A Performance Guarantee for Spectral Clustering*

March Boedihardjo[†], Shaofeng Deng[‡], and Thomas Strohmer[§]

Abstract. The two-step spectral clustering method, which consists of the Laplacian eigenmap and a rounding step, is a widely used method for graph partitioning. It can be seen as a natural relaxation to the NP-hard minimum ratio cut problem. In this paper we study the central question: when is spectral clustering able to find the global solution to the minimum ratio cut problem? First we provide a condition that naturally depends on the intra- and inter-cluster connectivities of a given partition under which we may certify that this partition is the solution to the minimum ratio cut problem. Then we develop a deterministic two-to-infinity norm perturbation bound for the the invariant subspace of the graph Laplacian that corresponds to the k smallest eigenvalues. Finally by combining these two results we give a condition under which spectral clustering is guaranteed to output the global solution to the minimum ratio cut problem, which serves as a performance guarantee for spectral clustering.

Key words. graph partitioning, ratio cut, spectral clustering, Laplacian eigenmap, matrix perturbation theory

AMS subject classifications. 90B10, 62H30, 47A55

1. Introduction.

The graph partitioning problem is ubiquitous in data analysis [4]: how to partition a graph into a given number of subgraphs so that the connections among them are weak? One popular measurement for how well the graph is partitioned is the ratio cut of this partition. Let G be an undirected graph with vertex set $V = \{v_1, \cdots, v_n\}$. We assume that the graph G is weighted, that is each edge between two vertices v_i and v_j carries a non-negative weight $w_{ij} \geq 0$ ($w_{ii} = 0$). The weighted adjacency matrix of the graph is the symmetric matrix $W = (w_{ij})$. Given a k-way partition of the vertices $\{V_i\}_{i=1}^k$ ($\sqcup_{i=1}^k V_i = V$), the ratio cut of this partition is defined to be

$$\operatorname{RatioCut}\left(\{V_i\}_{i=1}^k\right) = \sum_{i=1}^k \frac{\operatorname{Cut}(V_i, V_i^c)}{|V_i|},$$

where

$$\operatorname{Cut}(V_i, V_i^c) = \sum_{v_j \in V_i, v_k \in V_i^c} w_{jk}$$

is the total weight between V_i and V_i^c. The ratio cut measures the connections among the subgraphs normalized by the size of the subgraphs. The purpose of the normalization is to discourage unbalanced partitions. Hence we are interested in finding a k-way partition that

---

*Submitted to the editors DATE.

Funding: M.B. acknowledges support from the NSF via grant DMS 1856221. S.D. and T.S. acknowledge support from the NSF via grants DMS 1620455, DMS 1737943, and CCF 1934568.

[†]Department of Mathematics, University of California Los Angeles (Email: march@math.ucla.edu).

[‡]Department of Mathematics, University of California Davis (Email: sfdeng@math.ucdavis.edu).

[§]Center of Data Science and Artificial Intelligence Research and Department of Mathematics, University of California at Davis (Email: strohmer@math.ucdavis.edu).

has the minimum ratio cut, which is presumed to be a NP-hard problem ([23]). Spectral clustering is a natural relaxation to this NP-hard problem. We begin by defining the graph Laplacian of G. Let

$$d_i = \deg(v_i) = \sum_{j \neq i} w_{ij}$$

denote the degree of vertex v_i. Let the diagonal matrix D be the degree matrix with the degrees $d_1, \cdots, d_n$ on the diagonal. The graph Laplacian of the graph is then defined to be

$$L = D - W.$$

Note that we can rewrite

$$\text{RatioCut}(\{V_i\}_{i=1}^k) = \sum_{i=1}^k \frac{\mathbb{1}_{V_i}^T L \mathbb{1}_{V_i}}{|V_i|} = \text{Tr}\left(U^T L U\right),$$

where $U \in \mathbb{R}^{n \times k}$ has its ith column $U_{\cdot i}$ being $\frac{1}{\sqrt{|V_i|}} \mathbb{1}_{V_i}$ and $\mathbb{1}_{V_i}$ is the indicator vector that take value 1 on the vertices in V_i and 0 elsewhere. Therefore the minimum ratio cut problem can be formulated as

$$\min_{\{V_i\}_{i=1}^k} \text{Tr}\left(U^T L U\right) \quad \text{s.t. } U_{\cdot i} = \frac{1}{\sqrt{|V_i|}} \mathbb{1}_{V_i} \text{ for } i \in [k]. \tag{1.1}$$

Spectral clustering relaxes the combinatorial constraint of U and instead seeks a solution among all matrices U with orthonormal columns. So the relaxed problem is

$$\min_{U \in \mathbb{R}^{n \times k}} \text{Tr}\left(U^T L U\right) \quad \text{s.t. } U^T U = I_k, \tag{1.2}$$

whose solution U can be shown to be the eigenvectors w.r.t. the k smallest eigenvalues of L. Since the columns of U are no longer a collection of indicator vectors, a rounding step is necessary to obtain the partition. The rounding step is performed on the rows of U. Namely one should treat the ith row $U_{i\cdot}$ as the embedding of vertex v_i in $\mathbb{R}^k$ and obtain the partition by clustering those points (usually through k-means) in $\mathbb{R}^k$. A justification for this idea is the following equivalence form of the relaxed problem (1.2):

$$\min_{U \in \mathbb{R}^{n \times k}} \sum_{i=1}^n \sum_{j=1}^n w_{ij} \left|\left|U_{i\cdot} - U_{j\cdot}\right|\right|_2^2 \quad \text{s.t. } U^T U = I_k. \tag{1.3}$$

Hence $U_{i\cdot}$ and $U_{j\cdot}$ tend to be close in $\mathbb{R}^k$ if v_i and v_j are strongly connected in G. For this reason we call U the Laplacian eigenmap of G. Spectral clustering, which consists of the Laplacian eigenmap and a rounding step, is shown in Algorithm 1.1.

In this paper we try to answer the fundamental question: under what condition is Algorithm 1.1, a relaxation of the minimum ratio cut problem (1.1), able to find the global minimum of (1.1)?

