A Note on Spectral Clustering

Pavel Kolev* Kurt Mehlhorn
Max-Planck-Institut für Informatik, Saarbrücken, Germany
{pkolev,mehlhorn}@mpi-inf.mpg.de

Abstract

Spectral clustering is a popular and successful approach for partitioning the nodes of a graph into clusters for which the ratio of outside connections compared to the volume (sum of degrees) is small. In order to partition into k clusters, one first computes an approximation of the first k eigenvectors of the (normalized) Laplacian of G, uses it to embed the vertices of G into k-dimensional Euclidean space $\mathbb{R}^k$, and then partitions the resulting points via a k-means clustering algorithm. It is an important task for theory to explain the success of spectral clustering.

Peng et al. (COLT, 2015) made an important step in this direction. They showed that spectral clustering provably works if the gap between the $(k+1)$-th and the k-th eigenvalue of the normalized Laplacian is sufficiently large. They prove a structural and an algorithmic result. The algorithmic result needs a considerably stronger gap assumption and does not analyze the standard spectral clustering paradigm; it replaces spectral embedding by heat kernel embedding and k-means clustering by locality sensitive hashing.

We extend their work in two directions. Structurally, we improve the quality guarantee for spectral clustering by a factor of k and simultaneously weaken the gap assumption. Algorithmically, we show that the standard paradigm for spectral clustering works. Moreover, it even works with the same gap assumption as required for the structural result.

*This work has been funded by the Cluster of Excellence "Multimodal Computing and Interaction" within the Excellence Initiative of the German Federal Government.

Contents

1 Introduction **2**

2 Notations **5**

3 Technical Advances in The Improved Quality Guarantee **7**

4 Vectors $\widehat{g}_i$ and f_i are Close **7**
4.1 Analyzing the Columns of Matrix $\mathbf{F}$ . . . 7
4.2 Analyzing Eigenvectors f in terms of $\widehat{f}_j$ . . . 9

5 Spectral Properties of Matrix B **10**
5.1 Analyzing the Column Space of Matrix $\mathbf{B}$ . . . 10
5.2 Analyzing the Row Space of Matrix $\mathbf{B}$ . . . 11

6 Vectors $p^{(i)}$ are Well-Spread **14**

7 Proof of Lemma 1.3 **15**

8 Proof of Lemma 1.4 **15**

9 The Normalized Spectral Embedding is ε-separated **17**

10 An Efficient Spectral Clustering Algorithm **19**
10.1 Notations . . . 19
10.2 Spectral Embedding Properties . . . 20
10.3 Spectral Embeddings, Gap Assumption and ε-separability . . . 23
10.4 Proof of Part (b) of Theorem 1.2 . . . 25

11 Parameterized Upper Bound on $\widehat{\rho}_{\mathrm{avr}}(k)$ **26**

1 Introduction

A *cluster* in an undirected graph $G = (V, E)$ is a set S of nodes whose volume is large compared to the number of outside connections. Formally, we define the *conductance* of S by $\phi(S) = \left|E(S, \overline{S})\right| / \mu(S)$, where $\mu(S) = \sum_{v \in S} \deg(u)$ is the *volume* of S. The k-way partitioning problem for graphs asks to partition the vertices of a graph such that the conductance of each block of the partition is small (formal definition below). This problem arises in many applications, e.g., image segmentation and exploratory data analysis. We refer to the survey [8] for additional information. A popular and very successful approach to clustering [3, 7, 8] is *spectral clustering*. One first computes an approximation of the first k eigenvectors of the (normalized) Laplacian of G, uses it to embed the vertices of G into k-dimensional Euclidean space $\mathbb{R}^k$, and then partitions the resulting points via a k-means clustering algorithm. It is an important task for theory to explain the success of spectral clustering. Peng et al. [6] made an important step in this direction recently. They showed that spectral clustering provably works if the $(k+1)$-th and the k-th eigenvalue of the normalized Laplacian differ sufficiently. In order to explain their result, we need some notation.

The *order k partition constant* $\widehat{\rho}(k)$ of G is defined by

$$\widehat{\rho}(k) \triangleq \min_{\text{partition } (P_1,\ldots,P_k) \text{ of } V} \Phi(P_1, \ldots P_k), \quad \text{where} \quad \Phi(Z_1, \ldots, Z_k) = \max_{i \in [1:k]} \phi(Z_i).$$

Let $\mathcal{L}_G = I - D^{-1/2} A D^{-1/2}$ be the normalized Laplacian matrix of G, where D is the diagonal degree matrix and A is the adjacency matrix, and let $f_j \in \mathbb{R}^V$ be the eigenvector corresponding to the j-th smallest eigenvalue λ_j of $\mathcal{L}_G$. The *spectral embedding map* $F : V \to \mathbb{R}^k$ is defined by

$$F(u) = \frac{1}{\sqrt{d_u}} \left(f_1(u), \ldots, f_k(u)\right)^{\mathrm{T}}, \quad \text{for all vertices } u \in V. \tag{1}$$

Peng et al. [6] construct a k-means instance $\mathcal{X}_V$ by inserting d_u many copies of the vector $F(u)$ into $\mathcal{X}_V$, for every vertex $u \in V$.

Let $\mathcal{X}$ be a set of vectors of the same dimension. Then

$$\triangle_k(\mathcal{X}) \triangleq \min_{\text{partition } (X_1,\ldots,X_k) \text{ of } \mathcal{X}} \sum_{i=1}^{k} \sum_{x \in X_i} \|x - c_i\|^2, \quad \text{where} \quad c_i = \frac{1}{|X|} \sum_{x \in X_i} x,$$

is the optimal cost of clustering $\mathcal{X}$ into k sets. An α-approximate clustering algorithm returns a k-way partition $(A_1, \ldots, A_k)$ and centers $c_1, \ldots, c_k$ such that

$$\mathrm{Cost}(\{A_i, c_i\}_{i=1}^k) \triangleq \sum_{i=1}^{k} \sum_{x \in A_i} \|x - c_i\|^2 \leqslant \alpha \cdot \triangle_k(\mathcal{X}). \tag{2}$$

Theorem 1.1 ([6]). *Let $k \geqslant 3$ and $(P_1, \ldots, P_k)$ be a k-way partition of V with $\Phi(P_1, \ldots, P_k) = \widehat{\rho}(k)$. Let G be a graph that satisfies the gap assumption*

$$\Upsilon = \frac{\lambda_{k+1}}{\widehat{\rho}(k)} = 2 \cdot 10^5 \cdot k^3 / \delta, \tag{3}$$

for some $\delta \in (0, 1/2]$. Let $(A_1, \ldots, A_k)$ be the k-way partition[1] *of V returned by an α-approximate k-means algorithm applied to $\mathcal{X}_V$. Then the following statements hold (after suitable renumbering of one of the partitions):*

$$1)\ \mu(A_i \triangle P_i) \leqslant \alpha\delta \cdot \mu(P_i) \quad \text{and} \quad 2)\ \phi(A_i) \leqslant (1 + 2\alpha\delta) \cdot \phi(P_i) + 2\alpha\delta.$$

[1] The k-means algorithm returns a partition of $\mathcal{X}_V$. One may assume w.l.o.g. that all copies of $F(u)$ are put into the same cluster of $\mathcal{X}_V$. Thus the algorithm also partitions V.

Under the stronger gap assumption $\Upsilon = 2 \cdot 10^5 \cdot k^5/\delta$, they showed how to obtain a partition with essentially the guarantee stated in Theorem 1.1 in time $O\left(m \cdot \operatorname{poly}\log(n)\right)$. We use $m = |E|$ for the number of edges of G and $n = |V|$ for the number of nodes. The algorithmic result does not use the standard paradigm for spectral embedding. It uses the heat-kernel for the geometric embedding and uses locality sensitive hashing for the clustering and hence does not really explain the success of the standard paradigm for spectral embedding.

Our Contribution: We strengthen the approximation guarantees in Theorem 1.1 by a factor of k and simultaneously weaken the gap assumption. As a consequence, the variant of Lloyd's k-means algorithm analyzed by Ostrovsky et al. [4] applied to[2] $\widetilde{\mathcal{X}_V}$ achieves the improved approximation guarantees in time $O(m(k^2 + (\ln)/\lambda_{k+1}))$ with constant probability. Table 1 summarizes these results.

Let $\mathcal{O}$ be the set of all k-way partitions $(P_1, \ldots, P_k)$ with $\Phi(P_1, \ldots, P_k) = \widehat{\rho}(k)$, i.e., the set of all partitions that achieve the order k partition constant. Let

$$\widehat{\rho}_{\mathrm{avr}}(k) \triangleq \min_{(P_1,\ldots,P_k)\in\mathcal{O}} \frac{1}{k}\sum_{i=1}^{k} \phi(P_i)$$

be the *minimal average conductance* over all k-way partitions in $\mathcal{O}$. Our gap assumption is defined in terms of

$$\Psi \triangleq \frac{\lambda_{k+1}}{\widehat{\rho}_{\mathrm{avr}}(k)}.$$

For the remainder of this paper we denote by $(P_1, \ldots, P_k)$ a k-way partition of V that achieves $\widehat{\rho}_{\mathrm{avr}}(k)$. We can now state our main result.

Theorem 1.2 (Main Theorem)**.** *a) (Existence of a Good Clustering) Let G be a graph satisfying*

$$\Psi = 20^4 \cdot k^3/\delta \tag{4}$$

for some $\delta \in (0, 1/2]$ and $k \geqslant 3$ and let $(A_1, \ldots, A_k)$ be the k-way partition output by an α-approximate clustering algorithm applied to the spectral embedding $\mathcal{X}_V$. Then for every $i \in [1:k]$ the following two statements hold (after suitable renumbering of one of the partitions):

$$1)\ \mu(A_i \triangle P_i) \leqslant \frac{\alpha\delta}{10^3 k} \cdot \mu(P_i) \quad \textit{and} \quad 2)\ \phi(A_i) \leqslant \left(1 + \frac{2\alpha\delta}{10^3 k}\right) \cdot \phi(P_i) + \frac{2\alpha\delta}{10^3 k}.$$

b) (An Efficient Algorithm) If in addition $k/\delta \geqslant 10^9$ and[3] $\triangle_k(\mathcal{X}_V) \geqslant n^{-O(1)}$, then the variant of Lloyd's algorithm analyzed by Ostrovsky et al. [4] applied to $\widetilde{\mathcal{X}_V}$ returns in time $O(m(k^2 + (\ln n)/\lambda_{k+1}))$ with constant probability a partition $(A_1, \ldots, A_k)$ such that for every $i \in [1:k]$ the following two statements hold (after suitable renumbering of one of the partitions):

$$3)\ \mu(A_i \triangle P_i) \leqslant \frac{2\delta}{10^3 k} \cdot \mu(P_i) \quad \textit{and} \quad 4)\ \phi(A_i) \leqslant \left(1 + \frac{4\delta}{10^3 k}\right) \cdot \phi(P_i) + \frac{4\delta}{10^3 k}.$$

Part (b) of the Theorem gives theoretical support for the practical success of spectral clustering based on spectral embedding followed by k-means clustering. Previous papers [2, 6] replaced k-means clustering by other techniques for their algorithmic results.

	Gap Assumption	Partition Quality	Running Time
Peng et al. [6]	$\Upsilon = 2 \cdot 10^5 \cdot k^3/\delta$	$\mu(A_i \triangle P_i) \leqslant \alpha\delta \cdot \mu(P_i)$ $\phi(A_i) \leqslant (1 + 2\alpha\delta)\, \phi(P_i) + 2\alpha\delta$	Existential result
This paper	$\Psi = 20^4 \cdot k^3/\delta$	$\mu(A_i \triangle P_i) \leqslant \frac{\alpha\delta}{10^3 k} \cdot \mu(P_i)$ $\phi(A_i) \leqslant \left(1 + \frac{2\alpha\delta}{10^3 k}\right) \phi(P_i) + \frac{2\alpha\delta}{10^3 k}$	Existential result
Peng et al. [6]	$\Upsilon = 2 \cdot 10^5 \cdot k^5/\delta$	$\mu(A_i \triangle P_i) \leqslant \frac{\delta \log^2 k}{k^2} \cdot \mu(P_i)$ $\phi(A_i) \leqslant \left(1 + \frac{2\delta \log^2 k}{k^2}\right) \phi(P_i) + \frac{2\delta \log^2 k}{k^2}$	$O\left(m \cdot \mathrm{poly}\log(n)\right)$
This paper	$\Psi = 20^4 \cdot k^3/\delta$ $k/\delta \geqslant 10^9$ $\triangle_k(\mathcal{X}_V) \geqslant n^{-O(1)}$	$\mu(A_i \triangle P_i) \leqslant \frac{2\delta}{10^3 k} \cdot \mu(P_i)$ $\phi(A_i) \leqslant \left(1 + \frac{4\delta}{10^3 k}\right) \phi(P_i) + \frac{4\delta}{10^3 k}$	$O\left(m\left(k^2 + \frac{\ln n}{\lambda_{k+1}}\right)\right)$

Table 1: A comparison of the results in Peng et al. [6] and our results. The parameter $\delta \in (0, 1/2]$ relates the approximation guarantees with the gap assumption.

If $k \leqslant \mathrm{poly}(\log \mathrm{n})$ and $\lambda_{k+1} \geqslant \mathrm{poly}(\log n)$, our algorithm works in nearly linear time.

The k-means algorithm in [4] is efficient only for inputs $\mathcal{X}$ for which some partition into k clusters is much better than any partition into $k-1$ clusters; formally, for inputs $\mathcal{X}$ satisfying $\triangle_k(\mathcal{X}) \leqslant \varepsilon^2 \cdot \triangle_{k-1}(\mathcal{X})$ for some $\varepsilon \in (0, 6 \cdot 10^{-7}]$. For the proof of Part (b), we show in Section 10 that $\widetilde{\mathcal{X}_V}$ satisfies this assumption.

The *order k conductance constant* $\rho(k)$ is defined by

$$\rho(k) = \min_{\text{disjoint nonempty } Z_1,\ldots,Z_k} \Phi(Z_1, \ldots, Z_k), \quad \text{where} \quad \Phi(Z_1, \ldots, Z_k) = \max_{i \in [1:k]} \phi(Z_i). \tag{5}$$

Lee et al. [2] connected $\rho(k)$ and the k-th smallest eigenvalue of the normalized Laplacian matrix $\mathcal{L}_G$ through the relation

$$\lambda_k/2 \leqslant \rho(k) \leqslant O(k^2)\sqrt{\lambda_k}, \tag{6}$$

and Oveis Gharan and Trevisan [5] showed

$$\widehat{\rho}(k) \leqslant k\rho(k). \tag{7}$$

In Section 11, we establish an analogous relation for $\widehat{\rho}_{\mathrm{avr}}(k)$.

The Proof of Part (a) of the Main Theorem: The proof of Part $(a.1)$ builds upon the following Lemmas that we will prove in Section 7 and Section 8, respectively. Recall that $\mathcal{X}_V$ contains d_u copies of $F(u)$ for each $u \in V$. W.l.o.g. we may restrict attention to clusterings of $\mathcal{X}_V$ that put all copies of $F(u)$ into the same cluster and hence induce a clustering of V. Let $(A_1, \ldots, A_k)$ with cluster centers c_1 to c_k be a clustering of V. Its k-means cost is

$$\mathrm{Cost}(\{A_i, c_i\}_{i=1}^k) = \sum_{i=1}^k \sum_{u \in A_i} d_u \left\| F(u) - c_i \right\|^2.$$

[2] $\widetilde{\mathcal{X}_V}$ is defined as $\mathcal{X}_V$ but in terms of approximate eigenvectors, see Section 10.1

[3] The case $\triangle_k(\mathcal{X}_V) \leqslant n^{-O(1)}$ constitutes a trivial clustering problem. For technical reasons, we have to exclude too easy inputs.

