $\boldsymbol{D}^{-1/2}\mathbf{u}_1, \ldots, \boldsymbol{D}^{-1/2}\mathbf{u}_{k-1}$; they are called $(k-1)$-dimensional representatives of the vertices. The *weighted k-variance* of these representatives is defined as

$$\tilde{S}_k^2 = \min_{(V_1, \ldots, V_k) \in \mathcal{P}_k} \sum_{i=1}^{k} \sum_{v \in V_i} d_v \|\mathbf{r}_v - \mathbf{c}_i\|^2,$$

where $\mathbf{c}_i = \frac{1}{\text{Vol}(V_i)} \sum_{v \in U_i} d_v \mathbf{r}_v$ is the weighted center of the cluster V_i. Actually, $\tilde{S}_k^2$ is the minimal distance between the subspace spanned by the vectors $\mathbf{1}, \boldsymbol{D}^{-1/2}\mathbf{u}_1, \ldots, \boldsymbol{D}^{-1/2}\mathbf{u}_{k-1}$ and that spanned by step-vectors over the k-partitions in $\mathcal{P}_k$, see [5]. It is the *weighted k-means algorithm* that provides this minimum. This is the generalization of the k-means algorithm, for which there are polynomial time approximating schemes (PTAS), see [12]. The so-called spectral relaxation (base of the spectral clustering) means that we can approximately find discrepancy minimizing clustering via applying the weighted k-means algorithm to the $(k-1)$-dimensional vertex representatives.

In Theorem 1, for a general term of a graph sequence having 'no dominant vertices', we estimate $\text{md}_k(G)$ with $|\mu_k|, \tilde{S}_k$, and exact constants by strengthening the estimate of Theorem 2.1 of [4] in that there instead of exact constants only big-O, and instead of $\text{md}_k(G)$ the weaker discrepancy measure (1) for the volume-regularity of cluster pairs was used.

Conversely, in [6] we estimated $|\mu_k|$ with $\text{md}_k(G)$ as

$$|\mu_k| \leq 9\text{md}_k(G)(k + 2 - 9k \ln \text{md}_k(G)). \tag{2}$$

Putting these back and forth statements together, in Section 3 we establish an important relation between the gaps in the spectral and in the discrepancy view for expanding graph sequences, see Corollary 2. Hence, we justify for the discrepancy minimizing spectral clustering, and contribute to the characterization of the generalized quasirandom graphs (see [11]) by spectra and spectral subspaces.

So we are able to prove back and forth relations between k-clusterings of the vertices and the inner spectral gap (at the kth eigenvalue) of the normalized adjacency spectrum, with the same k, thus also indicating the choice of the optimal k. Relations between the overall discrepancy of a graph and the spectral gap at the edge of the spectrum have long been investigated together with quasirandomness, e.g., Thomason [13], Chung–Graham–Wilson [10], Bollobás–Nikiforov [8], Bilu–Linial [3], Butler [9]. However, their results apply to the $k = 1$ case. Similar results for bipartite graphs, e.g., Alon [1] and Thomason [14], correspond to the $k = 2$ case, but their findings do not extend trivially to a general k. For simulation results and applications see [7].

2. The strengthened discrepancy estimate

Theorem 1. *Let G_n be the general term of a connected simple graph sequence, G_n has n vertices. (We do not denote the dependence of the vertex-set V and adjacency matrix $\boldsymbol{A}$ of G_n on n.) Assume that there are constants $0 < c < C < 1$ such that except $o(n)$ vertices, the degrees satisfy $cn \leq d_v \leq Cn$, $v = 1, \ldots, n$. Let the eigenvalues of the normalized adjacency matrix of G_n, enumerated in decreasing absolute values, be*

$$1 = \mu_0 \geq |\mu_1| \geq \cdots \geq |\mu_{k-1}| > \varepsilon \geq |\mu_k| \geq \cdots \geq |\mu_{n-1}| = 0.$$

The partition $(V_1, \ldots, V_k)$ of V is defined so that it minimizes the weighted k-variance $s^2 = \tilde{S}_k^2$ of the optimal $(k-1)$-dimensional vertex representatives of G_n. Assume that this k-partition $(V_1, \ldots, V_k)$ satisfies the balancing condition: $\frac{|V_i|}{n} \to r_i$ for $i = 1, \ldots, k$ as $n \to \infty$ with some positive reals $r_1, \ldots, r_k$. Then

$$\text{md}_k(G_n) \leq \text{md}(G_n; V_1, \ldots, V_k) \leq 2\left(\frac{C}{c} + o(1)\right)(\sqrt{2k}s + \varepsilon).$$

Proof. At the beginning, we follow the proof of [4], the idea of which is summarized briefly. Denote by $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$ the unit-norm, pairwise orthogonal eigenvectors of $\boldsymbol{A_D}$ corresponding to the k largest (in absolute value) eigenvalues $\mu_0, \mu_1, \ldots, \mu_{k-1}$ of $\boldsymbol{A_D}$. The $(k-1)$-dimensional representatives of the vertices of G_n are row vectors of the matrix with column vectors $\mathbf{x}_i = \boldsymbol{D}^{-1/2}\mathbf{u}_i$ ($i = 1, \ldots, k-1$). The representatives can as well be regarded as k-dimensional ones, as by inserting the vector $\mathbf{x}_0 = \boldsymbol{D}^{-1/2}\mathbf{u}_0$ will not change the weighted k-variance $s^2 = \tilde{S}_k^2$, because it is the vector of all 1 coordinates. Suppose that the minimum weighted k-variance is attained at the k-partition $(V_1, \ldots, V_k)$ of the vertices. By an easy analysis of variance argument (see [5]) it follows that $s^2 = \sum_{i=0}^{k-1} \text{dist}^2(\mathbf{u}_i, F)$, where $F = \text{Span}\{\boldsymbol{D}^{1/2}\mathbf{z}_1, \ldots, \boldsymbol{D}^{1/2}\mathbf{z}_k\}$ with the so-called normalized partition vectors $\mathbf{z}_1, \ldots, \mathbf{z}_k$ of coordinates $z_{ji} = \frac{1}{\sqrt{\text{Vol}(V_i)}}$ if $j \in V_i$ and 0, otherwise ($i = 1, \ldots, k$).