Algorithm 1.1 Spectral clustering

1: **Input:** Weighted adjacency matrix W and the number of clusters k.
2: Compute the graph Laplacian $L = D - W$.
3: Compute $U \in \mathbb{R}^{n\times k}$ whose columns are the eigenvectors correspond to the k smallest eigenvalues of L.
4: Treat $U_{i\cdot}$ as the embedding of vertex v_i in $\mathbb{R}^k$ and apply clustering method (k-means etc.,) on the points $\{U_{i\cdot}\}_{i=1}^n$.
5: Obtain the partition $\{V_i\}_{i=1}^k$ of V based on the result form step 4.

1.1. Related work.

Spectral clustering is a popular graph partition method. We refer the readers to [22] for an excellent survey on this subject, whose topics include basic properties of the graph Laplacian, variants of spectral clustering methods, constructing similarity graphs from non-graph data, different perspectives of spectral clustering, etc. Even though we have yet to fully understand the mechanism of spectral clustering, some excellent research has been done about its theoretical analysis. One of the most prominent ones is the work on (higher-order) Cheeger-type inequalities [6, 13]. Another closely related work is [16] which gives performance guarantees for a SDP relaxation to (1.1). In fact our Theorem 2.2 is a direct improvement to their work. For an analysis of the spectral clustering method on random graphs we refer to [10, 9, 14, 20, 21, 1].

The technical tool we use is the invariant subspace perturbation theory which studies the change to the invariant subspace of a self-adjoint matrix after the matrix is perturbed. One of the most celebrated works is the classic Davis-Kahan theorem [8] which bounds the invariant subspace perturbation in term of canonical angle. Recent years have witnessed a surge of research on the two-to-infinity norm bound of the invariant subspace perturbation, which is more suitable in many applications. The result we use for this paper is from the remarkable paper by A. Damle and Y. Sun [7]. Other related work on this topic includes [1, 11, 10, 5].

1.2. Notation.

We introduce some notation which will be used throughout this paper. For any matrix $M \in \mathbb{C}^{n\times m}$, we denote by $M_{i\cdot}$ and $M_{\cdot i}$ its ith row vector and ith column vector respectively. Moreover, $||M||_2$ denotes the $\ell^2 \to \ell^2$ induced norm, $||M||_\infty = \max_i ||M_{i\cdot}||_1$ denotes the $\ell^\infty \to \ell^\infty$ induced norm and $||M||_{2,\infty} = \max_i ||M_{i\cdot}||_2$ is the $\ell^2 \to \ell^\infty$ induced norm. We denote by $\mathbb{1}_n$ the vector of length n with all entries being 1 and let $J_{n\times m} = \mathbb{1}_n \mathbb{1}_m^\top$ be the $n \times m$ matrix of all ones. If S is a subset of the vertex set V, then $\mathbb{1}_S$ is the indicator vector such that $(\mathbb{1}_S)_i = 1$ if $v_i \in S$ and $(\mathbb{1}_S)_i = 0$ if $v_i \notin S$. If $M \in \mathbb{C}^{n\times n}$ is self-adjoint, then we arrange its eigenvalues in increasing order:

$$\lambda_1(M) \le \lambda_2(M) \le \cdots \le \lambda_n(M).$$

2. Main results.

2.1. Certifying the global minimum of ratio cut.

Suppose the partition $\{V_i\}_{i=1}^k$ achieves the minimum ratio cut. If we see each V_i as a planted cluster, then the connectivity within each cluster should be strong and the connections between them should be weak. To quantify this, let $L_i \in \mathbb{R}^{|V_i|\times|V_i|}$ be the graph Laplacian of the induced subgraph $G[V_i]$. We measure the connectivity of $G[V_i]$ by $\lambda_2(L_i)$, which is the second smallest eigenvalue of L_i. The second smallest eigenvalue of a graph Laplacian is also called the algebraic connectivity of the graph. The larger it is, the stronger the graph is connected. In the case the graph is disconnected, the algebraic connectivity drops to 0. One way to interpret the algebraic connectivity is that it provides a lower bound for the edge density of the graph (see Lemma 2.1 below). The proof of this result and subsequent results will be presented in Section 4.

Lemma 2.1. *Let G be a weighted undirected graph with vertex set V. Let L be the graph Laplacian of G. Let S be a subset of V. Then*

$$Cut(S, V-S) \geq \lambda_2(L) \frac{|S| \cdot |V-S|}{|V|}.$$

To measure the inter-cluster connectivity, we define for each vertex v_i,

$$d_\delta^{(i)} = \sum_{v_k \in V_j^c} w_{ik}$$

where V_j is the cluster that contains v_i. In other words, $d_\delta^{(i)}$ is the total weight between v_i and outside clusters. With such definitions for intra- and inter-cluster connectivity, we are able to certify when a partition is optimal.

Theorem 2.2. *Suppose a partition $\{V_i\}_{i=1}^k$ satisfies*

$$\max_{1\leq i\leq n} d_\delta^{(i)} \leq \frac{1}{2} \min_{1\leq i\leq k} \lambda_2(L_i), \tag{2.1}$$

then $\{V_i\}_{i=1}^k$ achieves the minimum ratio cut among all k-way partitions of V. If (2.1) holds with the strict inequality, then $\{V_i\}_{i=1}^k$ is also the unique partition (up to relabeling) that achieves the minimum ratio cut.

Theorem 2.2 is a direct improvement of the result in [16], which has a constant $\frac{1}{4}$ instead of $\frac{1}{2}$. The following example shows that the constant $\frac{1}{2}$ cannot be further improved.

Example 2.3. Let $W \in \mathbb{R}^{4n\times 4n}$,

$$W = \begin{pmatrix} J_{n\times n} & J_{n\times n} & cJ_{n\times n} & 0 \\ J_{n\times n} & J_{n\times n} & 0 & cJ_{n\times n} \\ cJ_{n\times n} & 0 & J_{n\times n} & J_{n\times n} \\ 0 & cJ_{n\times n} & J_{n\times n} & J_{n\times n} \end{pmatrix} - I_{4n}.$$

Consider the partition $V_1 = \{v_1, \cdots, v_{2n}\}$, $V_2 = \{v_{2n+1}, \cdots, v_{4n}\}$. The corresponding

$$\min_{1\leq i\leq 2} \lambda_2(L_i) = 2n \quad \text{and} \quad \max_{1\leq i\leq 4n} d_\delta^{(i)} = cn.$$

If $c > 1$ then the condition in Theorem 2.2 is violated for this partition. One can check that in this case a different partition $V^{(1)} = \{v_1, \cdots, v_n, v_{2n+1}, \cdots, v_{3n}\}$, $V^{(2)} = V - V^{(1)}$ has a smaller ratio cut.