Lemma 1.3 (($P_1, \dots, P_k$) is a good k-means partition)**.** *If $\Psi > 4 \cdot k^{3/2}$ then there are vectors $\{p^{(i)}\}_{i=1}^k$ such that*

$$\mathrm{Cost}(\{P_i, p^{(i)}\}_{i=1}^k) \leqslant \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{k^2}{\Psi}.$$

Lemma 1.4 (Only partitions close to $(P_1, \dots, P_k)$ are good)**.** *Under the hypothesis of Theorem 1.2, the following holds. If for every permutation $\sigma : [1:k] \to [1:k]$ there exists an index $i \in [1:k]$ such that*

$$\mu(A_i \triangle P_{\sigma(i)}) \geqslant \frac{8\alpha\delta}{10^4 k} \cdot \mu(P_{\sigma(i)}).$$

Then it holds that

$$\mathrm{Cost}(\{A_i, c_i\}_{i=1}^k) > \frac{2\alpha k^2}{\Psi}.$$

Substituting these bounds into (2) yields a contradiction, since

$$\frac{2\alpha k^2}{\Psi} < \mathrm{Cost}(\{A_i, c_i\}_{i=1}^k) \leqslant \alpha \cdot \triangle_k(\mathcal{X}_V) \leqslant \alpha \cdot \mathrm{Cost}(\{P_i, p^{(i)}\}_{i=1}^k) \leqslant \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{\alpha k^2}{\Psi}.$$

Therefore, there exists a permutation π (the identity after suitable renumbering of one of the partitions) such that $\mu(A_i \triangle P_i) < \frac{8\alpha\delta}{10^4 k} \cdot \mu(P_i)$ for all $i \in [1:k]$.

Part $(a.2)$ follows from Part $(a.1)$. Indeed, for $\delta' = 8\delta/10^4$ we have

$$\mu(A_i) \geqslant \mu(P_i \cap A_i) = \mu(P_i) - \mu(P_i \setminus A_i) \geqslant \mu(P_i) - \mu(A_i \triangle P_i) \geqslant \left(1 - \frac{\alpha\delta'}{k}\right) \cdot \mu(P_i)$$

and $|E(A_i, \overline{A_i})| \leqslant |E(P_i, \overline{P_i})| + \mu(A_i \Delta P_i)$ since every edge that is counted in $|E(A_i, \overline{A_i})|$ but not in $|E(P_i, \overline{P_i})|$ must have an endpoint in $A_i \Delta P_i$. Thus

$$\Phi(A_i) = \frac{|E(A_i, \overline{A_i})|}{\mu(A_i)} \leqslant \frac{|E(P_i, \overline{P_i})| + \frac{\alpha\delta'}{k} \cdot \mu(P_i)}{(1 - \frac{\alpha \cdot \delta'}{k}) \cdot \mu(P_i)} \leqslant \left(1 + \frac{2\alpha\delta'}{k}\right) \cdot \phi(P_i) + \frac{2\alpha\delta'}{k}.$$

This completes the proof of Part (a) of Theorem 1.2.

2 Notations

We use the notation adopted by Peng et al. [6] and restate it below for completeness. Let $\mathcal{L}_G = I - D^{-1/2} A D^{-1/2}$ be a normalized Laplacian matrix, where D is diagonal degree matrix and A is adjacency matrix. We refer to the j-th eigenvalue of matrix $\mathcal{L}_G$ by $\lambda_j \triangleq \lambda_j(\mathcal{L}_G)$. The (unit) eigenvector corresponding to λ_j is denoted by f_j.

Let $\overline{g_i} = \frac{D^{1/2}\chi_{P_i}}{\|D^{1/2}\chi_{P_i}\|}$, where χ_{P_i} is the characteristic vector of a subset $P_i \subseteq V$. Note $\overline{g_i}$ is the normalized characteristic vector of P_i and that $\left\|D^{1/2}\chi_{P_i}\right\|^2 = \sum_{v \in P_i} \deg_v = \mu(P_i)$. We will write μ_i instead of $\mu(P_i)$. The Rayleigh quotient is defined by and satisfies that

$$\mathcal{R}\left(\overline{g_i}\right) \triangleq \frac{\overline{g_i}^{\mathrm{T}} \mathcal{L}_G \overline{g_i}}{\overline{g_i}^{\mathrm{T}} \overline{g_i}} = \frac{1}{\mu(P_i)} \chi_{P_i}^{\mathrm{T}} L \chi_{P_i} = \frac{|E(S, \overline{S})|}{\mu(P_i)} = \phi_{P_i},$$

where $L = D - A$ is the graph Laplacian matrix.

$$
\begin{tikzcd}
\widehat{f_i} = \sum_{j=1}^{k} \alpha_j^{(i)} f_j \arrow[r, dash, "\text{Lemma 4.3}", "\|\widehat{f_i} - \overline{g_i}\|^2 \leqslant \phi_{P_i}/\lambda_{k+1}"'] \arrow[d, dash] & \overline{g_i} = \frac{D^{1/2}\chi_{P_i}}{\sqrt{\mu(P_i)}} = \sum_{j=1}^{n} \alpha_j^{(i)} f_j \arrow[d, dash] \\
f_i = \sum_{j=1}^{k} \beta_j^{(i)} \widehat{f_j} \arrow[r, dash, "\text{Theorem 4.1}", "\|f_i - \widehat{g_i}\|^2 \leqslant (1+3k/\Psi)\cdot k/\Psi"'] & \widehat{g_i} = \sum_{j=1}^{k} \beta_j^{(i)} \overline{g_j}
\end{tikzcd}
$$

Figure 1: The relation between the vectors f_i, $\widehat{f_i}$, $\widehat{g_i}$ and $\overline{g_i}$. The vectors $\{f_i\}_{i=1}^n$ are eigenvectors of the normalized Laplacian matrix $\mathcal{L}_G$ of a graph G satisfying $\Psi > 4 \cdot k^{3/2}$. The vectors $\{\overline{g}_i\}_{i=1}^k$ are the normalized characteristic vectors of an optimal partition $(P_1, \ldots, P_k)$. For each $i \in [1:k]$ the vector $\widehat{f_i}$ is the projection of vector $\overline{g_i}$ onto $\mathrm{span}(f_1, \ldots, f_k)$. By Lemma 4.3 the vectors $\widehat{f_i}$ and $\overline{g}_i$ are close for $i \in [1:k]$. By Lemma 4.2 it holds $\mathrm{span}(f_1, \ldots, f_k) = \mathrm{span}(\widehat{f_1}, \ldots, \widehat{f_k})$ when $\Psi > 4 \cdot k^{3/2}$, and thus we can write $f_i = \sum_{j=1}^k \beta_j^{(i)} \widehat{f}_j$. Moreover, by Theorem 4.1 the vectors f_i and $\widehat{g_i} = \sum_{j=1}^k \beta_j^{(i)} \overline{g_j}$ are close for $i \in [1:k]$.

The eigenvectors $\{f_i\}_{i=1}^n$ form an orthonormal basis of $\mathbb{R}^n$. Thus each characteristic vector $\overline{g_i}$ can be expressed as $\overline{g_i} = \sum_{j=1}^n \alpha_j^{(i)} f_j$ for all $i \in [1:k]$. We define its *projection* onto the first k eigenvectors by $\widehat{f_i} = \sum_{j=1}^k \alpha_j^{(i)} f_j$.

Peng et al. [6] showed that $\mathrm{span}(\{\widehat{f_i}\}_{i=1}^k) = \mathrm{span}(\{f_i\}_{i=1}^k)$ if the gap parameter Υ is large enough. In Lemma 4.2 we demonstrate that similar statement holds with substituted gap parameter Ψ. This implies that each of the first k eigenvectors can be expressed by $f_i = \sum_{j=1}^k \beta_j^{(i)} \widehat{f_j}$. Moreover, Peng et al. [6] showed that each vector

$$\widehat{g_i} = \sum_{j=1}^{k} \beta_j^{(i)} \overline{g_j}$$

approximates the eigenvector f_i for all $i \in [1:k]$, if Υ is large. We prove in Theorem 4.1 that it suffices to have a large gap parameter Ψ.

In the proof of Lemma 1.3, we will use the vectors

$$p^{(i)} = \frac{1}{\sqrt{\mu(P_i)}} \left(\beta_i^{(1)}, \ldots, \beta_i^{(k)}\right)^{\mathrm{T}}. \tag{8}$$

For any vertex $u \in P_i$, we have

$$p^{(i)} = \left(\left[D^{-1/2}\widehat{g_1}\right](u), \ldots, \left[D^{-1/2}\widehat{g_k}\right](u)\right). \tag{9}$$

Indeed, for any $h \in [1:k]$,

$$D^{-1/2}\widehat{g}_h(u) = \sum_{1 \leqslant j \leqslant k} \beta_j^{(h)} D^{-1/2} \frac{D^{1/2}\chi_{P_i}}{\sqrt{\mu(P_i)}}(u) = \frac{1}{\sqrt{\mu(P_i)}} \beta_i^{(h)}.$$

Our analysis builds upon the following two matrices. Let $\mathbf{F}, \mathbf{B} \in \mathbb{R}^{k \times k}$ be square matrices such that for all indices $i, j \in [1:k]$ we have

$$\mathbf{F}_{j,i} = \alpha_j^{(i)} \quad \text{and} \quad \mathbf{B}_{j,i} = \beta_j^{(i)}. \tag{10}$$

3 Technical Advances in The Improved Quality Guarantee

In Section 4, we show that if $\Psi > 4 \cdot k^{3/2}$ then the vectors $\widehat{g}_i$ and f_i are close for all $i \in [1:k]$, i.e.,

$$\|f_i - \widehat{g}_i\|^2 \leqslant \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{k}{\Psi}.$$

The proof follows [6] but our analysis depends on the less restrictive gap parameter Ψ.

In contrast to [6] we exhibit in Section 5 key spectral properties of the matrices $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ and $\mathbf{B}\mathbf{B}^{\mathrm{T}}$. More precisely, we show that they are close to the identity matrix. For our improved quality guarantee we use the fact that if $\Psi \geqslant 10^4 \cdot k^3/\varepsilon^2$ and $\varepsilon \in (0,1)$ then for all distinct $i, j \in [1:k]$ it holds

$$1 - \varepsilon \leqslant \langle \mathbf{B}_{i,:}, \mathbf{B}_{i,:} \rangle \leqslant 1 + \varepsilon \quad \text{and} \quad |\langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:} \rangle| \leqslant \sqrt{\varepsilon}. \tag{11}$$

Peng et al. (c.f. [6, Lemma 4.2]) proved that the square Euclidean distance between any distinct estimation centers satisfies

$$\left\| p^{(i)} - p^{(j)} \right\|^2 \geqslant \left[10^3 \cdot k \cdot \min\{\mu(P_i), \mu(P_j)\}\right]^{-1}.$$

In Section 6, we improve their result by a factor of k. Our analysis depends on the less restrictive gap assumption $\Psi \geqslant 20^4 \cdot k^3$ and builds upon (11). We show in Lemma 6.2 that for all distinct $i, j \in [1:k]$ it holds

$$\left\| p^{(i)} - p^{(j)} \right\|^2 \geqslant \left[3 \cdot \min\{\mu(P_i), \mu(P_j)\}\right]^{-1}.$$

We prove Lemma 1.3 in Section 7 and Lemma 1.4 in Section 8. The analysis of these Lemmas builds upon the results from Section 4 to Section 6.

4 Vectors $\widehat{g}_i$ and f_i are Close

In this section we prove Theorem 4.1. We argue in a similar manner as in [6], however, in terms of Ψ instead of Υ. For completeness, we show in Subsection 4.1 that the span of the first k eigenvectors is equal to the span of the projections of the characteristic vectors of subsets P_i onto the first k eigenvectors. Then in Subsection 4.2 by expressing the eigenvectors f_i in terms of the vectors $\widehat{f}_i$ we conclude the proof of Theorem 4.1.

Theorem 4.1. *If $\Psi > 4 \cdot k^{3/2}$ then the vectors $\widehat{g}_i = \sum_{j=1}^{k} \beta_j^{(i)} \overline{g_j}$, $i \in [1:k]$, satisfy*

$$\|f_i - \widehat{g}_i\|^2 \leqslant \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{k}{\Psi}.$$

4.1 Analyzing the Columns of Matrix F

We prove in this subsection the following result that depends on gap parameter Ψ.

Lemma 4.2. *If $\Psi > k^{3/2}$ then the $\mathrm{span}(\{\widehat{f}_i\}_{i=1}^k) = \mathrm{span}(\{f_i\}_{i=1}^k)$ and thus each eigenvector can be expressed as $f_i = \sum_{j=1}^{k} \beta_j^{(i)} \cdot \widehat{f}_j$ for every $i \in [1:k]$.*

To prove Lemma 4.2 we build upon the following result shown by Peng et al. [6].

Lemma 4.3. *[6, Theorem 1.1 Part 1] For $P_i \subset V$ let $\overline{g_i} = \frac{D^{1/2}\chi_{P_i}}{\|D^{1/2}\chi_{P_i}\|}$. Then any $i \in [1:k]$ it holds that*

$$\left\|\overline{g_i} - \widehat{f}_i\right\|^2 = \sum_{j=k+1}^{n}\left(\alpha_j^{(i)}\right)^2 \leqslant \frac{\mathcal{R}\left(\overline{g_i}\right)}{\lambda_{k+1}} = \frac{\phi(P_i)}{\lambda_{k+1}}.$$

Based on the following two results we prove Lemma 4.2.

Lemma 4.4. *For every $i \in [1:k]$ and $p \neq q \in [1:k]$ it holds that*

$$1 - \phi(P_i)/\lambda_{k+1} \leqslant \left\|\widehat{f}_i\right\|^2 = \left\|\alpha^{(i)}\right\|^2 \leqslant 1 \quad \textit{and} \quad \left|\left\langle \widehat{f}_p, \widehat{f}_q\right\rangle\right| = \left|\langle \alpha^p, \alpha^q\rangle\right| \leqslant \frac{\sqrt{\phi(P_p)\cdot\phi(P_q)}}{\lambda_{k+1}}.$$

Proof. The first part follows by Lemma 4.3 and the following chain of inequalities

$$1 - \frac{\phi(P_i)}{\lambda_{k+1}} \leqslant 1 - \sum_{j=k+1}^{n}\left(\alpha_j^{(i)}\right)^2 = \left\|\widehat{f}_i\right\|^2 = \sum_{j=1}^{k}\left(\alpha_j^{(i)}\right)^2 \leqslant \sum_{j=1}^{n}\left(\alpha_j^{(i)}\right)^2 = 1.$$

We show now the second part. Since $\{f_i\}_{i=1}^n$ are orthonormal eigenvectors we have for all $p \neq q$ that

$$\langle f_p, f_q\rangle = \sum_{l=1}^{n}\alpha_l^{(p)}\cdot\alpha_l^{(q)} = 0. \tag{12}$$

We combine (12) and Cauchy-Schwarz to obtain

$$\begin{aligned}\left|\left\langle \widehat{f}_p, \widehat{f}_q\right\rangle\right| &= \left|\sum_{l=1}^{k}\alpha_l^{(p)}\cdot\alpha_l^{(q)}\right| = \left|\sum_{l=k+1}^{n}\alpha_l^{(p)}\cdot\alpha_l^{(q)}\right| \\ &\leqslant \sqrt{\sum_{l=k+1}^{n}\left(\alpha_l^{(p)}\right)^2}\cdot\sqrt{\sum_{l=k+1}^{n}\left(\alpha_l^{(q)}\right)^2} \leqslant \frac{\sqrt{\phi(P_p)\cdot\phi(P_q)}}{\lambda_{k+1}}.\end{aligned}$$

∎

Lemma 4.5. *If $\Psi > k^{3/2}$ then the columns $\{\mathbf{F}_{:,i}\}_{i=1}^k$ are linearly independent.*

Proof. We show that the columns of matrix $\mathbf{F}$ are almost orthonormal. Consider the symmetric matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$. It is known that $ker\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}\right) = ker(\mathbf{F})$ and that all eigenvalues of matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ are real numbers. We proceeds by showing that the smallest eigenvalue $\lambda_{\min}(\mathbf{F}^{\mathrm{T}}\mathbf{F}) > 0$. This would imply that $ker(\mathbf{F}) = \emptyset$ and hence yields the statement.