Theorem 2.2 is algorithm independent and can be useful in many ways. For example one can use it to check in polynomial time if a given partition is optimal. It can also serve as a benchmark for comparing different algorithms. In [16] the authors propose a SDP relaxation to the minimum ratio cut problem (1.1) and show that it is able to find the optimal partition if it satisfies $\max_{1\le i\le n} d_\delta^{(i)} < \frac{1}{4}\min_{1\le i\le k}\lambda_2(L_i)$. In this paper we prove that Algorithm 1.1 is able to find the optimal partition if $\max_{1\le i\le n} d_\delta^{(i)} \lesssim \frac{1}{\ln n}\min_{1\le i\le k}\lambda_2(L_i)$. The notation "$\lesssim$" hides a term that does not depend on n.

2.2. A two-to-infinity norm bound for the Laplacian eigenmap. Algorithm 1.1 can be understood from a perturbation perspective. Suppose we try to recover the planted partition $\{V_i\}_{i=1}^k$. Let W_i, D_i, L_i denote the weighted adjacency matrix, degree matrix and graph Laplacian of the induced subgraph $G[V_i]$ respectively. Let

$$W_{\text{iso}} = \begin{pmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_k \end{pmatrix}, W_\delta = W - W_{\text{iso}}.$$

Let D_{iso}, L_{iso}, D_δ, L_δ be the corresponding degree matrices or graph Laplacians (here we suppose $\lambda_{k+1}(L_{\text{iso}}) > 0$). Let U (U_{iso}) be a matrix with orthonormal columns whose range is the invariant subspace of L (L_{iso}) that corresponds to the k smallest eigenvalues. Then the k smallest eigenvalues of L_{iso} are 0 and U_{iso}, up to a multiplication of orthogonal matrix from the right, is

$$U_{\text{iso}} = \begin{pmatrix} \frac{1}{\sqrt{|V_1|}}\mathbb{1}_{V_1} & \frac{1}{\sqrt{|V_2|}}\mathbb{1}_{V_2} & \cdots & \frac{1}{\sqrt{|V_k|}}\mathbb{1}_{V_k} \end{pmatrix}.$$

Hence the rows of U_{iso} reduce to k different points in $\mathbb{R}^k$ with one cluster at each point. Any rounding method will recover the planted clusters perfectly. Here we also point out that a multiplication of orthogonal matrix from the right transforms all the rows simultaneously and thus preserves the geometry of the embedding. If W_δ is small, then U should be close to U_{iso}. A reasonable measurement for the closeness is

$$\min_{V\in\mathbf{O}^k} ||UV - U_{\text{iso}}||_{2,\infty} = \min_{V\in\mathbf{O}^k}\max_{1\le i\le n} ||(UV - U_{\text{iso}})_{i\cdot}||_2 ,$$

where the minimization is taken over all $k\times k$ orthogonal matrices. This error measures the maximum distance of a point $U_{i\cdot}$ away from its origin $(U_{\text{iso}})_{i\cdot}$ after some global orthogonal transformation. If this error is small enough then the rounding step should be able to recover the planted clusters perfectly. We present our bound for $\min_{V\in\mathbf{O}^k}||UV - U_{\text{iso}}||_{2,\infty}$ in Theorem 2.4 below. The result is stated in terms of $||U\tilde{V} - U_{\text{iso}}||_{2,\infty}$ where $\tilde{V}$ solves the orthogonal Procrustes problem

$$\tilde{V} = \arg\min_{V\in\mathbf{O}^k} ||UV - U_{\text{iso}}||_F .$$

Note that $\min_{V\in\mathbf{O}^k}||UV - U_{\text{iso}}||_{2,\infty} \le ||U\tilde{V} - U_{\text{iso}}||_{2,\infty}$. The matrices V and $\tilde{V}$ are only defined for the sake of analysis and are not required by the actual algorithm.

Theorem 2.4. *Suppose each* $|V_i| \geq 3$. *Let*

$$c = \max_{1\leq i\leq k} \frac{n}{|V_i|} \quad \textit{and} \quad r = \frac{\max_{1\leq i\leq n} d_\delta^{(i)}}{\min_{1\leq i\leq k} \lambda_2(L_i)}$$

be the unbalanceness and the perturbation/eigengap ratio respectively. If $r \leq \frac{1}{16(1+c)\ln n}$, *then*

$$\left|\left|U\tilde{V} - U_{iso}\right|\right|_{2,\infty} \leq 32\sqrt{c}\left(r^2 + r\ln n\right)\frac{1}{\sqrt{n}}.$$

The rest of this section is dedicated to the technical details of the proof. Discussions and applications regarding this bound are deferred to Section 3. The tool we use is Corollary 3.3 in [7] which gives a two-to-infinity norm perturbation bound for the invariant subspace. We cite this result in Lemma 2.5 below. The definition of the separation of two matrices (denoted by *sep*) that arises in Lemma 2.5, is stated below the lemma.

Lemma 2.5. *Let* $L_{iso} = U_{iso}\Lambda_1 U_{iso}^T + U_2\Lambda_2 U_2^T$ *be the spectral decomposition of* L_{iso} *where* $\Lambda_1 \in \mathbb{R}^{k\times k}$ *is a zero matrix and* $\Lambda_2 \in \mathbb{R}^{(n-k)\times(n-k)}$ *whose diagonal contains all the positive eigenvalues of* L_{iso}. *Let* $\operatorname{gap} = \min\left\{\operatorname{sep}_2(\Lambda_1,\Lambda_2), \operatorname{sep}_{(2,\infty),U_2}(\Lambda_1, U_2\Lambda_2U_2^T)\right\}$ *and* $\mu = \sqrt{n}\,||U_{iso}||_{2,\infty}$. *If* $||L_\delta||_2 \leq \frac{\operatorname{gap}}{5}$ *and* $||L_\delta||_\infty \leq \operatorname{gap}/(4+4\mu^2)$ *then*

$$\left|\left|U\tilde{V} - U_{iso}\right|\right|_{2,\infty} \leq 8\,||U_{iso}||_{2,\infty}\left(\frac{||L_\delta||_2}{\operatorname{sep}_2(\Lambda_1,\Lambda_2)}\right)^2 + 4\frac{\left|\left|U_2U_2^TL_\delta U_{iso}\right|\right|_{2,\infty}}{\operatorname{gap}}.$$

Classical perturbation theory like Davis-Kahan usually bounds the invariant subspace perturbation in terms of the perturbation/eigengap ratio. Lemma 2.5 is similar but with the classical eigengap replaced by the *gap* term defined therein. Here the separation of two matrices is defined as

$$\operatorname{sep}_{*,W}(B,C) = \inf\left\{||ZB - CZ||_* : Z \in \mathbb{R}^{m\times l}, \operatorname{ran} Z \subseteq \operatorname{ran} W, ||Z||_* = 1\right\}$$

where $B \in \mathbb{R}^{l\times l}$, $C \in \mathbb{R}^{m\times m}$, $\operatorname{ran} W$ is a linear subspace of $\mathbb{R}^m$ and $||\cdot||_*$ is a norm on $\mathbb{R}^{m\times l}$. When $\operatorname{ran} W = \mathbb{R}^m$ we denote $\operatorname{sep}_*(B,C) = \operatorname{sep}_{*,W}(B,C)$. $||L_\delta||_2$ and $||L_\delta||_\infty$ in Lemma 2.5 can be bounded by $2\max_{1\leq i\leq n} d_\delta^{(i)}$. Moreover, the two matrix separation terms $\operatorname{sep}_2(\Lambda_1,\Lambda_2)$ and $\operatorname{sep}_{(2,\infty),U_2}(\Lambda_1, U_2\Lambda_2U_2^T)$ are closely related to the eigengap $\min_{1\leq i\leq k}\lambda_2(L_i)$. In fact,

$$\begin{aligned}\operatorname{sep}_2(\Lambda_1,\Lambda_2) &= \inf\left\{||Z\Lambda_1 - \Lambda_2 Z||_2 : Z \in \mathbb{R}^{(n-k)\times k}, ||Z||_2 = 1\right\}\\ &= \inf\left\{||\Lambda_2 Z||_2 : Z \in \mathbb{R}^{(n-k)\times k}, ||Z||_2 = 1\right\}\\ &= \min_{1\leq i\leq k}\lambda_2(L_i) \qquad (2.2)\end{aligned}$$

is exactly the eigengap. Furthermore,

$$\begin{aligned}\operatorname{sep}_{(2,\infty),U_2}(\Lambda_1, U_2\Lambda_2U_2^T) &= \inf\left\{||0 - L_{\text{iso}}Z||_{2,\infty} : Z \in \mathbb{R}^{n\times k}, \operatorname{ran} Z \subseteq \operatorname{ran} U_2, ||Z||_{2,\infty} = 1\right\}\\ &= \min_{1\leq i\leq k}\inf\left\{||L_iZ||_{2,\infty} : Z \in \mathbb{R}^{|V_i|\times k}, \operatorname{ran} Z \subseteq \left\{\mathbb{1}_{|V_i|}\right\}^\perp, ||Z||_{2,\infty} = 1\right\}\\ &= \min_{1\leq i\leq k}\inf_{x\perp\mathbb{1}_{|V_i|}}\frac{||L_ix||_\infty}{||x||_\infty}\end{aligned}$$

can be understood as the "eigengap" in terms of the ℓ^∞ norm. The third equality holds because for any Z if we let $x \neq 0$ be the vector that $||Zx||_\infty = ||x||_2$, then

$$||L_i Z||_{2,\infty} \geq \frac{||L_i Z x||_\infty}{||x||_2} = \frac{||L_i Z x||_\infty}{||Zx||_\infty} \geq \inf_{x \perp \mathbb{1}_{|V_i|}} \frac{||L_i x||_\infty}{||x||_\infty}.$$

And on the other hand we can pick a $\tilde{Z}$ such that its first column satisfies

$$\left|\left|\tilde{Z}_{\cdot 1}\right|\right|_\infty = 1\ , \quad \left|\left|L_i \tilde{Z}_{\cdot 1}\right|\right|_\infty = \inf_{x \perp \mathbb{1}_{|V_i|}} \frac{||L_i x||_\infty}{||x||_\infty}$$

and its other columns are 0 so that

$$\left|\left|L_i \tilde{Z}\right|\right|_{2,\infty} = \inf_{x \perp \mathbb{1}_{|V_i|}} \frac{||L_i x||_\infty}{||x||_\infty}.$$

Note that we always have

$$\inf_{x \perp \mathbb{1}_{|V_i|}} \frac{||L_i x||_\infty}{||x||_\infty} \leq \lambda_2(L_i).$$

Therefore the the *gap* term in Lemma 2.5 is simplified to

$$\text{gap} = \min_{1 \leq i \leq k} \inf_{x \perp \mathbb{1}_{|V_i|}} \frac{||L_i x||_\infty}{||x||_\infty}. \tag{2.3}$$

There is a trivial bound that relates *gap* to the eigengap:

$$\inf_{x \perp \mathbb{1}_{|V_i|}} \frac{||L_i x||_\infty}{||x||_\infty} \geq \frac{\lambda_2(L_i)}{\sqrt{|V_i|}}.$$

But we will show that due to the diagonally dominant structure of the graph Laplacian, the $\sqrt{|V_i|}$ factor can be improved to $\ln |V_i|$. The following theorem, which is also of independent interest, is essential in this context.

Theorem 2.6. *Let B be a self-adjoint $n \times n$ matrix, $n \geq 3$ such that $B_{i,i} \geq \sum_{j \in \{1,\ldots,n\} \setminus \{i\}} |B_{i,j}|$ for all $1 \leq i \leq n$. Let $\mathcal{M}$ be a subspace of $\mathbb{C}^n$ such that $B\mathcal{M} \subset \mathcal{M}$. Then*

$$\|Bx\|_\infty \geq \frac{\lambda_{\min}(B|_{\mathcal{M}})\|x\|_\infty}{2 \ln n},$$

for all $x \in \mathcal{M}$.

Corollary 2.7. *Suppose that L is the Laplacian of a graph with n vertices, $n \geq 3$. Then*

$$\frac{\lambda_2(L)}{2 \ln n} \leq \inf_{x \perp \mathbb{1}_n} \frac{\|Lx\|_\infty}{\|x\|_\infty} \leq \frac{4M}{D},$$

where the second inequality holds for unweighted graphs with M being the maximum degree and D being the diameter.