By combining Gersgorin Circle Theorem, Lemma 4.4 and Cauchy-Schwarz it holds that

$$\begin{aligned}\lambda_{\min}(\mathbf{F}^{\mathrm{T}}\mathbf{F}) &\geqslant \min_{i\in[1:k]}\left\{\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}\right)_{ii} - \sum_{j\neq i}^{k}\left|\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}\right)_{ij}\right|\right\} = \min_{i\in[1:k]}\left\{\left\|\alpha^{(i)}\right\|^2 - \sum_{j\neq i}^{k}\left|\left\langle\alpha^{(j)},\alpha^{(i)}\right\rangle\right|\right\} \\ &\geqslant 1 - \sum_{j=1}^{k}\sqrt{\frac{\phi(P_j)}{\lambda_{k+1}}}\sqrt{\frac{\phi(P_{i^\star})}{\lambda_{k+1}}} \geqslant 1 - \sqrt{k}\sqrt{\sum_{j=1}^{k}\frac{\phi(P_j)}{\lambda_{k+1}}}\sqrt{\frac{\phi(P_{i^\star})}{\lambda_{k+1}}} \geqslant 1 - \frac{k^{3/2}}{\Psi} > 0,\end{aligned}$$

where $i^\star \in [1:k]$ is the index that minimizes the expression above. ∎

We present now the proof of Lemma 4.2.

Proof of Lemma 4.2. Let λ be an arbitrary non-zero vector. Notice that

$$\sum_{i=1}^{k} \lambda_i \cdot \widehat{f}_i = \sum_{i=1}^{k} \lambda_i \sum_{j=1}^{k} \alpha_j^{(i)} f_j = \sum_{j=1}^{k} \left(\sum_{i=1}^{k} \lambda_i \alpha_j^{(i)} \right) f_j = \sum_{j=1}^{k} \gamma_j f_j, \quad \text{where} \quad \gamma_j = \langle \mathbf{F}_{j,:}, \lambda \rangle . \tag{13}$$

By Lemma 4.5 the columns $\{\mathbf{F}_{:,i}\}_{i=1}^{k}$ are linearly independent and since $\gamma = \mathbf{F}\lambda$, it follows at least one component $\gamma_j \neq 0$. Therefore the vectors $\left\{ \widehat{f}_i \right\}_{i=1}^{k}$ are linearly independent and span $\mathbb{R}^k$. ∎

4.2 Analyzing Eigenvectors f in terms of $\widehat{f}_j$

To prove Theorem 4.1 we establish next the following result.

Lemma 4.6. *If $\Psi > k^{3/2}$ then for $i \in [k]$ it holds*

$$\left(1 + \frac{2k}{\Psi}\right)^{-1} \leqslant \sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2 \leqslant \left(1 - \frac{2k}{\Psi}\right)^{-1}.$$

Proof. We show now the upper bound. By Lemma 4.2 $f_i = \sum_{j=1}^{k} \beta_j^{(i)} \widehat{f}_j$ for all $i \in [1:k]$ and thus

$$\begin{aligned} 1 &= \|f_i\|^2 = \left\langle \sum_{a=1}^{k} \beta_a^{(i)} \widehat{f}_a, \sum_{b=1}^{k} \beta_b^{(i)} \widehat{f}_b \right\rangle \\ &= \sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2 \left\| \widehat{f}_j \right\|^2 + \sum_{a=1}^{k} \sum_{b \neq a}^{k} \beta_a^{(i)} \beta_b^{(i)} \left\langle \widehat{f}_a, \widehat{f}_b \right\rangle \\ &\overset{(\star)}{\geqslant} \left(1 - \frac{2k}{\Psi}\right) \cdot \sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2. \end{aligned}$$

To prove the inequality $(\star)$ we consider the two terms separately.

By Lemma 4.4, $\left\| \widehat{f}_j \right\|^2 \geqslant 1 - \phi(P_j)/\lambda_{k+1}$. We then apply $\sum_i a_i b_i \leqslant (\sum_i a_i)(\sum_i b_i)$ for all non-negative vectors a, b and obtain

$$\sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2 \left(1 - \frac{\phi(P_j)}{\lambda_{k+1}}\right) = \sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2 - \sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2 \frac{\phi(P_j)}{\lambda_{k+1}} \geqslant \left(1 - \frac{k}{\Psi}\right) \sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2.$$

Again by Lemma 4.4, we have $\left| \left\langle \widehat{f}_a, \widehat{f}_b \right\rangle \right| \leqslant \sqrt{\phi(P_a)\phi(P_b)}/\lambda_{k+1}$, and by Cauchy-Schwarz it holds

$$\begin{aligned} \sum_{a=1}^{k} \sum_{b \neq a}^{k} \beta_a^{(i)} \beta_b^{(i)} \left\langle \widehat{f}_a, \widehat{f}_b \right\rangle &\geqslant -\sum_{a=1}^{k} \sum_{b \neq a}^{k} \left|\beta_a^{(i)}\right| \cdot \left|\beta_b^{(i)}\right| \cdot \left| \left\langle \widehat{f}_a, \widehat{f}_b \right\rangle \right| \\ &\geqslant -\frac{1}{\lambda_{k+1}} \sum_{a=1}^{k} \sum_{b \neq a}^{k} \left|\beta_a^{(i)}\right| \sqrt{\phi(P_a)} \cdot \left|\beta_b^{(i)}\right| \sqrt{\phi(P_b)} \\ &\geqslant -\frac{1}{\lambda_{k+1}} \left(\sum_{j=1}^{k} \left|\beta_j^{(i)}\right| \sqrt{\phi(P_j)} \right)^2 \geqslant -\frac{k}{\Psi} \cdot \sum_{j=1}^{k} \left(\beta_j^{(i)}\right)^2. \end{aligned}$$

The lower bound follows by analogous arguments. ∎

We are ready now to prove Theorem 4.1.

Proof of Theorem 4.1. By Lemma 4.2, we have $f_i = \sum_{j=1}^{k} \beta_j^{(i)} \widehat{f_j}$ and recall that $\widehat{g_i} = \sum_{j=1}^{k} \beta_j^{(i)} \overline{g_j}$ for all $i \in [1:k]$. We combine triangle inequality, Cauchy-Schwarz, Lemma 4.3 and Lemma 4.6 to obtain

$$\begin{aligned} \|f_i - \widehat{g_i}\|^2 &= \left\| \sum_{j=1}^{k} \beta_j^{(i)} \left(\widehat{f_j} - \overline{g_j} \right) \right\|^2 \leqslant \left(\sum_{j=1}^{k} |\beta_j^i| \cdot \left\| \widehat{f_j} - \overline{g_j} \right\| \right)^2 \\ \leqslant & \left(\sum_{j=1}^{k} \left(\beta_j^{(i)} \right)^2 \right) \cdot \left(\sum_{j=1}^{k} \left\| \widehat{f_j} - \overline{g_j} \right\|^2 \right) \leqslant \left(1 - \frac{2k}{\Psi} \right)^{-1} \left(\frac{1}{\lambda_{k+1}} \sum_{j=1}^{k} \phi(P_j) \right) \\ = & \left(1 - \frac{2k}{\Psi} \right)^{-1} \cdot \frac{k}{\Psi} \leqslant \left(1 + \frac{3k}{\Psi} \right) \cdot \frac{k}{\Psi}, \end{aligned}$$

where the last inequality uses $\Psi > 4 \cdot k$. ■

5 Spectral Properties of Matrix B

In this section we bound the inner product of any two rows of matrix $\mathbf{B}$ (c.f. Equation 10).

Theorem 5.1. *If $\Psi \geqslant 10^4 \cdot k^3/\varepsilon^2$ and $\varepsilon \in (0,1)$ then for all distinct $i, j \in [1:k]$ it holds*

$$1 - \varepsilon \leqslant \langle \mathbf{B}_{i,:}, \mathbf{B}_{i,:} \rangle \leqslant 1 + \varepsilon \quad \textit{and} \quad |\langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:} \rangle| \leqslant \sqrt{\varepsilon}.$$

The proof is divided into two parts. We show in Lemma 5.4 that $1 - \varepsilon \leqslant \langle \mathbf{B}_{i,:}, \mathbf{B}_{i,:} \rangle \leqslant 1 + \varepsilon$, and we establish the second statement $|\langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:} \rangle| \leqslant \sqrt{\varepsilon}$ in Lemma 5.5.

5.1 Analyzing the Column Space of Matrix B

We show below that the matrix $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ is close to the identity matrix.

Lemma 5.2. *(Columns) If $\Psi > 4 \cdot k^{3/2}$ then for all distinct $i, j \in [1:k]$ it holds*

$$1 - \frac{3k}{\Psi} \leqslant \langle \mathbf{B}_{:,i}, \mathbf{B}_{:,i} \rangle \leqslant 1 + \frac{3k}{\Psi} \quad \textit{and} \quad |\langle \mathbf{B}_{:,i}, \mathbf{B}_{:,j} \rangle| \leqslant 4\sqrt{\frac{k}{\Psi}}.$$

Proof. By Lemma 4.6 it holds that

$$1 - \frac{3k}{\Psi} \leqslant \langle \mathbf{B}_{:,i}, \mathbf{B}_{:,i} \rangle = \sum_{j=1}^{k} \left(\beta_j^{(i)} \right)^2 \leqslant 1 + \frac{3k}{\Psi}.$$

Recall that $\widehat{g_i} = \sum_{j=1}^{k} \beta_j^{(i)} \cdot \overline{g_j}$. Moreover, since the eigenvectors $\{f_i\}_{i=1}^{k}$ and the characteristic vectors $\{\overline{g_i}\}_{i=1}^{k}$ are orthonormal by combing Cauchy-Schwarz and by Theorem 4.1 it holds

$$\begin{aligned} |\langle \mathbf{B}_{:,i}, \mathbf{B}_{:,j} \rangle| &= \sum_{l=1}^{k} \beta_l^{(i)} \beta_l^{(j)} = \left\langle \sum_{a=1}^{k} \beta_a^{(i)} \cdot \overline{g_a}, \sum_{b=1}^{k} \beta_b^{(j)} \cdot \overline{g_b} \right\rangle = \langle \widehat{g_i}, \widehat{g_j} \rangle \\ &= \langle (\widehat{g_i} - f_i) + f_i, (\widehat{g_j} - f_j) + f_j \rangle \\ &= \langle \widehat{g_i} - f_i, \widehat{g_j} - f_j \rangle + \langle \widehat{g_i} - f_i, f_j \rangle + \langle f_i, \widehat{g_j} - f_j \rangle \\ &\leqslant \|\widehat{g_i} - f_i\| \cdot \|\widehat{g_j} - f_j\| + \|\widehat{g_i} - f_i\| + \|\widehat{g_j} - f_j\| \\ &\leqslant \left(1 + \frac{3k}{\Psi} \right) \cdot \frac{k}{\Psi} + 2\sqrt{\left(1 + \frac{3k}{\Psi} \right) \cdot \frac{k}{\Psi}} \leqslant 4\sqrt{\frac{k}{\Psi}}. \end{aligned}$$

$\blacksquare$

Using a stronger gap assumption we show that the columns of matrix $\mathbf{B}$ are linearly independent.

Lemma 5.3. *If $\Psi > 25 \cdot k^3$ then the columns $\{\mathbf{B}_{:,i}\}_{i=1}^{k}$ are linearly independent.*

Proof. Since $ker(\mathbf{B}) = ker\left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)$ and $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ is SPSD[4] matrix, it suffices to show that the smallest eigenvalue

$$\lambda(\mathbf{B}^{\mathrm{T}}\mathbf{B}) = \min_{x \neq 0} \frac{x^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}x}{x^{\mathrm{T}}x} > 0.$$

By Lemma 5.2,

$$\sum_{i=1}^{k}\sum_{j\neq i}^{k} |x_i|\,|x_j| \left|\left\langle \beta^{(i)}, \beta^{(j)} \right\rangle\right| \leqslant 4\sqrt{\frac{k}{\Psi}} \left(\sum_{i=1}^{k} |x_i|\right)^2 \leqslant \|x\|^2 \cdot 4k\sqrt{\frac{k}{\Psi}},$$

and

$$\begin{aligned} x^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{B}x &= \left\langle \sum_{i=1}^{k} x_i\beta^{(i)}, \sum_{j=1}^{k} x_j\beta^{(j)} \right\rangle = \sum_{i=1}^{k} x_i^2 \left\|\beta^{(i)}\right\|^2 + \sum_{i=1}^{k}\sum_{j\neq i}^{k} x_i x_j \left\langle \beta^{(i)}, \beta^{(j)} \right\rangle \\ &\geqslant \left(1 - \frac{3k}{\Psi}\right)\|x\|^2 - \sum_{i=1}^{k}\sum_{j\neq i}^{k} |x_i|\,|x_j| \left|\left\langle \beta^{(i)}, \beta^{(j)} \right\rangle\right| \geqslant \left(1 - 5k\sqrt{\frac{k}{\Psi}}\right) \cdot \|x\|^2. \end{aligned}$$

Therefore $\lambda(\mathbf{B}^{\mathrm{T}}\mathbf{B}) > 0$ and the statement follows. $\blacksquare$

5.2 Analyzing the Row Space of Matrix B

In this subsection we show that the matrix $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ is close to the identity matrix. We bound now the squared L_2 norm of the rows in matrix $\mathbf{B}$, i.e. the diagonal entries in matrix $\mathbf{B}\mathbf{B}^{\mathrm{T}}$.

Lemma 5.4. *(Rows) If $\Psi \geqslant 400 \cdot k^3/\varepsilon^2$ and $\varepsilon \in (0,1)$ then for all distinct $i, j \in [1:k]$ it holds*

$$1 - \varepsilon \leqslant \langle \mathbf{B}_{i,:}, \mathbf{B}_{i,:} \rangle \leqslant 1 + \varepsilon.$$

Proof. We show that the eigenvalues of matrix $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ are concentrated around 1. This would imply that $\chi_i^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\chi_i = \langle \mathbf{B}_{i,:}, \mathbf{B}_{i,:} \rangle \approx 1$, where χ_i is a characteristic vector. By Lemma 5.2 we have

$$\left(1 - \frac{3k}{\Psi}\right)^2 \leqslant \left(\beta^{(i)}\right)^{\mathrm{T}} \cdot \mathbf{B}\mathbf{B}^{\mathrm{T}} \cdot \beta^{(i)} = \left\|\beta^{(i)}\right\|^4 + \sum_{j\neq i}^{k} \left\langle \beta^{(j)}, \beta^{(i)} \right\rangle^2 \leqslant \left(1 + \frac{3k}{\Psi}\right)^2 + \frac{16k^2}{\Psi} \leqslant 1 + \frac{23k^2}{\Psi}$$

and

$$\left|\left(\beta^{(i)}\right)^{\mathrm{T}} \cdot \mathbf{B}\mathbf{B}^{\mathrm{T}} \cdot \beta^{(j)}\right| \leqslant \sum_{l=1}^{k} \left|\left\langle \beta^{(i)}, \beta^{(l)} \right\rangle\right| \left|\left\langle \beta^{(l)}, \beta^{(j)} \right\rangle\right| \leqslant 8\left(1 + \frac{3k}{\Psi}\right)\sqrt{\frac{k}{\Psi}} + 16\frac{k^2}{\Psi} \leqslant 11\sqrt{\frac{k}{\Psi}}.$$

[4] We denote by SPSD the class of symmetric positive semi-definite matrices.

By Lemma 5.3 every vector $x \in \mathbb{R}^k$ can be expressed as $x = \sum_{i=1}^k \gamma_i \beta^{(i)}$.

$$\begin{aligned} x^{\mathrm{T}} \mathbf{B}\mathbf{B}^{\mathrm{T}} x &= \sum_{i=1}^k \gamma_i \left(\beta^{(i)}\right)^{\mathrm{T}} \cdot \mathbf{B}\mathbf{B}^{\mathrm{T}} \cdot \sum_{j=1}^k \gamma_j \beta^{(j)} \\ &= \sum_{i=1}^k \gamma_i^2 \left(\beta^{(i)}\right)^{\mathrm{T}} \cdot \mathbf{B}\mathbf{B}^{\mathrm{T}} \cdot \beta^{(i)} + \sum_{i=1}^k \sum_{j \neq i}^k \gamma_i \gamma_j \left(\beta^{(i)}\right)^{\mathrm{T}} \cdot \mathbf{B}\mathbf{B}^{\mathrm{T}} \cdot \beta^{(j)} \\ &\geqslant \left(1 - \frac{23k^2}{\Psi} - 11k\sqrt{\frac{k}{\Psi}}\right) \|\gamma\|^2 \geqslant \left(1 - 14k\sqrt{\frac{k}{\Psi}}\right) \|\gamma\|^2 . \end{aligned}$$

and

$$x^{\mathrm{T}} x = \sum_{i=1}^k \sum_{j=1}^k \gamma_i \gamma_j \left\langle \beta^{(i)}, \beta^{(j)} \right\rangle = \sum_{i=1}^k \gamma_i^2 \left\| \beta^{(i)} \right\|^2 + \sum_{i=1}^k \sum_{j \neq i}^k \gamma_i \gamma_j \left\langle \beta^{(i)}, \beta^{(j)} \right\rangle$$

By Lemma 5.2 we have $\left| \sum_{i=1}^k \sum_{j \neq i}^k \gamma_i \gamma_j \left\langle \beta^{(i)}, \beta^{(j)} \right\rangle \right| \leqslant \|\gamma\|^2 \cdot 4k\sqrt{\frac{k}{\Psi}}$ and $\left\| \beta^{(i)} \right\|^2 \leqslant 1 + \frac{3k}{\Psi}$. Thus it holds

$$\left(1 - 5k\sqrt{\frac{k}{\Psi}}\right) \|\gamma\|^2 \leqslant x^{\mathrm{T}} x \leqslant \left(1 + 5k\sqrt{\frac{k}{\Psi}}\right) \|\gamma\|^2 .$$

Therefore

$$1 - 20k\sqrt{\frac{k}{\Psi}} \leqslant \lambda(\mathbf{B}\mathbf{B}^{\mathrm{T}}) \leqslant 1 + 20k\sqrt{\frac{k}{\Psi}}.$$

■

We have now established the first part of Theorem 5.1. We turn to the second part and restate it in the following Lemma.

Lemma 5.5. *(Rows) If $\Psi \geqslant 10^4 \cdot k^3/\varepsilon^2$ and $\varepsilon \in (0,1)$ then for all distinct $i, j \in [1:k]$ it holds*

$$|\langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:} \rangle| \leqslant \sqrt{\varepsilon}.$$

To prove Lemma 5.5 we establish the following three Lemmas. Before stating them we need some notation that is inspired by Lemma 5.2.

Definition 5.6. *Let $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I} + \mathbf{E}$, where $|\mathbf{E}_{ij}| \leqslant 4\sqrt{\frac{k}{\Psi}}$ and $\mathbf{E}$ is symmetric matrix. Then we have*

$$\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)^2 = \mathbf{B}\left(\mathbf{I} + \mathbf{E}\right)\mathbf{B}^{\mathrm{T}} = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}.$$

Lemma 5.7. *If $\Psi \geqslant 40^2 \cdot k^3/\varepsilon^2$ and $\varepsilon \in (0,1)$ then all eigenvalues of matrix $\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}$ satisfy*

$$\left|\lambda(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}})\right| \leqslant \varepsilon/5.$$

Proof. Let $z = \mathbf{B}^{\mathrm{T}} x$. We upper bound the quadratic form

$$\left|x^{\mathrm{T}} \mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}} x\right| = \left|z^{\mathrm{T}} \mathbf{E} z\right| \leqslant \sum_{ij} |\mathbf{E}_{ij}| \, |z_i| \, |z_j| \leqslant 4\sqrt{\frac{k}{\Psi}} \cdot \left(\sum_{i=1}^k |z_i|\right)^2 \leqslant \|z\|^2 \cdot 4k\sqrt{\frac{k}{\Psi}}.$$

By Lemma 5.4 we have $1-\varepsilon \leqslant \lambda(\mathbf{B}\mathbf{B}^{\mathrm{T}}) \leqslant 1+\varepsilon$ and since $\|z\|^2 = \frac{x\mathbf{B}\mathbf{B}^{\mathrm{T}}x}{x^{\mathrm{T}}x} \cdot \|x\|^2$ it follows that

$$\frac{\|z\|^2}{1+\varepsilon} \leqslant \|x\|^2 \leqslant \frac{\|z\|^2}{1-\varepsilon},$$

and hence

$$\left|\lambda(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}})\right| \leqslant \max_x \frac{\left|x^{\mathrm{T}}\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}x\right|}{x^{\mathrm{T}}x} \leqslant 4\,(1+\varepsilon) \cdot k\sqrt{\frac{k}{\Psi}} \leqslant \varepsilon/5.$$

■

Lemma 5.8. *Suppose $\{u_i\}_{i=1}^k$ is orthonormal basis and the square matrix $\mathbf{U}$ has u_i as its i-th column. Then $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$.*

Proof. Notice that by the definition of $\mathbf{U}$ it holds $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$. Moreover, the matrix $\mathbf{U}^{-1}$ exists and thus $\mathbf{U}^{\mathrm{T}} = \mathbf{U}^{-1}$. Therefore, we have $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$ as claimed. ■

Lemma 5.9. *If $\Psi \geqslant 40^2 \cdot k^3/\varepsilon^2$ and $\varepsilon \in (0,1)$ then it holds $|(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}})_{ij}| \leqslant \varepsilon/5$ for every $i,j \in [1:k]$.*

Proof. Notice that $\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}$ is symmetric matrix, since E is symmetric. By SVD Theorem there is an orthonormal basis $\{u_i\}_{i=1}^k$ such that $\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}} = \sum_{i=1}^k \lambda_i(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}) \cdot u_i u_i^{\mathrm{T}}$. Thus, it suffices to bound the expression

$$|(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}})_{ij}| \leqslant \sum_{l=1}^{k} |\lambda_l(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}})| \cdot |(u_l u_l^{\mathrm{T}})_{ij}|.$$

By Lemma 5.8 we have

$$\sum_{l=1}^{k} |(u_l)_i| \cdot |(u_l)_j| \leqslant \sqrt{\|\mathbf{U}_{i,:}\|^2}\sqrt{\|\mathbf{U}_{j,:}\|^2} = 1.$$

We apply now Lemma 5.7 to obtain

$$\sum_{l=1}^{k} |\lambda_l(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}})| \cdot |(u_l u_l^{\mathrm{T}})_{ij}| \leqslant \frac{\varepsilon}{5} \cdot \sum_{l=1}^{k} |(u_l)_i| \cdot |(u_l)_j| \leqslant \frac{\varepsilon}{5}.$$

■

We are ready now to prove Lemma 5.5, i.e. $|\langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:}\rangle| \leqslant \sqrt{\varepsilon}$ for all $i \neq j$.

Proof of Lemma 5.5. By Definition 5.6 we have $\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)^2 = \mathbf{B}\mathbf{B}^{\mathrm{T}} + \mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}$. Observe that the (i,j)-th entry of matrix $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ is equal to the inner product between the i-th and j-th row of matrix $\mathbf{B}$, i.e. $\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)_{ij} = \langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:}\rangle$. Moreover, we have

$$\left[\left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)^2\right]_{ij} \;=\; \sum_{l=1}^{k} \left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)_{i,l} \left(\mathbf{B}\mathbf{B}^{\mathrm{T}}\right)_{l,j} = \sum_{l=1}^{k} \langle \mathbf{B}_{i,:}, \mathbf{B}_{l,:}\rangle \langle \mathbf{B}_{l,:}, \mathbf{B}_{j,:}\rangle\,.$$

For the entries on the main diagonal, it holds

$$\langle \mathbf{B}_{i,:}, \mathbf{B}_{i,:}\rangle^2 + \sum_{l\neq i}^{k} \langle \mathbf{B}_{i,:}, \mathbf{B}_{l,:}\rangle^2 = [(\mathbf{B}\mathbf{B}^{\mathrm{T}})^2]_{ii} = [\mathbf{B}\mathbf{B}^{\mathrm{T}} + \mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}]_{ii} = \langle \mathbf{B}_{i,:}, \mathbf{B}_{i,:}\rangle + \left(\mathbf{B}\mathbf{E}\mathbf{B}^{\mathrm{T}}\right)_{ii},$$

and hence by applying Lemma 5.4 with $\varepsilon' = \varepsilon/5$ and Lemma 5.9 with $\varepsilon' = \varepsilon$ we obtain

$$\langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:}\rangle^2 \leqslant \sum_{l\neq i} \langle \mathbf{B}_{i,:}, \mathbf{B}_{l,:}\rangle^2 \leqslant \left(1+\frac{\varepsilon}{5}\right) + \frac{\varepsilon}{5} - \left(1-\frac{\varepsilon}{5}\right)^2 \leqslant \varepsilon.$$

■

6 Vectors $p^{(i)}$ are Well-Spread

Peng et al. (c.f. [6, Lemma 4.2]) showed for $\Upsilon \geqslant \Omega(k^3)$ that the square Euclidean distance between any distinct estimation center vectors (c.f. Equation 8) is lower bounded by

$$\left\|p^{(i)} - p^{(j)}\right\|^2 \geqslant \left[10^3 \cdot k \cdot \min\left\{\mu(P_i), \mu(P_j)\right\}\right]^{-1}.$$

Under a less restrictive gap assumption $\Psi \geqslant \Omega(k^3)$ we improve [6, Lemma 4.2] by a factor of k. Our analysis builds upon Theorem 5.1 and bounds a summation of k terms, instead of applying [6, Lemma 4.2] to a single component. We show now a statement similar to [6, Lemma B.1] that depends on Ψ.

Lemma 6.1. *If $\Psi = 20^4 \cdot k^3/\delta$ for some $\delta \in (0,1]$ then for every $i \in [1:k]$ it holds*

$$\left\|p^{(i)}\right\|^2 \in \frac{1}{\mu(P_i)}\left[1 \pm \frac{\sqrt{\delta}}{4}\right].$$

Proof. By definition $p_i = \frac{1}{\sqrt{\mu(P_i)}} \cdot \mathbf{B}_{i,:}$ and by Theorem 5.1 we have $\|\mathbf{B}_{i,:}\|^2 \in [1 \pm \sqrt{\delta}/4]$. ∎

We present now our statement.

Lemma 6.2. *If $\Psi = 20^4 \cdot k^3/\delta$ for some $\delta \in (0,1/2]$ then for any distinct $i,j \in [1:k]$ it holds that*

$$\left\|p^{(i)} - p^{(j)}\right\|^2 \geqslant \left[2 \cdot \min\left\{\mu(P_i), \mu(P_j)\right\}\right]^{-1}.$$

Suppose c_i is the center of a cluster A_i. If $\left\|c_i - p^{(i_1)}\right\| \geqslant \left\|c_i - p^{(i_2)}\right\|$ then it holds

$$\left\|c_i - p^{(i_1)}\right\|^2 \geqslant \frac{1}{4}\left\|p^{(i_1)} - p^{(i_2)}\right\|^2 \geqslant \left[8 \cdot \min\left\{\mu(P_{i_1}), \mu(P_{i_2})\right\}\right]^{-1}.$$

Proof. We argue in a similar manner as in [6] but in contrast apply Theorem 5.1 with $\varepsilon = \sqrt{\delta}/4$ to obtain

$$\left\langle \frac{p^{(i)}}{\left\|p^{(i)}\right\|}, \frac{p^{(j)}}{\left\|p^{(j)}\right\|} \right\rangle = \frac{\langle \mathbf{B}_{i,:}, \mathbf{B}_{j,:}\rangle}{\|\mathbf{B}_{i,:}\|\,\|\mathbf{B}_{j,:}\|} \leqslant \frac{\sqrt{\varepsilon}}{1-\varepsilon} = \frac{2\delta^{1/4}}{3}.$$

W.l.o.g. assume that $\left\|p^{(i)}\right\|^2 \geqslant \left\|p^{(j)}\right\|^2$. Then by Lemma 6.1 we have

$$\left\|p^{(i)}\right\|^2 \geqslant \left(1 - \frac{\sqrt{\delta}}{4}\right) \cdot \left[\min\left\{\mu(P_i), \mu(P_j)\right\}\right]^{-1}.$$

Let $\left\|p^{(j)}\right\| = \alpha \cdot \left\|p^{(i)}\right\|$ for some $\alpha \in (0,1]$. Then

$$\begin{aligned}
\left\|p^{(i)} - p^{(j)}\right\|^2 &= \left\|p^{(i)}\right\|^2 + \left\|p^{(j)}\right\|^2 - 2\left\langle \frac{p^{(i)}}{\left\|p^{(i)}\right\|}, \frac{p^{(j)}}{\left\|p^{(j)}\right\|} \right\rangle \left\|p^{(i)}\right\| \left\|p^{(j)}\right\| \\
&\geqslant \left(\alpha^2 - \frac{4\delta^{1/4}}{3} \cdot \alpha + 1\right) \left\|p^{(i)}\right\|^2 \geqslant \left[2 \cdot \min\left\{\mu(P_i), \mu(P_j)\right\}\right]^{-1}.
\end{aligned}$$

The second claim follows immediately from the first. ∎

7 Proof of Lemma 1.3

By Theorem 4.1 we have $\|f_i - \widehat{g}_i\|^2 \leqslant \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{k}{\Psi}$ and thus

$$\sum_{i=1}^{k} \sum_{u \in P_i} d_u \left\|F(u) - c_i^\star\right\|^2 \leqslant \sum_{i=1}^{k} \sum_{u \in P_i} d_u \left\|F(u) - p^{(i)}\right\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{u \in P_i} d_u \left(F(u)_j - p_j^{(i)}\right)^2$$
$$= \sum_{j=1}^{k} \sum_{i=1}^{k} \sum_{u \in P_i} \left(f_j(u) - \widehat{g}_j(u)\right)^2 = \sum_{j=1}^{k} \|f_j - \widehat{g}_j\|^2 \leqslant \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{k^2}{\Psi},$$

where the k-way partition $(P_1, \ldots, P_k)$ achieving $\widehat{\rho}_{\mathrm{avr}}(k)$ has corresponding centers $c_1^\star, \ldots, c_k^\star$.