The following example shows that the $\ln n$ factor in Corollary 2.7 is necessary. However, at this point we do not know whether it must carry over to $||U\tilde{V} - U_{\text{iso}}||_{2,\infty}$ as well.

Example 2.8. Suppose that L is the Laplacian of a d-regular Ramanujan graph with n vertices. This means that $\lambda_2(L) \geq d - 2\sqrt{d-1}$. Note that $n \leq (d+1)^D$ where D is the diameter of the graph. This follows from the fact that for a fixed vertex u_0, every vertex can be connected to u_0 via a path of length at most D and that there are at most $1+d+d^2+\ldots+d^D \leq (d+1)^D$ paths of length at most D. Thus, $D \geq \frac{\ln n}{\ln(d+1)}$. By Corollary 2.7,

$$\frac{d - 2\sqrt{d-1}}{2\ln n} \leq \inf_{x \perp \mathbb{1}_n} \frac{\|Lx\|_\infty}{\|x\|_\infty} \leq \frac{4d}{D} \leq \frac{4d\ln(d+1)}{\ln n}.$$

For example, if L is the Laplacian of a 5-regular Ramanujan graph with n vertices, then

$$\frac{1}{2\ln n} \leq \inf_{x \perp \mathbb{1}_n} \frac{\|Lx\|_\infty}{\|x\|_\infty} \leq \frac{36}{\ln n}.$$

For every $d \geq 3$, there exist infinitely many d-regular Ramanujan graphs by [19]. This shows that the $\ln n$ factor in Corollary 2.7 is necessary.

3. Discussions.

Algorithm 1.1 consists of two steps: the Laplacian eigenmap and a rounding step. We have bounded the Laplacian eigenmap in Theorem 2.4. The next question is whether the rounding step will successfully recover the planted clusters based on the embedded points in $\mathbb{R}^k$. The answer depends on our understanding of the choice of the rounding method and it is beyond the scope of this paper to present a survey on this subject. But since the rows of U_{iso} have a natural magnitude of $O(1/\sqrt{n})$, if we have $\left|\left|U\tilde{V} - U_{\text{iso}}\right|\right|_{2,\infty} < C/\sqrt{n}$ for some sufficiently small C, then it means the rows of U (after some proper global rotation) are close enough to the ideal U_{iso} and thus should be nicely separable. Indeed, we will show through several examples that the condition

$$\left|\left|U\tilde{V} - U_{\text{iso}}\right|\right|_{2,\infty} < \frac{C}{\sqrt{n}} \tag{3.1}$$

can imply successful recovery, where C depends on the specific choice of the rounding method and (possibly) the number of clusters k.

- **A simple bisector for two clusters.** When $k = 2$, the Fiedler eigenvector (i.e., the eigenvector $u_2(L)$ that corresponds to the second smallest eigenvalue of L) is a popular tool to partition the graph. One way to do this is to first put the entries of the Fiedler eigenvector in algebraic order. Then out of all $n-1$ possible linear bisections of the entries we pick the one that gives the smallest ratio cut. This method is equivalent to finding the best linear bisection of the embedded points in $\mathbb{R}^2$. For this rounding method we can let $C = 1$ in (3.1). To see why, first note that $\tilde{V}$ is the solution to an orthogonal Procrustes problem and therefore has a closed form solution

$$\tilde{V} = V_1 V_2^T$$

where $U^T U_{\text{iso}} = V_1 \Sigma V_2^T$ is the singular value decomposition of $U^T U_{\text{iso}}$. Given that $u_1(L) = u_1(L_{\text{iso}}) = \frac{1}{\sqrt{n}} \mathbb{1}_n$, it is easy to check that

$$\left|\left|U\tilde{V} - U_{\text{iso}}\right|\right|_{2,\infty} = ||u_2(L) - u_2(L_{\text{iso}})||_{\infty}$$

where the sign of $u_2(L)$ is chosen so that $\langle u_2(L_{\text{iso}}), u_2(L)\rangle > 0$. Note that the distance between the two embedded unperturbed clusters is $\sqrt{1/|V_1| + 1/|V_2|}$. To ensure the separation of the two clusters in $u_2(L)$ we require

$$||u_2(L) - u_2(L_{\text{iso}})||_{\infty} < \frac{1}{2}\sqrt{\frac{1}{|V_1|} + \frac{1}{|V_2|}},$$

which is guaranteed by $\left|\left|U\tilde{V} - U_{\text{iso}}\right|\right|_{2,\infty} < 1/\sqrt{n}$.

- **An SDP type of k-means algorithm.** The k-means algorithms are a family of algorithms that seek the k-way partition $\{\Gamma_i\}_{i=1}^k$ of n points in $\mathbb{R}^d$ by minimizing the following k-means objective function:

$$\min_{\{\Gamma_i\}_{i=1}^k} \sum_{i=1}^k \sum_{x \in \Gamma_i} ||x - \mu_i||_2^2$$

where μ_i is the mean of points in Γ_i. Note that the optimization is shown to be NP-hard ([18, 2]) so no polynomial-time algorithm is guaranteed to find the optimal partition in general. The most famous and widely used k-means algorithm is the Lloyd's algorithm ([17]). But its heuristic nature and random initial starts make the analysis of exact recovery difficult. Here we consider a SDP type of k-means algorithm proposed in [15]. The proposed algorithm comes with a proximity condition for the planted partition under which the algorithm is guaranteed to recover the planted partition. Let $n_i = |\Gamma_i|$ and $X_i \in \mathbb{R}^{n_i \times d}$ be the data matrix of the i-th cluster with each row being a point in Γ_i. Let $\overline{X}_i = X_i - \mathbb{1}_{n_i}\mu_i^T$ be the centered data matrix of the i-th cluster. For each pair of $i \neq j$, let $M_{i,j}$ denote the bisecting hyperplane that passes through $\frac{\mu_i + \mu_j}{2}$ and is perpendicular to the line segment that joins μ_i and μ_j. The proximity condition is then stated as follows: for all $i \neq j$, (i) Γ_i and Γ_j are separated by $M_{i,j}$ and (ii)

$$\xi_{i,j} > \frac{1}{2}\sqrt{\sum_{i=1}^k ||\overline{X}_i||_2^2 \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

where $\xi_{i,j} = \text{dist}\{M_{i,j}, \Gamma_i \cup \Gamma_j\}$ is the margin between the clusters and the bisecting hyperplane (see also Figure 1). We now claim that if $C = 1/5$ in (3.1) then the SDP k-means algorithm is guaranteed to recover the planted partition. To see why, note that (3.1) implies the points in Γ_i and Γ_j, together with their means, are confined within two balls of radius $C/\sqrt{n}$ whose centers are $\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$ apart. By Lemma 4.3 in Section 4.3 we have

$$\xi_{i,j} > \frac{1}{2}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} - \frac{3C}{\sqrt{n}} \geq \frac{1 - 3C}{2}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

And on the other hand

$$\sum_{i=1}^{k} ||\overline{X}_i||_2^2 \leq \sum_{i=1}^{k} ||\overline{X}_i||_F^2 < 4C^2.$$

Therefore, if $C \leq 1/5$ then the proximity condition is satisfied.