8 Proof of Lemma 1.4

Our main result in this section improves [6, Lemma 4.4] by a factor of k. We argue in a similar manner as in [6], but in contrast our result relies on Lemma 6.2 and the gap parameter Ψ.

We begin our discussion by restating [6, Lemma B.2] whose analysis crucially relies on a function σ defined by

$$\sigma(l) = \arg \max_{j \in [1:k]} \frac{\mu(A_l \cap P_j)}{\mu(P_j)}. \tag{14}$$

Lemma 8.1. *[6, Lemma B.2] Let $(P_1, \ldots, P_k)$ and $(A_1, \ldots, A_k)$ be partitions of the vector set. Suppose for every permutation $\pi : [1:k] \to [1:k]$ there is an index $i \in [1:k]$ such that*

$$\mu(A_i \triangle P_{\pi(i)}) \geqslant 2\varepsilon \cdot \mu(P_{\pi(i)}), \tag{15}$$

where $\varepsilon \in (0, 1/2)$ is a parameter. Then one of the following three statements holds:
1. If σ is a permutation and $\mu(P_{\sigma(i)} \backslash A_i) \geqslant \varepsilon \cdot \mu(P_{\sigma(i)})$, then for every index $j \neq i$ there is a real $\varepsilon_j \geqslant 0$ such that

$$\mu(A_j \cap P_{\sigma(j)}) \geqslant \mu(A_j \cap P_{\sigma(i)}) \geqslant \varepsilon_j \cdot \min\{\mu(P_{\sigma(j)}), \mu(P_{\sigma(i)})\},$$

and $\sum_{j \neq i} \varepsilon_j \geqslant \varepsilon$.
2. If σ is a permutation and $\mu(A_i \backslash P_{\sigma(i)}) \geqslant \varepsilon \cdot \mu(P_{\sigma(i)})$, then for every $j \neq i$ there is a real $\varepsilon_j \geqslant 0$ such that

$$\mu(A_i \cap P_{\sigma(i)}) \geqslant \varepsilon_j \cdot \mu(P_{\sigma(i)}), \quad \mu(A_i \cap P_{\sigma(j)}) \geqslant \varepsilon_j \cdot \mu(P_{\sigma(i)}),$$

and $\sum_{j \neq i} \varepsilon_j \geqslant \varepsilon$.
3. If σ is not a permutation, then there is an index $\ell \notin \{\sigma(1), \ldots, \sigma(k)\}$ and for every index j there is a real $\varepsilon_j \geqslant 0$ such that

$$\mu(A_j \cap P_{\sigma(j)}) \geqslant \mu(A_j \cap P_\ell) \geqslant \varepsilon_j \cdot \min\{\mu(P_{\sigma(j)}), \mu(P_\ell)\},$$

and $\sum_{j=1}^{k} \varepsilon_j = 1$.

We prove now our main technical result that yields an improved lower bound by a factor of k.

Lemma 8.2. *Suppose the hypothesis of Lemma 8.1 is satisfied and $\Psi = 20^4 \cdot k^3/\delta$ for some $\delta \in (0, 1/2]$. Then it holds*

$$\mathrm{Cost}(\{A_i, c_i\}_{i=1}^{k}) \geqslant \frac{\varepsilon}{16} - \frac{2k^2}{\Psi}.$$

Proof. By definition

$$\mathrm{Cost}(\{A_i, c_i\}_{i=1}^{k}) = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{u \in A_i \cap P_j} d_u \left\| F(u) - c_i \right\|^2 \triangleq \Lambda. \tag{16}$$

Since for every vectors $x, y, z \in \mathbb{R}^k$ it holds

$$2\left(\|x-y\|^2 + \|z-y\|^2\right) \geqslant (\|x-y\| + \|z-y\|)^2 \geqslant \|x-z\|^2,$$

we have for all indices $i, j \in [1:k]$ that

$$\|F(u) - c_i\|^2 \geqslant \frac{\left\|p^{(j)} - c_i\right\|^2}{2} - \left\|F(u) - p^{(j)}\right\|^2. \tag{17}$$

Our proof proceeds by considering three cases. Let $i \in [1:k]$ be the index from the hypothesis in Lemma 8.1.

Case 1. Suppose the first conclusion of Lemma 8.1 holds. For every index $j \neq i$ let

$$p^{\gamma(j)} = \begin{cases} p^{\sigma(j)} & \text{, if } \left\|p^{\sigma(j)} - c_j\right\| \geqslant \left\|p^{\sigma(i)} - c_j\right\|; \\ p^{\sigma(i)} & \text{, otherwise.} \end{cases}$$

Then by combining (17), Lemma 6.2 and Lemma 1.3, we have

$$\begin{aligned} \Lambda &\geqslant \frac{1}{2} \sum_{j \neq i} \sum_{u \in A_j \cap P_{\gamma(j)}} d_u \left\|p^{\gamma(j)} - c_j\right\|^2 - \sum_{j \neq i} \sum_{u \in A_j \cap P_{\gamma(j)}} \left\|F(u) - p^{\gamma(j)}\right\|^2 \\ &\geqslant \frac{1}{16} \sum_{j \neq i} \frac{\mu(A_j \cap P_{\gamma(j)})}{\min\{\mu(P_{\sigma(i)}), \mu(P_{\sigma(j)})\}} - \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{k^2}{\Psi} \geqslant \frac{\varepsilon}{16} - \frac{2k^2}{\Psi}. \end{aligned}$$

Case 2. Suppose the second conclusion of Lemma 8.1 holds. Notice that if $\mu(A_i \cap P_{\sigma(i)}) \leqslant (1-\varepsilon) \cdot \mu(P_{\sigma(i)})$ then $\mu(P_{\sigma(i)} \backslash A_i) \geqslant \varepsilon \cdot \mu(P_{\sigma(i)})$ and thus we can argue as in Case 1. Hence, we can assume that it holds

$$\mu(A_i \cap P_{\sigma(i)}) \geqslant (1-\varepsilon) \cdot \mu(P_{\sigma(i)}). \tag{18}$$

We proceed by analyzing two subcases.

a) If $\left\|p^{\sigma(j)} - c_i\right\| \geqslant \left\|p^{\sigma(i)} - c_i\right\|$ holds for all $j \neq i$ then by combining (17), Lemma 6.2 and Lemma 1.3 it follows

$$\begin{aligned} \Lambda &\geqslant \frac{1}{2} \sum_{j \neq i} \sum_{u \in A_i \cap P_{\sigma(j)}} d_u \left\|p^{\sigma(j)} - c_i\right\|^2 - \sum_{j \neq i} \sum_{u \in A_i \cap P_{\sigma(j)}} \left\|F(u) - p^{\sigma(j)}\right\|^2 \\ &\geqslant \frac{1}{2} \sum_{j \neq i} \frac{\mu(A_i \cap P_{\sigma(j)})}{\min\{\mu(P_{\sigma(i)}), \mu(P_{\sigma(j)})\}} - \left(1 + \frac{3k}{\Psi}\right) \cdot \frac{k^2}{\Psi} \geqslant \frac{\varepsilon}{16} - \frac{2k^2}{\Psi}. \end{aligned}$$

b) Suppose there is an index $j \neq i$ such that $\left\|p^{\sigma(j)} - c_i\right\| < \left\|p^{\sigma(i)} - c_i\right\|$. Then by triangle inequality combined with Lemma 6.2 we have

$$\left\|p^{\sigma(i)} - c_i\right\|^2 \geqslant \frac{1}{4} \left\|p^{\sigma(i)} - p^{\sigma(j)}\right\| \geqslant \left[8 \cdot \min\{\mu(P_{\sigma(i)}), \mu(P_{\sigma(j)})\}\right]^{-1}.$$

Thus, by combining (17), (18) and Lemma 1.3 we obtain

$$\begin{aligned}\Lambda &\geqslant \frac{1}{2}\sum_{u\in A_i\cap P_{\sigma(i)}} d_u \left\|p^{\sigma(i)}-c_i\right\|^2 - \sum_{u\in A_i\cap P_{\sigma(i)}} d_u \left\|F(u)-p^{\sigma(i)}\right\|^2 \\ &\geqslant \frac{1}{16}\cdot\frac{\mu(A_i\cap P_{\sigma(i)})}{\min\{\mu(P_{\sigma(i)}),\mu(P_{\sigma(j)})\}} - \left(1+\frac{3k}{\Psi}\right)\cdot\frac{k^2}{\Psi} \geqslant \frac{1-\varepsilon}{16}-\frac{2k^2}{\Psi}.\end{aligned}$$

Case 3. Suppose the third conclusion of Lemma 8.1 holds, i.e., σ is not a permutation. Then there is an index $\ell\in[1:k]\setminus\{\sigma(1),\ldots,\sigma(k)\}$ and for every index $j\in[1:k]$ let

$$p^{\gamma(j)} = \begin{cases} p^{\ell} & \text{, if } \left\|p^{\ell}-c_j\right\| \geqslant \left\|p^{\sigma(j)}-c_j\right\|; \\ p^{\sigma(j)} & \text{, otherwise.}\end{cases}$$

By combining (17), Lemma 6.2 and Lemma 1.3 it follows that

$$\begin{aligned}\Lambda &\geqslant \frac{1}{2}\sum_{j=1}^{k}\sum_{u\in A_j\cap P_{\gamma(j)}} d_u \left\|p^{\gamma(j)}-c_j\right\|^2 - \sum_{j=1}^{k}\sum_{u\in A_j\cap P_{\gamma(j)}} d_u \left\|F(u)-p^{\gamma(j)}\right\|^2 \\ &\geqslant \frac{1}{16}\sum_{j=1}^{k}\frac{\mu(A_j\cap P_{\gamma(j)})}{\min\{\mu(P_{\sigma(j)}),\mu(P_{\ell})\}} - \left(1+\frac{3k}{\Psi}\right)\cdot\frac{k^2}{\Psi} \geqslant \frac{1}{16}-\frac{2k^2}{\Psi}.\end{aligned}$$

■

Based on Lemma 8.2 we improve [6, Lemma 4.4] by a factor of k and condition our analysis on a less restrictive gap assumption that depends on Ψ.

Corollary 8.3. *Let $(P_1,\ldots,P_k)$ and $(A_1,\ldots,A_k)$ are partitions of the vector set. Suppose for every permutation $\pi:[1:k]\to[1:k]$ there is an index $i\in[1:k]$ such that*

$$\mu(A_i\triangle P_{\pi(i)}) \geqslant \frac{2\varepsilon}{k}\cdot\mu(P_{\pi(i)}), \tag{19}$$

where $\varepsilon\in(0,1)$ is a parameter. If $\Psi=20^4\cdot k^3/\delta$ for some $\delta\in(0,1/2]$, and $\varepsilon\geqslant 64\cdot\alpha\cdot k^3/\Psi$ then

$$\mathrm{Cost}(\{A_i,c_i\}_{i=1}^{k}) > \frac{2k^2}{\Psi}\alpha.$$

Proof. We apply Lemma 8.1 with $\varepsilon'=\varepsilon/k$. Then by Lemma 8.2 we have

$$\mathrm{Cost}(\{A_i,c_i\}_{i=1}^{k}) \geqslant \frac{\varepsilon}{16k}-\frac{2k^2}{\Psi},$$

and the desired result follows by setting $\varepsilon\geqslant 64\cdot\alpha\cdot k^3/\Psi$. ■

We note that Lemma 1.4 follows directly by applying Corollary 8.3 with $\varepsilon=64\cdot\alpha\cdot k^3/\Psi$.

9 The Normalized Spectral Embedding is ε-separated

In this section, we prove that the normalized spectral embedding $\mathcal{X}_V$ is ε-separated.

Theorem 9.1. *Let G be a graph that satisfies $\Psi=20^4\cdot k^3/\delta$, $\delta\in(0,1/2]$ and $k/\delta\geqslant 10^9$. Then for $\varepsilon=6\cdot 10^{-7}$ it holds*

$$\triangle_k(\mathcal{X}_V) \leqslant \varepsilon^2\triangle_{k-1}(\mathcal{X}_V). \tag{20}$$

Proof of Theorem 9.1 We establish first a lower bound on $\triangle_{k-1}(\mathcal{X}_V)$.

Lemma 9.2. *Let G be a graph that satisfies $\Psi = 20^4 \cdot k^3/\delta$ for some $\delta \in (0, 1/2]$. Then for $\delta' = 2\delta/20^4$ it holds*

$$\triangle_{k-1}(\mathcal{X}_V) \geqslant \frac{1}{12} - \frac{\delta'}{k}. \tag{21}$$

Before we present the proof of Lemma 9.2 we show that it implies (20). By Lemma 1.3 we have

$$\triangle_k(\mathcal{X}_V) \leqslant \frac{2k^2}{\Psi} = \frac{\delta'}{k}.$$

Moreover, by applying Lemma 9.2 with $k/\delta \geqslant 10^9$ and $\varepsilon = 6 \cdot 10^{-7}$ we obtain

$$\triangle_{k-1}(\mathcal{X}_V) \geqslant \frac{1}{12} - \frac{\delta'}{k} = \frac{1}{12} - \frac{2}{20^4} \cdot \frac{\delta}{k} \geqslant \frac{10^{10}}{9 \cdot 2^5} \cdot \frac{\delta}{k} = \frac{1}{\varepsilon^2} \cdot \frac{\delta'}{k} \geqslant \frac{1}{\varepsilon^2} \cdot \triangle_k(\mathcal{X}_V).$$

Proof of Lemma 9.2 We argue in a similar manner as in Lemma 8.2 (c.f. Case 3). We start by giving some notations. Then we prove Lemma 9.3 which is later used in the proof of Lemma 9.2.

We redefine the function σ (c.f. Equation 14) such that for any two partitions $(P_1, \ldots, P_k)$ and $(Z_1, \ldots, Z_{k-1})$ of V, we define a function $\sigma : [1 : k-1] \mapsto [1 : k]$ by

$$\sigma(i) = \arg \max_{j \in [1:k]} \frac{\mu(Z_i \cap P_j)}{\mu(P_j)}, \quad \text{for every } i \in [1 : k-1].$$

The next statement is similar to the third conclusion of Lemma 8.1, but in contrast lower bounds the overlapping (in terms of the volume) between any k-way and $(k-1)$-way partitions of V.

Lemma 9.3. *Suppose $(P_1, \ldots, P_k)$ and $(Z_1, \ldots, Z_{k-1})$ are partitions of V. Then for any index $\ell \in [1 : k] \setminus \{\sigma(1), \ldots, \sigma(k-1)\}$ (there is at least one such ℓ) and for every $i \in [1 : k-1]$ it holds*

$$\left\{\mu(Z_i \cap P_{\sigma(i)}), \mu(Z_i \cap P_\ell)\right\} \geqslant \tau_i \cdot \min\left\{\mu(P_\ell), \mu(P_{\sigma(i)})\right\},$$

where $\sum_{i=1}^{k-1} \tau_i = 1$ and $\tau_i \geqslant 0$.

Proof. By pigeonhole principle there is an index $\ell \in [1 : k]$ such that $\ell \notin \{\sigma(1), \ldots, \sigma(k-1)\}$. Thus, for every $i \in [1 : k-1]$ we have $\sigma(i) \neq \ell$ and

$$\frac{\mu(Z_i \cap P_{\sigma(i)})}{\mu(P_{\sigma(i)})} \geqslant \frac{\mu(Z_i \cap P_\ell)}{\mu(P_\ell)} \triangleq \tau_i,$$

where $\sum_{i=1}^{k-1} \tau_i = 1$ and $\tau_i \geqslant 0$ for all i. Hence, the statement follows. ∎

We present now the proof of Lemma 9.2.