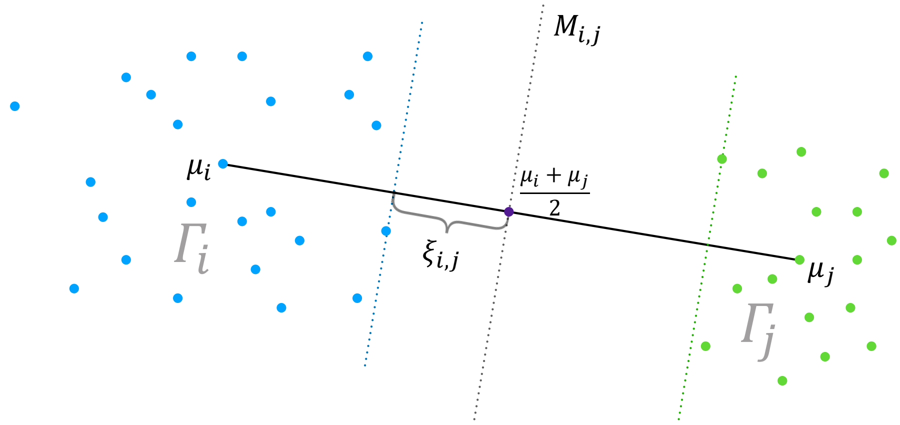


Figure 1: Proximity condition for the SDP k-means algorithm. If the partition satisfies the proximity condition, then each pair of clusters Γ_i and Γ_j are separated by and sufficiently bounded away from the bisecting hyperplane of the line segment that joins μ_i and μ_j.

- **Two projective k-means algorithms.** In [12] and [3] two projective k-means algorithms are proposed which consist of an SVD-based projection followed by iterative Lloyd steps with informed initial starts. By our notation, the algorithm in [12] is guaranteed to recover the planted partition if for any $i \neq j$,

$$\xi_{i,j} > \tilde{C}k\left(\frac{1}{\sqrt{n_i}} + \frac{1}{\sqrt{n_j}}\right)||\overline{W}||_2$$

where $\tilde{C} > 0$ is an absolute constant and $\overline{W} = [\overline{X}_1^T, \cdots, \overline{X}_k^T]^T$. The algorithm in [3] is guaranteed to recover the planted partition if for any $i \neq j$,

$$\xi_{i,j} > \frac{1}{2}\tilde{C}\left(\frac{1}{\sqrt{n_i}} + \frac{1}{\sqrt{n_j}}\right)||\overline{W}||_2$$

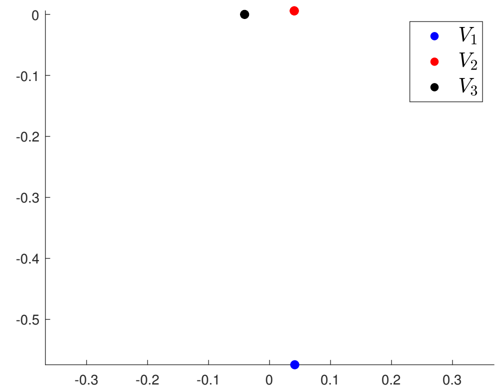
and

$$||\mu_i - \mu_j||_2 > \tilde{C}\sqrt{k}\left(\frac{1}{\sqrt{n_i}} + \frac{1}{\sqrt{n_j}}\right)||\overline{W}||_2.$$

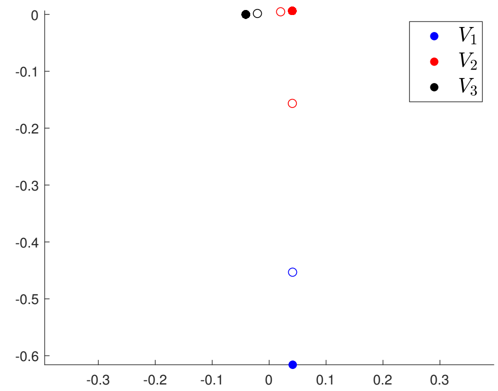
Then C in (3.1) can be similarly derived for both methods.

Thus, by Theorem 2.4 and the discussion above, Algorithm 1.1 finds the planted partition if $r \lesssim 1/\ln n$ where the hidden term depends on k and the unbalanceness term c. By Theorem 2.2 when $r \leq 1/2$, we can certify that the planted partition is optimal. Therefore we may claim that Algorithm 1.1 finds the optimal partition when $r \lesssim 1/\ln n$.

Note that when the unbalanceness term c gets arbitrarily large, Theorem 2.4 gets arbitrarily bad. This is to be expected. We illustrate the effect of unbalanceness on the Laplacian eigenmap using a simple numerical example. Consider $|V_1| = 3$ and $|V_2| = |V_3| = 300$. Let $W_i = J_{|V_i|\times|V_i|}/|V_i|$ so $\lambda_2(L_i) = 1$ for $1 \le i \le 3$. We perturb the graph by adding a weight 0.5 between a vertex in V_1 and a vertex in V_2. We then add another weight 0.5 between a vertex in V_3 and another vertex in V_2. The Laplacian eigenmap of both the unperturbed and the perturbed graphs are shown in Figure 2. As seen from the figure, $\min_{V\in\mathbf{O}^k} ||UV - U_{\text{iso}}||_{2,\infty}$ is large and will be even more so as the clusters get more unbalanced. Despite the error being large, the Laplacian eigenmap still separates the three clusters well. The reason is that $\min_{V\in\mathbf{O}^k} ||UV - U_{\text{iso}}||_{2,\infty}$ only measures the magnitude and thus ignores the directions of the perturbation. A more refined analysis on the Laplacian eigenmap should consider the direction of the perturbation as well as the magnitude. It remains an open problem whether there exists a constant C such that $r \le C$ implies successful recovery of the planted clusters by Algorithm 1.1. The constant C should only depend on k or, better yet, not even on k.



(a) Unperturbed graph



(b) Perturbed graph

Figure 2: Laplacian eigenmap for both the unperturbed and the perturbed graphs. The embedded points live on a common plane in $\mathbb{R}^3$ so we can visualize them in $\mathbb{R}^2$. The four empty circles are the four perturbed vertices.

4. Proofs.

4.1. Proofs for Section 2.1.

Proof of Lemma 2.1. Consider $M = L - \lambda_2(L)P$ where $P = I - \frac{1}{|V|}J_{|V|\times|V|}$. Then M is positive semi-definite. Therefore

$$\mathbb{1}_S^T M \mathbb{1}_S = \mathbb{1}_S^T L \mathbb{1}_S - \lambda_2(L)\mathbb{1}_S^T P \mathbb{1}_S = \text{Cut}\,(S, V-S) - \lambda_2(L)\frac{|S|\cdot|V-S|}{|V|} \ge 0 \quad \blacksquare$$

Proof of Theorem 2.2. Suppose without loss of generality that the partition $\{V_i\}_{i=1}^k$ satisfies

$$\max_{1\le i\le n} d_\delta^{(i)} \le \frac{1}{2} \quad \text{and} \quad \min_{1\le i\le k} \lambda_2(L_i) \ge 1.$$

Let $\left\{V^{(j)}\right\}_{j=1}^k$ be another partition of V. We aim to show

$$\text{RatioCut}\left(\{V_i\}_{i=1}^k\right) \le \text{RatioCut}\left(\left\{V^{(j)}\right\}_{j=1}^k\right).$$

Let $n_i = |V_i|$, $n^{(j)} = |V^{(j)}|$ and $m_i^{(j)} = |V_i \cap V^{(j)}|$. We have

$$\sum_i m_i^{(j)} = n^{(j)}, \quad \sum_j m_i^{(j)} = n_i.$$

Let

$$W_{i,i'}^{(j,j')} = \sum_{v_k\in V_i\cap V^{(j)}, v_l \in V_{i'}\cap V^{(j')}} w_{kl}.$$

Thus we have divided the weighted adjacency matrix W into $k^2 \times k^2$ rectangular areas and $W_{i,i'}^{(j,j')}$ is the total weight in one of the areas. Lemma 2.1 gives

$$\sum_{j'\neq j} W_{i,i}^{(j,j')} \ge \min_{1\le a\le k} \lambda_2(L_a) \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i} \ge \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i} \tag{4.1}$$

for all i, j. We also know each $d_\delta^{(i)}$ is at most $1/2$. This implies

$$\sum_{i'\neq i} W_{i,i'}^{(j,j)} + \sum_{j'\neq j}\sum_{i'\neq i} W_{i,i'}^{(j,j')} = \sum_{j'}\sum_{i'\neq i} W_{i,i'}^{(j,j')} \le \frac{1}{2} m_i^{(j)} \tag{4.2}$$

for all i, j. Moreover

$$\sum_{i'\neq i} W_{i,i'}^{(j,j)} \le \frac{1}{2}\min\{m_i^{(j)}, n^{(j)} - m_i^{(j)}\} \tag{4.3}$$

for all i, j because the summands together represent a rectangular area in W with length $m_i^{(j)}$ and width $n^{(j)} - m_i^{(j)}$. Therefore we need to show

$$\begin{aligned}
&\text{RatioCut}\left(\{V_i\}_{i=1}^k\right) - \text{RatioCut}\left(\left\{V^{(j)}\right\}_{j=1}^k\right)\\
=&\sum_j\sum_{j'}\sum_i\sum_{i'\neq i} \frac{1}{n_i} W_{i,i'}^{(j,j')} - \sum_i\sum_{i'}\sum_j\sum_{j'\neq j} \frac{1}{n^{(j)}} W_{i,i'}^{(j,j')}\\
=&\sum_i\sum_j\sum_{i'\neq i}\frac{1}{n_i}W_{i,i'}^{(j,j)} + \sum_i\sum_j\left(\frac{1}{n_i} - \frac{1}{n^{(j)}}\right)\left(\sum_{j'\neq j}\sum_{i'\neq i} W_{i,i'}^{(j,j')}\right) - \sum_i\sum_j\sum_{j'\neq j}\frac{1}{n^{(j)}} W_{i,i}^{(j,j')}\\
=&A_1 + A_2 - A_3 \le 0.
\end{aligned}$$

For A_1,

$$
\begin{aligned}
A_1 =& \sum_i \sum_j \sum_{i' \neq i} \frac{1}{n_i} W_{i,i'}^{(j,j)} \\
=& \sum_i \sum_j \sum_{i' \neq i} \frac{1}{n_i} W_{i,i'}^{(j,j)} - \frac{1}{2} \sum_i \sum_j \frac{1}{n_i} \min\{m_i^{(j)}, n^{(j)} - m_i^{(j)}\} \\
& + \frac{1}{2} \sum_i \sum_j \frac{1}{n_i} \min\{m_i^{(j)}, n^{(j)} - m_i^{(j)}\} \\
=& A_{11} - A_{12} + A_{13}.
\end{aligned}
$$

For A_2, when $\frac{1}{n_i} - \frac{1}{n^{(j)}} \leq 0$, the summand is upper bounded by 0. When $\frac{1}{n_i} - \frac{1}{n^{(j)}} > 0$, the summand is upper bounded by (by using (4.2))

$$
\left(\frac{1}{n_i} - \frac{1}{n^{(j)}} \right) \left(\sum_{j' \neq j} \sum_{i' \neq i} W_{i,i'}^{(j,j')} \right) \leq \left(\frac{1}{n_i} - \frac{1}{n^{(j)}} \right) \left(\frac{1}{2} m_i^{(j)} - \sum_{i' \neq i} W_{i,i'}^{(j,j)} \right).
$$