Proof of Lemma 9.2. Let $(Z_1, \ldots, Z_{k-1})$ be a $(k-1)$-way partition of V with centers $c'_1, \ldots, c'_{k-1}$ that achieves $\triangle_{k-1}(\mathcal{X}_V)$, and $(P_1, \ldots, P_k)$ be a k-way partition of V achieving $\widehat{\rho}_{\mathrm{avr}}(k)$. Our goal now is to lower bound the optimal $(k-1)$-means cost

$$\triangle_{k-1}(\mathcal{X}_V) = \sum_{i=1}^{k-1} \sum_{j=1}^{k} \sum_{u \in Z_i \cap P_j} d_u \left\|F(u) - c'_i\right\|^2. \tag{22}$$

By Lemma 9.3 there is an index $\ell \in [1:k] \setminus \{\sigma(1),\ldots,\sigma(k-1)\}$. For $i \in [1:k-1]$ let

$$p^{\gamma(i)} = \begin{cases} p^{\ell} & \text{, if } \left\|p^{\ell} - c_i'\right\| \geqslant \left\|p^{\sigma(i)} - c_i'\right\|; \\ p^{\sigma(i)} & \text{, otherwise.} \end{cases}$$

Then by combining Lemma 6.2 and Lemma 9.3, we have

$$\left\|p^{\gamma(i)} - c_i'\right\|^2 \geqslant \left[8 \cdot \min\left\{\mu(P_\ell), \mu(P_{\sigma(i)})\right\}\right]^{-1} \text{ and } \mu(Z_i \cap P_{\gamma(i)}) \geqslant \tau_i \cdot \min\left\{\mu(P_\ell), \mu(P_{\sigma(i)})\right\}, \quad (23)$$

where $\sum_{i=1}^{k-1} \tau_i = 1$. We now lower bound the expression in (22). Since

$$\left\|F(u) - c_i'\right\|^2 \geqslant \frac{1}{2}\left\|p^{\gamma(i)} - c_i'\right\|^2 - \left\|F(u) - p^{\gamma(i)}\right\|^2,$$

it follows for $\delta' = 2\delta/20^4$ that

$$\begin{aligned}
\triangle_{k-1}(\mathcal{X}_V) &= \sum_{i=1}^{k-1}\sum_{j=1}^{k}\sum_{u\in Z_i\cap P_j} d_u \left\|F(u) - c_i'\right\|^2 \geqslant \sum_{i=1}^{k-1}\sum_{u\in Z_i\cap P_{\gamma(i)}} d_u \left\|F(u) - c_i'\right\|^2 \\
&\geqslant \frac{1}{2}\sum_{i=1}^{k-1}\sum_{u\in Z_i\cap P_{\gamma(i)}} d_u \left\|p^{\gamma(i)} - c_i'\right\|^2 - \sum_{i=1}^{k-1}\sum_{u\in Z_i\cap P_{\gamma(i)}} d_u \left\|F(u) - p^{\gamma(i)}\right\|^2 \\
&\geqslant \frac{1}{2}\sum_{i=1}^{k-1} \frac{\mu(Z_i \cap P_{\gamma(i)})}{8 \cdot \min\left\{\mu(P_{\gamma(i)}), \mu(P_{\sigma(i)})\right\}} - \sum_{i=1}^{k}\sum_{u\in P_i} d_u \left\|F(u) - p^i\right\|^2 \\
&\geqslant \frac{1}{16} - \frac{\delta'}{k},
\end{aligned}$$

where the last inequality holds due to (23) and Lemma 1.3. ∎

10 An Efficient Spectral Clustering Algorithm

In this section, we prove Part (b) of Theorem 1.2. We start by stating in Subsection 10.1 the notations used in our proof. Then we describe the proof-overview of our approach. The proof itself is divided into three parts, each of which is covered in Subsection 10.2, 10.3 and 10.4, respectively.

10.1 Notations

Let $Z \in \mathbb{R}^{n\times k}$ be a matrix whose rows represent n vectors that are to be partitioned into k clusters. For every k-way partition we associate an indicator matrix $X \in \mathbb{R}^{n\times k}$ that satisfies $X_{ij} = 1/\sqrt{|C_j|}$ if the i-th row $Z_{i,:}$ belongs to the j-th cluster C_j, and $X_{ij} = 0$ otherwise. We denote the optimal indicator matrix X_{opt} by

$$X_{\text{opt}} = \arg\min_{X\in\mathbb{R}^{n\times k}} \left\|Z - XX^{\mathrm{T}}Z\right\|_F^2 = \arg\min_{X\in\mathbb{R}^{n\times k}} \sum_{j=1}^{k}\sum_{u\in X_j} \left\|Z_{u,:} - c_j\right\|_2^2, \quad (24)$$

where $c_j = (1/|X_j|)\sum_{u\in X_j} Z_{u,:}$ is the center point of cluster C_j.

Let $\mathcal{L}_G = I - \mathcal{A}_N$ be the normalized Laplacian matrix of a graph G and $\mathcal{A}_N = D^{1/2}AD^{-1/2}$ be the corresponding normalized adjacency matrix. Let matrix U_k be composed of the top k

orthonormal eigenvectors of $\mathcal{A}_N$ or equivalently the bottom eigenvectors of $\mathcal{L}_G$. We define by $Y \triangleq U_k$ the canonical spectral embedding.

We describe now the "Power Method" that is used to compute an approximate spectral embedding. Let $S \in \mathbb{R}^{n\times k}$ be a matrix whose entries are i.i.d. samples from the standard Gaussian distribution $N(0,1)$ and p be a positive integer. Then the approximate spectral embedding $\widetilde{Y}$ is defined by the following two-step process:

$$1) \quad B \triangleq \mathcal{A}_N^{2p+1} \cdot S = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^{\mathrm{T}}; \quad \text{and} \quad 2) \quad \widetilde{Y} \triangleq \widetilde{U} \in \mathbb{R}^{n\times k}. \tag{25}$$

We proceed by defining the normalized (approximate) spectral embedding. We construct a matrix $Y' \in \mathbb{R}^{m\times k}$ such that for every vertex $u \in V$ we add $d(u)$ many copies of the normalized row $U_k(u,:)/\sqrt{d(u)}$ to Y'. Formally, the normalized (approximate) spectral embedding Y' ($\widetilde{Y'}$) is defined by

$$Y' = \begin{pmatrix} \mathbf{1}_{d(1)} \frac{U_k(1,:)}{\sqrt{d(1)}} \\ \cdots \\ \mathbf{1}_{d(n)} \frac{U_k(n,:)}{\sqrt{d(n)}} \end{pmatrix}_{m\times k} \quad \text{and} \quad \widetilde{Y'} = \begin{pmatrix} \mathbf{1}_{d(1)} \frac{\widetilde{U}(1,:)}{\sqrt{d(1)}} \\ \cdots \\ \mathbf{1}_{d(n)} \frac{\widetilde{U}(n,:)}{\sqrt{d(n)}} \end{pmatrix}_{m\times k}, \tag{26}$$

where $\mathbf{1}_{d(i)}$ is all-one column vector with dimension $d(i)$.

Similarly to (24) we associate to Y' ($\widetilde{Y'}$) an indicator matrix X' ($\widetilde{X'}$) that satisfies $X'_{ij} = 1/\sqrt{\mu(C_j)}$ if the i-th row $Y'_{i,:}$ belongs to the j-th cluster C_j, and $X'_{ij} = 0$ otherwise. We may assume w.l.o.g. that a k-means algorithm outputs an indicator matrix X' such that all copies of row $U_k(v,:)/\sqrt{d(v)}$ belong to the same cluster, for every vertex $v \in V$.

We associate to matrices Y' and $\widetilde{Y'}$ sets of points which we denote by $\mathcal{X}_V$ and $\widetilde{\mathcal{X}_V}$, respectively. We present now a key connection between the spectral embedding map $F(\cdot)$, the optimal k-means cost $\triangle_k(\mathcal{X}_V)$ and matrices Y', X'_{opt}:

$$\left\| Y' - X'_{\mathrm{opt}} \left(X'_{\mathrm{opt}}\right)^{\mathrm{T}} Y' \right\|_F^2 = \sum_{j=1}^{k} \sum_{v\in C_j^\star} d(v) \left\| F(v) - c_j^\star \right\|_F^2 = \triangle_k(\mathcal{X}_V), \tag{27}$$

where each center satisfies $c_j^\star = \mu(C_j^\star)^{-1} \cdot \sum_{v\in C_j^\star} d(v)F(v)$ and $F(v) = Y_{v,:}/\sqrt{d(v)}$.

Proof Overview of Theorem 1.2 Building upon the work of Boutsidis et al [1] we prove that any α-approximate k-means algorithm that runs on an approximate normalized spectral embedding $\widetilde{Y'}$ computed by the "power method", yields an approximate clustering $\widetilde{X'_\alpha}$ of the normalized spectral embedding Y'. Under our gap assumption, we prove that $\widetilde{Y'}$ is ε-separated. This allows us to apply the variant of Lloyd's k-means algorithm analyzed by Ostrovsky et al. [4] that efficiently computes $\widetilde{X'_\alpha}$. Then we use Part (a) of Theorem 1.2 to establish the desired statement.

10.2 Spectral Embedding Properties

Boutsidis et al [1] showed that running an approximate k-means algorithm on an approximate spectral embedding $\widetilde{Y}$ computed by the "power method", yields an approximate clustering of the canonical spectral embedding Y. Here, we extend their result (c.f. [1, Theorem 6]) and prove that it is applicable to the normalized (approximate) spectral embedding Y' ($\widetilde{Y'}$).

Theorem 10.1. *[1, Theorem 6] Compute matrix $\widetilde{Y'}$ via the power method with*

$$p \geqslant \frac{\frac{1}{2} \cdot \ln\left(4 \cdot n \cdot \varepsilon^{-1} \cdot \delta_p^{-1} \cdot \sqrt{k}\right)}{\ln \gamma_k}, \text{ where } \gamma_k = \frac{1-\lambda_k(\mathcal{L}_G)}{1-\lambda_{k+1}(\mathcal{L}_G)}.$$

Run on the rows of $\widetilde{Y'}$ an α-approximate k-means algorithm with failure probability δ_α. Let the outcome be a clustering indicator matrix $\widetilde{X'_\alpha} \in \mathbb{R}^{n\times k}$. Then with probability at least $1-e^{-2n}-3\delta_p-\delta_\alpha$ it holds

$$\left\|Y' - \widetilde{X'_\alpha}\left(\widetilde{X'_\alpha}\right)^{\mathrm{T}} Y'\right\|_F^2 \leqslant (1+4\varepsilon)\cdot\alpha\cdot\left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}} Y'\right\|_F^2 + 4\varepsilon^2.$$

Our analysis builds upon the following key Lemma proved in [1].

Lemma 10.2. *[1, Lemma 5] Construct $\widetilde{Y}$ via the power method with*

$$p \geqslant \frac{\frac{1}{2} \cdot \ln\left(4 \cdot n \cdot \varepsilon^{-1} \cdot \delta_p^{-1} \cdot \sqrt{k}\right)}{\ln \gamma_k}, \text{ where } \gamma_k = \frac{1-\lambda_k(\mathcal{L}_G)}{1-\lambda_{k+1}(\mathcal{L}_G)}.$$

Then with probability at least $1-\exp\{-2n\}-3\delta_p$ it holds

$$\left\|YY^{\mathrm{T}} - \widetilde{Y}\widetilde{Y}^{\mathrm{T}}\right\|_F \leqslant \varepsilon.$$

The rest of this subsection is devoted to the proof of Theorem 10.1. We start by establishing several useful Lemmas that allows us to argue in a similar manner as in [1].

Useful Lemmas

Lemma 10.3. *$X'X'^{\mathrm{T}}$ is a projection matrix.*

Proof. By construction there are $d(v)$ many copies of row $U_k(v,:)/\sqrt{d(v)}$ in Y', for every vertex $v \in V$. We may assume w.l.o.g. that a k-means algorithm outputs an indicator matrix X' such that all copies of row $U_k(v,:)/\sqrt{d(v)}$ belong to the same cluster, for every $v \in V$. Moreover, by definition $X'_{ij} = 1/\sqrt{\mu(C_j)}$ if row $Y'_{i,:}$ belongs to the j-th cluster C_j and $X'_{ij} = 0$ otherwise, where matrix $X' \in \mathbb{R}^{m\times k}$. Therefore, it follows that $X'^{\mathrm{T}}X' = I_{k\times k}$ and thus $(X'X'^{\mathrm{T}})^2 = X'X'^{\mathrm{T}}$. ∎

Lemma 10.4. *It holds that $Y'^{\mathrm{T}}Y' = I_{k\times k} = \widetilde{Y'}^{\mathrm{T}}\widetilde{Y'}$.*

Proof. We prove now $Y'^{\mathrm{T}}Y' = I_{k\times k}$, but the equality $\widetilde{Y'}^{\mathrm{T}}\widetilde{Y'} = I_{k\times k}$ follows similarly. Since

$$\begin{aligned}\left(Y'^{\mathrm{T}}Y'\right)_{ij} &= \begin{pmatrix} \frac{U_k(1,i)}{\sqrt{d(1)}}\mathbf{1}_{d(1)}^{\mathrm{T}} & \cdots & \frac{U_k(n,i)}{\sqrt{d(n)}}\mathbf{1}_{d(n)}^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \frac{U_k(1,j)}{\sqrt{d(1)}}\mathbf{1}_{d(1)} \\ \cdots \\ \frac{U_k(n,j)}{\sqrt{d(n)}}\mathbf{1}_{d(n)} \end{pmatrix} \\ &= \sum_{\ell=1}^{n} d(\ell)\frac{U_k(\ell,i)}{\sqrt{d(\ell)}}\frac{U_k(\ell,j)}{\sqrt{d(\ell)}} = \langle U_k(:,i), U_k(:,j)\rangle = \delta_{ij},\end{aligned}$$

the statement follows. ∎

Lemma 10.5. *It holds that $\left\|Y'Y'^{\mathrm{T}} - \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}}\right\|_F = \left\|YY^{\mathrm{T}} - \widetilde{Y}\widetilde{Y}^{\mathrm{T}}\right\|_F$.*

Proof. By definition

$$Y'Y'^{\mathrm{T}} = \sum_{\ell=1}^{k} Y'_{:,\ell} Y'^{T}_{:,\ell} \quad \text{where} \quad Y'_{:,\ell} = \begin{pmatrix} \frac{U_k(1,\ell)}{\sqrt{d(1)}} \mathbf{1}_{d(1)} \\ \cdots \\ \frac{U_k(n,\ell)}{\sqrt{d(n)}} \mathbf{1}_{d(n)} \end{pmatrix}_{m \times 1}$$

and

$$\left(Y'_{:,\ell} Y'^{T}_{:,\ell}\right)_{d(i)d(j)} = \frac{U_k(i,\ell) U_k(j,\ell)}{\sqrt{d(i)d(j)}} \cdot \mathbf{1}_{d(1)} \mathbf{1}^{\mathrm{T}}_{d(j)}.$$

The statement follows by establishing the following chain of equalities

$$\begin{aligned}
\left\| Y'Y'^{\mathrm{T}} - \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} \right\|_F^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n} \left\| \left(Y'Y'^{\mathrm{T}} - \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} \right)_{d(i)d(j)} \right\|_F^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \left\| \sum_{\ell=1}^{k} \left(Y'_{:,\ell} Y'^{T}_{:,\ell} - \widetilde{Y'}_{:,\ell} \widetilde{Y'}^{\mathrm{T}}_{:,\ell} \right)_{d(i)d(j)} \right\|_F^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \left\| \left\{ \sum_{\ell=1}^{k} \left(\frac{U_k(i,\ell)U_k(j,\ell)}{\sqrt{d(i)d(j)}} - \frac{\widetilde{U}(i,\ell)\widetilde{U}(j,\ell)}{\sqrt{d(i)d(j)}} \right) \right\} \cdot \mathbf{1}_{d(i)} \mathbf{1}^{\mathrm{T}}_{d(j)} \right\|_F^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} d(i)d(j) \left[\sum_{\ell=1}^{k} \left(\frac{U_k(i,\ell)U_k(j,\ell)}{\sqrt{d(i)d(j)}} - \frac{\widetilde{U}(i,\ell)\widetilde{U}(j,\ell)}{\sqrt{d(i)d(j)}} \right) \right]^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \left[\sum_{\ell=1}^{k} \left(U_k(i,\ell)U_k(j,\ell) - \widetilde{U}(i,\ell)\widetilde{U}(j,\ell) \right) \right]^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \left(U_k U_k^{\mathrm{T}} - \widetilde{U}\widetilde{U}^{\mathrm{T}} \right)_{ij}^2 \\
&= \left\| U_k U_k^{\mathrm{T}} - \widetilde{U}\widetilde{U}^{\mathrm{T}} \right\|_F^2 = \left\| YY^{\mathrm{T}} - \widetilde{Y}\widetilde{Y}^{\mathrm{T}} \right\|_F^2.
\end{aligned}$$

■

Lemma 10.6. *For any matrix U with orthonormal columns and every matrix A it holds*

$$\left\| UU^{\mathrm{T}} - AA^{\mathrm{T}}UU^{\mathrm{T}} \right\|_F = \left\| U - AA^{\mathrm{T}}U \right\|_F. \tag{28}$$

Proof. The statement follows by the Frobenius norm property $\|B\|_F^2 = \mathrm{Tr}[B^{\mathrm{T}}B]$, the cyclic property of trace $\mathrm{Tr}[UB^{\mathrm{T}}BU^{\mathrm{T}}] = \mathrm{Tr}[B^{\mathrm{T}}B \cdot U^{\mathrm{T}}U]$ and the orthogonality of matrix U. ■

Proof of Theorem 10.1

Using Lemma 10.2 and Lemma 10.5 with probability at least $1 - \exp\{-2n\} - 3\delta_p$ we have

$$\left\| Y'Y'^{\mathrm{T}} - \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} \right\|_F = \left\| YY^{\mathrm{T}} - \widetilde{Y}\widetilde{Y}^{\mathrm{T}} \right\|_F \leqslant \varepsilon.$$

Let $Y'Y'^{\mathrm{T}} = \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} + E$ such that $\|E\|_F \leqslant \varepsilon$. Based on Lemma 10.4 and Lemma 10.6 we have that equation (28) holds for the matrices Y' and $\widetilde{Y'}$. Therefore, by Lemma 10.3 we can apply the proof in [1, Theorem 6] to obtain

$$\left\| Y' - \widetilde{X'_\alpha} \left(\widetilde{X'_\alpha} \right)^{\mathrm{T}} Y' \right\|_F \leqslant \sqrt{\alpha} \cdot \left(\left\| Y' - X'_{\text{opt}} \left(X'_{\text{opt}} \right)^{\mathrm{T}} Y' \right\|_F + 2\varepsilon \right). \tag{29}$$

After a simple manipulation, (29) yields the desired statement.

10.3 Spectral Embeddings, Gap Assumption and ε-separability

In this subsection, we prove under the gap assumption that the approximate normalized spectral embedding $\widetilde{Y'}$ is ε-separated, i.e. $\triangle_k(\widetilde{\mathcal{X}_V}) < 5\varepsilon^2 \cdot \triangle_{k-1}(\widetilde{\mathcal{X}_V})$. Our analysis builds upon Theorem 9.1, Theorem 10.1 and the proof techniques in [1].

Theorem 10.7 (Approximate Normalized Spectral Embedding is ε-separated)**.** *Suppose the gap assumption satisfies $\Psi = 20^4 \cdot k^3/\delta$, $k/\delta \geqslant 10^9$ for some $\delta \in (0, 1/2]$ and the optimum cost $\left\| Y' - X'_{\text{opt}}(X'_{\text{opt}})^{\mathrm{T}} Y' \right\|_F \geqslant n^{-O(1)}$. Construct matrix $\widetilde{Y'}$ via the power method with $p \geqslant \Omega(\frac{1}{\lambda_{k+1}} \ln n)$. Then for $\varepsilon = 6 \cdot 10^{-7}$ with high probability it holds*

$$\triangle_k \left(\widetilde{\mathcal{X}_V} \right) < 5\varepsilon^2 \cdot \triangle_{k-1} \left(\widetilde{\mathcal{X}_V} \right).$$

Before we present the proof of Theorem 10.7 we will establish two technical results.

Lemma 10.8. *If $\Psi \geqslant 20^4 \cdot k^3/\delta$ for $\delta \in (0, 1/2]$ it holds*

$$\ln \left(\frac{1 - \lambda_k}{1 - \lambda_{k+1}} \right) \geqslant \left(1 - \frac{4\delta}{20^4 k^2} \right) \lambda_{k+1}.$$

Proof. Lee et al. [2] proved that higher order Cheeger's inequality satisfies

$$\lambda_k / 2 \leqslant \rho(k) \leqslant O(k^2) \cdot \sqrt{\lambda_k}. \tag{30}$$

Using the LHS of (30) we have

$$k^3 \widehat{\rho}_{\text{avr}}(k) = k^2 \sum_{i=1}^{k} \phi(P_i) \geqslant k^2 \max_{i \in [1:k]} \phi(P_i) \geqslant k^2 \cdot \rho(k) \geqslant \frac{k^2 \lambda_k}{2}$$

and thus we can upper bound the k-th smallest eigenvalue of $\mathcal{L}_G$ by

$$\lambda_k \leqslant 2k \cdot \widehat{\rho}_{\text{avr}}(k).$$

Moreover, by the gap assumption it follows that

$$\lambda_{k+1} \geqslant \frac{20^4 k^2}{2\delta} \cdot 2k \cdot \widehat{\rho}_{\text{avr}}(k) \geqslant \frac{20^4 k^2}{2\delta} \cdot \lambda_k.$$

The statement follows by

$$\frac{1 - \lambda_k}{1 - \lambda_{k+1}} \geqslant \frac{1 - \frac{2\delta}{20^4 k^2} \lambda_{k+1}}{1 - \lambda_{k+1}} \geqslant \exp \left\{ \left(1 - \frac{4\delta}{20^4 k^2} \right) \lambda_{k+1} \right\}.$$

∎

To state our next results we need some notations. We use interchangeably X'_{opt} with $X'^{(k)}_{\mathrm{opt}}$ to denote the optimal indicator matrix for the k-means problem on $\mathcal{X}_V$ that is induced by the rows of matrix Y'. Similarly, we denote by $X'^{(k-1)}_{\mathrm{opt}}$ the optimal indicator matrix for the $(k-1)$-means problem on $\mathcal{X}_V$.

Based on Lemma 1.3 and the definition of Y' and $X'^{(k)}_{\mathrm{opt}}$ we obtain the following statement.

Corollary 10.9. *Let G be a graph that satisfies $\Psi = 20^4 \cdot k^3/\delta$, $\delta \in (0, 1/2]$ and $k/\delta \geqslant 10^9$. Then it holds*

$$\left\| Y' - X'^{(k)}_{\mathrm{opt}} \left(X'^{(k)}_{\mathrm{opt}} \right)^{\mathrm{T}} Y' \right\|_F^2 \leqslant \frac{1}{8 \cdot 10^{13}}.$$

We are now ready to present to proof of Theorem 10.7.

Proof of Theorem 10.7. By Theorem 9.1 we have

$$\left\| Y' - X'^{(k)}_{\mathrm{opt}} \left(X'^{(k)}_{\mathrm{opt}} \right)^{\mathrm{T}} Y' \right\|_F \leqslant \varepsilon \left\| Y' - X'^{(k-1)}_{\mathrm{opt}} \left(X'^{(k-1)}_{\mathrm{opt}} \right)^{\mathrm{T}} Y' \right\|_F. \tag{31}$$

We set the approximation parameter in Theorem 10.1 to

$$\varepsilon' \triangleq \frac{1}{4} \sqrt{\triangle_k(\mathcal{X}_V)} = \frac{1}{4} \left\| Y' - X'^{(k)}_{\mathrm{opt}} \left(X'^{(k)}_{\mathrm{opt}} \right)^{\mathrm{T}} Y' \right\|_F \geqslant n^{-O(1)}, \tag{32}$$

and we note that by Theorem 9.1 it holds

$$\varepsilon' \leqslant \frac{\varepsilon}{4} \sqrt{\triangle_{k-1}(\mathcal{X}_V)}. \tag{33}$$

Construct the matrix $\widetilde{Y}$ via the power method with $p \geqslant \Omega(\frac{1}{\lambda_{k+1}} \ln n)$. By combining Lemma 10.2 and Lemma 10.5 we obtain with high probability

$$\left\| Y'Y'^{\mathrm{T}} - \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} \right\|_F = \left\| YY^{\mathrm{T}} - \widetilde{Y}\widetilde{Y}^{\mathrm{T}} \right\|_F \leqslant \varepsilon'.$$

Let $Y'Y'^{\mathrm{T}} = \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} + E$ such that $\|E\|_F \leqslant \varepsilon'$. By Lemma 10.4 we have $Y'^{\mathrm{T}}Y' = I_{k \times k} = \widetilde{Y'}^{\mathrm{T}}\widetilde{Y'}$ and thus (28) in Lemma 10.6 holds for the orthonormal matrices Y' and $\widetilde{Y'}$. Therefore, by Lemma 10.3 we have

$$\begin{aligned}
\sqrt{\triangle_k\left(\widetilde{\mathcal{X}_V}\right)} &= \left\| \widetilde{Y'} - \widetilde{X'^{(k)}_{\mathrm{opt}}} \left(\widetilde{X'^{(k)}_{\mathrm{opt}}} \right)^{\mathrm{T}} \widetilde{Y'} \right\|_F = \left\| \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} - \widetilde{X'^{(k)}_{\mathrm{opt}}} \left(\widetilde{X'^{(k)}_{\mathrm{opt}}} \right)^{\mathrm{T}} \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} \right\|_F \\
&= \left\| Y'Y'^{\mathrm{T}} - \widetilde{X'^{(k)}_{\mathrm{opt}}} \left(\widetilde{X'^{(k)}_{\mathrm{opt}}} \right)^{\mathrm{T}} Y'Y'^{\mathrm{T}} - \left(I - \widetilde{X'^{(k)}_{\mathrm{opt}}} \left(\widetilde{X'^{(k)}_{\mathrm{opt}}} \right)^{\mathrm{T}} \right) E \right\|_F \\
&\leqslant \|E\|_F + \left\| Y' - \widetilde{X'^{(k)}_{\mathrm{opt}}} \left(\widetilde{X'^{(k)}_{\mathrm{opt}}} \right)^{\mathrm{T}} Y' \right\|_F
\end{aligned}$$

By Lemma 10.8 we can apply Theorem 10.1 which yields

$$\left\| Y' - \widetilde{X'^{(k)}_{\mathrm{opt}}} \left(\widetilde{X'^{(k)}_{\mathrm{opt}}} \right)^{\mathrm{T}} Y' \right\|_F^2 \leqslant (1 + 4\varepsilon') \cdot \left\| Y' - X'^{(k)}_{\mathrm{opt}} \left(X'^{(k)}_{\mathrm{opt}} \right)^{\mathrm{T}} Y' \right\|_F^2 + 4\varepsilon'^2.$$

By Corollary 10.9 we derive an upper bound on the optimal k-means cost of $\mathcal{X}_V$

$$\left\| Y' - X'^{(k)}_{\text{opt}} \left(X'^{(k)}_{\text{opt}} \right)^{\mathrm{T}} Y' \right\|_F^2 \leqslant \frac{1}{8 \cdot 10^{13}} \tag{34}$$

that combined with the definition of ε' gives

$$\begin{aligned} \sqrt{\triangle_k\left(\widetilde{\mathcal{X}_V}\right)} &\leqslant \varepsilon' + \sqrt{(1+4\varepsilon') \left\| Y' - X'^{(k)}_{\text{opt}} \left(X'^{(k)}_{\text{opt}} \right)^{\mathrm{T}} Y' \right\|_F^2 + 4\varepsilon'^2} \\ &\leqslant 2 \left\| Y' - X'^{(k)}_{\text{opt}} \left(X'^{(k)}_{\text{opt}} \right)^{\mathrm{T}} Y' \right\|_F = 2\sqrt{\triangle_k(\mathcal{X}_V)} \\ &\leqslant 2\varepsilon \cdot \sqrt{\triangle_{k-1}(\mathcal{X}_V)}. \end{aligned} \tag{35}$$

Moreover, it holds that

$$\begin{aligned} \sqrt{\triangle_{k-1}(\mathcal{X}_V)} &= \left\| Y' - X'^{(k-1)}_{\text{opt}} \left(X'^{(k-1)}_{\text{opt}} \right)^{\mathrm{T}} Y' \right\|_F \leqslant \left\| Y' - \widetilde{X'^{(k-1)}_{\text{opt}}} \left(\widetilde{X'^{(k-1)}_{\text{opt}}} \right)^{\mathrm{T}} Y' \right\|_F \\ &= \left\| Y'Y'^{\mathrm{T}} - \widetilde{X'^{(k-1)}_{\text{opt}}} \left(\widetilde{X'^{(k-1)}_{\text{opt}}} \right)^{\mathrm{T}} Y'Y'^{\mathrm{T}} \right\|_F \\ &= \left\| \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} - \widetilde{X'^{(k-1)}_{\text{opt}}} \left(\widetilde{X'^{(k-1)}_{\text{opt}}} \right)^{\mathrm{T}} \widetilde{Y'}\widetilde{Y'}^{\mathrm{T}} + \left(I - \widetilde{X'^{(k-1)}_{\text{opt}}} \left(\widetilde{X'^{(k-1)}_{\text{opt}}} \right)^{\mathrm{T}} \right) E \right\|_F \\ &\leqslant \left\| \widetilde{Y'} - \widetilde{X'^{(k-1)}_{\text{opt}}} \left(\widetilde{X'^{(k-1)}_{\text{opt}}} \right)^{\mathrm{T}} \widetilde{Y'} \right\|_F + \|E\|_F \\ &\leqslant \sqrt{\triangle_{k-1}\left(\widetilde{\mathcal{X}_V}\right)} + \frac{\varepsilon}{4}\sqrt{\triangle_{k-1}(\mathcal{X}_V)} \end{aligned}$$

and thus

$$\sqrt{\triangle_{k-1}(\mathcal{X}_V)} \leqslant \left(1 + \frac{\varepsilon}{2}\right) \sqrt{\triangle_{k-1}\left(\widetilde{\mathcal{X}_V}\right)}. \tag{36}$$

Therefore, by combining (35) and (36) we obtain the desired statement

$$\sqrt{\triangle_k\left(\widetilde{\mathcal{X}_V}\right)} \leqslant 2\varepsilon \cdot \sqrt{\triangle_{k-1}(\mathcal{X}_V)} \leqslant (2+\varepsilon) \cdot \varepsilon \cdot \sqrt{\triangle_{k-1}\left(\widetilde{\mathcal{X}_V}\right)}.$$

■

10.4 Proof of Part (b) of Theorem 1.2

Our analysis crucially depends on the following variant of Lloyd's k-means algorithm analyzed by Ostrovsky et al. [4].

Theorem 10.10. *[4, Theorem 4.15] Assuming that $\triangle_k(\mathcal{X}) \leqslant \varepsilon^2 \triangle_{k-1}(\mathcal{X})$ for $\varepsilon \in (0, 6 \cdot 10^{-7}]$, there is an algorithm that returns a solution of cost at most*

$$\frac{1-\varepsilon^2}{1-37\varepsilon^2} \cdot \triangle_k(\mathcal{X})$$

with probability at least $1 - O(\sqrt{\varepsilon})$ in time $O(nkd + k^3 d)$.

Proof of Part (b) of Theorem 1.2. Let $p = \Theta((\ln n)/\lambda_{k+1})$. We can compute the matrix $B = \mathcal{A}_N^{2p+1}S$ in time $O(mkp)$ and its singular value decomposition $\widetilde{U}\widetilde{\Sigma}\widetilde{V}^{\mathrm{T}}$ in time $O(nk^2)$. Based on it we construct in time $O(mk)$ matrix $\widetilde{Y'}$ (c.f. (26)).

By Theorem 10.7, $\widetilde{\mathcal{X}_V}$ is ε-separated for $\varepsilon = 6 \cdot 10^{-7}$, i.e. $\triangle_k\left(\widetilde{\mathcal{X}_V}\right) < 5\varepsilon^2 \cdot \triangle_{k-1}\left(\widetilde{\mathcal{X}_V}\right)$. Hence, by Theorem 10.10 there is an algorithm that outputs a clustering with indicator matrix $\widetilde{X'_\alpha}$ that has a cost at most

$$\left\|\widetilde{Y'} - \widetilde{X'_\alpha}\left(\widetilde{X'_\alpha}\right)^{\mathrm{T}}\widetilde{Y'}\right\|_F^2 \leqslant \left(1 + \frac{1}{10^{10}}\right) \cdot \left\|\widetilde{Y'} - \widetilde{X'_{\mathrm{opt}}}\left(\widetilde{X'_{\mathrm{opt}}}\right)^{\mathrm{T}}\widetilde{Y'}\right\|_F^2$$

with constant probability (close to 1) in time $O(mk^2 + k^4)$, where $\alpha = 1 + 10^{-10}$.

We apply now Theorem 10.1 with $\varepsilon' = \frac{\sqrt{\delta_A}}{4}\left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}}Y'\right\|_F$, where $\delta_A \in (0,1)$ is to be determined soon. Moreover, by Corollary 10.9 we have

$$\left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}}Y'\right\|_F < \frac{1}{10^6}$$

and thus with constant probability it holds

$$\begin{aligned}
&\left\|Y' - \widetilde{X'_\alpha}\left(\widetilde{X'_\alpha}\right)^{\mathrm{T}}Y'\right\|_F^2 \leqslant (1+4\varepsilon')\alpha\left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}}Y'\right\|_F^2 + 4\varepsilon'^2 \\
= \;& \left[\left(1 + \sqrt{\delta_A}\left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}}Y'\right\|_F\right)\alpha + \frac{\delta_A}{4}\right] \cdot \left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}}Y'\right\|_F^2 \\
\leqslant \;& \left[\left(1 + \frac{\sqrt{\delta_A}}{10^6}\right) \cdot \left(1 + \frac{1}{10^{10}}\right) + \frac{\delta_A}{4}\right] \cdot \left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}}Y'\right\|_F^2.
\end{aligned}$$

The indicator matrix $\widetilde{X'_\alpha}$ yields a relative approximation of $\mathcal{X}_V$ that satisfies for $\delta_A = 1/10^6$

$$\left\|Y' - \widetilde{X'_\alpha}\left(\widetilde{X'_\alpha}\right)^{\mathrm{T}}Y'\right\|_F^2 \leqslant \left(1 + \frac{1}{10^6}\right)\left\|Y' - X'_{\mathrm{opt}}\left(X'_{\mathrm{opt}}\right)^{\mathrm{T}}Y'\right\|_F^2. \tag{37}$$

The statement follows by Part (a) of Theorem 1.2 applied to the partition $(A_1, \ldots, A_k)$ of V that is induced by the indicator matrix $\widetilde{X'_\alpha}$. ■

11 Parameterized Upper Bound on $\widehat{\rho}_{\mathrm{avr}}(k)$

A k-disjoint tuple Z is a k-tuple $(Z_1, \ldots, Z_k)$ of disjoint subsets of V. A k-way partition $(P_1, \ldots, P_k)$ of V is compatible with a k-disjoint tuple Z if $Z_i \subseteq P_i$ for all i. We then define $S_i = P_i \backslash Z_i$ and use $\mathcal{P}_Z$ to denote all partitions compatible with Z. We use $\mathcal{Z}_k$ to denote all k-tuples Z with $\rho(k) = \Phi(Z) = \Phi(Z_1, \ldots, Z_k)$. The elements of $\mathcal{Z}_k$ are called optimal (k-disjoint) tuples. We denote all partitions compatible with some optimal k-tuple by

$$\mathcal{P}_k = \cup_{Z \in \mathcal{Z}_k}\mathcal{P}_Z. \tag{38}$$

Oveis Gharan and Trevisan [5, Lemma 2.5] proved that for every k-disjoint tuple $Z \in \mathcal{Z}_k$ there is a k-way partition $(P_1, \ldots, P_k) \in \mathcal{P}_Z$ with

$$\Phi(P_1, \ldots, P_k) \leqslant k\rho(k). \tag{39}$$

Remark 11.1. *In this section, we assume that every partition* $(P_1, \ldots, P_k) \in \mathcal{P}_k$ *satisfies*

$$\Phi(P_1, \ldots, P_k) > \rho(k), \tag{40}$$

since otherwise $\widehat{\rho}(k) = \rho(k)$.

We refine the analysis in [5] and prove a parameterized upper bound on $\widehat{\rho}_{\mathrm{avr}}(k)$ that depends on a natural combinatorial parameter and the average conductance of a k-disjoint tuple $Z \in \mathcal{Z}_k$. Before we state our results, we need some notations.

We define the order k inter-connection constant of a graph G by

$$\rho_{\mathcal{P}}(k) \triangleq \min_{P_1,\ldots,P_k \in \mathcal{P}_k} \Phi_{IC}\left(P_1, \ldots, P_k\right) \tag{41}$$

where

$$\Phi_{IC}\left(P_1, \ldots, P_k\right) \triangleq \max_{S_i \neq \emptyset} \frac{|E(S_i, V \backslash P_i)| - |E(S_i, Z_i)|}{|E(P_i, V \backslash P_i)|}. \tag{42}$$

We will prove in Lemma 11.5 that $\rho_{\mathcal{P}}(k) \in (0, 1 - 1/(k-1)]$. Furthermore, let $\mathcal{O}_{\mathcal{P}}$ be the set of all k-way partitions $(P_1, \ldots, P_k) \in \mathcal{P}_k$ with $\Phi_{IC}(P_1, \ldots, P_k) = \rho_{\mathcal{P}}(k)$, i.e., the set of all partitions that achieve the order k inter-connection constant. Let

$$\widetilde{\rho}_{\mathrm{avr}}(k) = \min_{(P_1,\ldots,P_k) \in \mathcal{O}_{\mathcal{P}}} \frac{1}{k} \sum_{i=1}^{k} \phi(P_i) \tag{43}$$

be the minimal *average* conductance over all k-way partitions in $\mathcal{O}_{\mathcal{P}}$. By construction it holds that

$$\widehat{\rho}_{\mathrm{avr}}(k) \leqslant \widetilde{\rho}_{\mathrm{avr}}(k). \tag{44}$$

We present now our main result of this Section which upper bounds $\widetilde{\rho}_{\mathrm{avr}}(k)$.

Theorem 11.2. *For any graph* G *there exists a* k*-way partition* $(P_1, \ldots, P_k) \in \mathcal{O}_{\mathcal{P}}$ *compatible with a* k*-disjoint tuple* Z *with* $\Phi(Z_1, \ldots, Z_k) = \rho(k)$ *such that for* $\kappa_{\mathcal{P}} \triangleq [1 - \rho_{\mathcal{P}}(k)]^{-1} \in (1, k-1]$ *it holds*

$$\widetilde{\rho}_{\mathrm{avr}}(k) \leqslant \frac{\kappa_{\mathcal{P}}}{k} \sum_{i=1}^{k} \phi(Z_i)$$

and in addition, for every $i \in [1:k]$

$$\phi(P_i) \leqslant \kappa_{\mathcal{P}} \cdot \phi(Z_i).$$

Our goal now is to prove Theorem 11.2. We establish first a few useful Lemmas that will be used to prove Lemma 11.5 and Theorem 11.2.

Oveis Gharan and Trevisan [5, Algorithm 2 and Fact 2.4] showed that

Fact 11.3 ([5])**.** *For any* k*-disjoint tuple* Z*, there is a* k*-way partition* $(P_1, \ldots, P_k) \in \mathcal{P}_Z$ *such that*

1. *For every* $i \in [1:k]$, $Z_i \subseteq P_i$.
2. *For every* $i \in [1:k]$*, and every subset* $\emptyset \neq S \subseteq P_i \backslash Z_i$ *it holds*

$$|E(S, P_i \backslash S)| \geqslant \frac{1}{k} |E(S, V \backslash S)|.$$

Lemma 11.4. *For any k-disjoint tuple Z, there exists a k-way partition $(P_1, \ldots, P_k) \in \mathcal{P}_Z$ that satisfies*

$$\max_{S_i \neq \emptyset} \frac{|E(S_i, V \backslash P_i)| - |E(S_i, Z_i)|}{|E(P_i, V \backslash P_i)|} \leqslant 1 - \frac{1}{k-1}.$$

Proof. By Fact 11.3 there is a k-way partition $(P_1, \ldots, P_k) \in \mathcal{P}_Z$ such that for all i it holds

$$|E(S_i, Z_i)| = |E(S_i, P_i \backslash S_i)| \geqslant \frac{1}{k} |E(S_i, V \backslash S_i)| = \frac{1}{k} \left(|E(S_i, V \backslash P_i)| + |E(S_i, Z_i)|\right)$$

and hence

$$|E(S_i, Z_i)| \geqslant \frac{1}{k-1} |E(S_i, V \backslash P_i)|.$$

■

Lemma 11.5. *The order k inter-connection constant of a graph G is bounded by*

$$0 < \rho_{\mathcal{P}}(k) \leqslant 1 - \frac{1}{k-1}.$$

Proof. We prove first the upper bound. By Lemma 11.4 there is a k-way partition $(P_1, \ldots, P_k) \in \mathcal{P}_k$ compatible with a k-disjoint tuple Z such that

$$\max_{S_i \neq \emptyset} \frac{|E(S_i, V \backslash P_i)| - |E(S_i, Z_i)|}{|E(P_i, V \backslash P_i)|} \leqslant 1 - \frac{1}{k-1}.$$

Therefore,

$$\begin{aligned} \rho_{\mathcal{P}}(k) &= \min_{P'_1, \ldots, P'_k \in \mathcal{P}_k} \Phi_{IC}\left(P'_1, \ldots, P'_k\right) \leqslant \Phi_{IC}\left(P_1, \ldots, P_k\right) \\ &= \max_{S_i \neq \emptyset} \frac{|E(S_i, V \backslash P_i)| - |E(S_i, Z_i)|}{|E(P_i, V \backslash P_i)|} \leqslant 1 - \frac{1}{k-1}. \end{aligned}$$

We prove now the lower bound. Suppose for contradiction that $\rho_{\mathcal{P}}(k) \leqslant 0$. By definition we have

$$\begin{aligned} \phi(P_i) &= \frac{|E(P_i, V \backslash P_i)|}{\mu(P_i)} = \frac{|E(Z_i, V \backslash Z_i)| + |E(S_i, V \backslash P_i)| - |E(S_i, Z_i)|}{\mu(P_i)} \\ &\leqslant \phi(Z_i) + \frac{|E(S_i, V \backslash P_i)| - |E(S_i, Z_i)|}{\mu(P_i)} \end{aligned}$$

By (41), it holds for any $S_i \neq \emptyset$ that

$$|E(S_i, V \backslash P_i)| - |E(S_i, Z_i)| \leqslant \rho_{\mathcal{P}}(k) \cdot |E(P_i, V \backslash P_i)|$$

and thus

$$\phi(P_i) \quad \begin{cases} \leqslant \phi(Z_i) - |\rho_{\mathcal{P}}(k)| \cdot \phi(P_i) & \text{, if } S_i \neq \emptyset; \\ = \phi(Z_i) & \text{, otherwise.} \end{cases}$$

However, this contradicts $\Phi(P_1, \ldots, P_k) > \rho(k)$ and thus the statement follows. ■

We are now ready to prove Theorem 11.2.

Proof of Theorem 11.2. Let $(P_1, \ldots, P_k) \in \mathcal{O}_\mathcal{P}$ be a k-way partition compatible with a k-disjoint tuple $Z \in \mathcal{Z}_k$ that satisfies $\Phi(Z_1, \ldots, Z_k) = \rho(k)$. By Lemma 11.5 there is a real number such that

$$\kappa_\mathcal{P} \triangleq [1 - \rho_\mathcal{P}(k)]^{-1} \in (1, k-1]. \tag{45}$$

We argue in a similar manner as in Lemma 11.5 to obtain

$$\phi(P_i) \quad \begin{cases} \leqslant \phi(Z_i) - \rho_\mathcal{P}(k) \cdot \phi(P_i) & \text{, if } S_i \neq \emptyset; \\ = \phi(Z_i) & \text{, otherwise.} \end{cases} \tag{46}$$

By combining (45) and the first conclusion of (46) we have

$$\phi(P_i) \leqslant [1 - \rho_\mathcal{P}(k)]^{-1} \cdot \phi(Z_i) = \kappa_\mathcal{P} \cdot \phi(Z_i). \tag{47}$$

The statement follows by combining (43) and (47), since

$$\widetilde{\rho}_{\text{avr}}(k) \leqslant \frac{1}{k} \sum_{i=1}^{k} \phi(P_i) \leqslant \frac{\kappa_\mathcal{P}}{k} \sum_{i=1}^{k} \phi(Z_i).$$

■

References

[1] C. Boutsidis, P. Kambadur, and A. Gittens. Spectral clustering via the power method - provably. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 40–48, 2015. URL http://jmlr.org/proceedings/papers/v37/boutsidis15.html.

[2] J. R. Lee, S. Oveis Gharan, and L. Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 1117–1130, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1245-5. doi: 10.1145/2213977.2214078. URL http://doi.acm.org/10.1145/2213977.2214078.

[3] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, pages 849–856. MIT Press, 2002.

[4] R. Ostrovsky, Y. Rabani, L. J. Schulman, and C. Swamy. The effectiveness of lloyd-type methods for the k-means problem. *J. ACM*, 59(6):28:1–28:22, Jan. 2013. ISSN 0004-5411. doi: 10.1145/2395116.2395117. URL http://doi.acm.org/10.1145/2395116.2395117.

[5] S. Oveis Gharan and L. Trevisan. Partitioning into expanders. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1256–1266, 2014. doi: 10.1137/1.9781611973402.93. URL http://dx.doi.org/10.1137/1.9781611973402.93.

[6] R. Peng, H. Sun, and L. Zanetti. Partitioning well-clustered graphs: Spectral clustering works! In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pages 1423–1455, 2015. URL http://jmlr.org/proceedings/papers/v40/Peng15.html. see also http://arxiv.org/abs/1411.2021.

[7] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/34.868688.

[8] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, pages 395–416, 2007.