Therefore we can bound A_2 by

$$
\begin{aligned}
A_2 &= \sum_i \sum_j \left(\frac{1}{n_i} - \frac{1}{n^{(j)}} \right) \left(\sum_{j' \neq j} \sum_{i' \neq i} W_{i,i'}^{(j,j')} \right) \\
&\leq \sum_i \sum_j \max \left\{ \frac{1}{n_i} - \frac{1}{n^{(j)}}, 0 \right\} \left(\frac{1}{2} m_i^{(j)} - \sum_{i' \neq i} W_{i,i'}^{(j,j)} \right) \\
&= \sum_i \sum_j \left(\frac{1}{n_i} - \min \left\{ \frac{1}{n^{(j)}}, \frac{1}{n_i} \right\} \right) \left(\frac{1}{2} m_i^{(j)} - \sum_{i' \neq i} W_{i,i'}^{(j,j)} \right) \\
&\leq \frac{1}{2} \sum_i \sum_j \frac{1}{n_i} m_i^{(j)} - \sum_i \sum_j \sum_{i' \neq i} \frac{1}{n_i} W_{i,i'}^{(j,j)} - \frac{1}{2} \sum_i \sum_j \min \left\{ \frac{1}{n^{(j)}}, \frac{1}{n_i} \right\} \max \left\{ 0, 2m_i^{(j)} - n^{(j)} \right\} \\
&= A_{21} - A_{22} - A_{23},
\end{aligned}
$$

where in the last step we used (4.3). For A_3, we use (4.1):

$$
\begin{aligned}
A_3 = \sum_i \sum_j \sum_{j' \neq j} \frac{1}{n^{(j)}} W_{i,i}^{(j,j')} &\geq \sum_i \sum_j \min_{1 \leq a \leq k} \lambda_2(L_a) \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i n^{(j)}} \\
&\geq \sum_i \sum_j \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i n^{(j)}}.
\end{aligned} \tag{4.4}
$$

Therefore (here we introduce the shorthand notation min* for $\min\{m_i^{(j)}, n^{(j)} - m_i^{(j)}\}$ and max* for $\max\{m_i^{(j)}, n^{(j)} - m_i^{(j)}\}$)

$$
\begin{aligned}
A_{12} - A_3 &\leq \sum_i \sum_j \left(\frac{1}{2n_i} \min\{m_i^{(j)}, n^{(j)} - m_i^{(j)}\} - \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i n^{(j)}} \right) \\
&= \sum_i \sum_j \left(\frac{\min^*(\min^* + \max^*)}{2n_i n^{(j)}} - \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i n^{(j)}} \right) \\
&= \sum_i \sum_j \left(\frac{\min^* \max^*}{n_i n^{(j)}} - \frac{\min^*(\max^* - \min^*)}{2n_i n^{(j)}} - \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i n^{(j)}} \right) \\
&= \sum_i \sum_j \left(\frac{m_i^{(j)}(n^{(j)} - m_i^{(j)})}{n_i n^{(j)}} - \frac{m_i^{(j)}(n_i - m_i^{(j)})}{n_i n^{(j)}} \right) - \frac{1}{2} \sum_i \sum_j \frac{\min^*(\max^* - \min^*)}{n_i n^{(j)}} \\
&= \sum_i \sum_j \left(\frac{m_i^{(j)}(n^{(j)} - n_i)}{n_i n^{(j)}} \right) - A_4 \\
&= \sum_i \sum_j \frac{m_i^{(j)}}{n_i} - \sum_j \sum_i \frac{m_i^{(j)}}{n^{(j)}} - A_4 \\
&= k - k - A_4 = -A_4.
\end{aligned}
$$

The cancellation above is the reason why the constant in the condition (2.1) is 1/2. We also have $A_{11} - A_{22} = 0$. Finally

$$
\begin{aligned}
& A_1 + A_2 - A_3 \\
\leq & A_{21} - A_{12} - A_4 - A_{23} \\
= & \frac{1}{2} \sum_i \sum_j \left(\frac{m_i^{(j)}}{n_i} - \frac{\min^*}{n_i} - \frac{\min^*(\max^* - \min^*)}{n_i n^{(j)}} - \min\left\{\frac{1}{n^{(j)}}, \frac{1}{n_i}\right\} \max\left\{0, 2m_i^{(j)} - n^{(j)}\right\} \right) \\
= & \frac{1}{2} \sum_i \sum_j \left(\frac{m_i^{(j)}}{n_i} - \frac{\min^*(\min^* + \max^*)}{n_i n^{(j)}} - \frac{\min^*(\max^* - \min^*)}{n_i n^{(j)}} - \min\left\{\frac{1}{n^{(j)}}, \frac{1}{n_i}\right\} \max\left\{0, 2m_i^{(j)} - n^{(j)}\right\} \right) \\
= & \frac{1}{2} \sum_i \sum_j \left(\frac{m_i^{(j)}}{n_i} - \frac{2\min^* \max^*}{n_i n^{(j)}} - \min\left\{\frac{1}{n^{(j)}}, \frac{1}{n_i}\right\} \max\left\{0, 2m_i^{(j)} - n^{(j)}\right\} \right) \\
= & \frac{1}{2} \sum_i \sum_j \left(\frac{m_i^{(j)} n^{(j)}}{n_i n^{(j)}} - \frac{2m_i^{(j)}(n^{(j)} - m_i^{(j)})}{n_i n^{(j)}} - \min\left\{\frac{1}{n^{(j)}}, \frac{1}{n_i}\right\} \max\left\{0, 2m_i^{(j)} - n^{(j)}\right\} \right) \\
= & \frac{1}{2} \sum_i \sum_j \left(\frac{m_i^{(j)}(2m_i^{(j)} - n^{(j)})}{n_i n^{(j)}} - \min\left\{\frac{1}{n^{(j)}}, \frac{1}{n_i}\right\} \max\left\{0, 2m_i^{(j)} - n^{(j)}\right\} \right) \\
\leq & 0.
\end{aligned}
$$

∎

The last step is because for each summand,

$$\frac{m_i^{(j)}}{n_i n^{(j)}} \leq \min\left\{\frac{1}{n^{(j)}}, \frac{1}{n_i}\right\}.$$

This concludes the proof that $\{V_i\}_{i=1}^k$ achieves the minimum ratio cut. To show that it is also the unique global minimum when the strict inequality holds, we assume $\{V_i\}_{i=1}^k$ satisfies

$$\max_{1\leq i\leq n} d_\delta^{(i)} \leq \frac{1}{2} \quad \text{and} \quad \min_{1\leq i\leq k} \lambda_2(L_i) > 1.$$

All claims above still hold but we will show (4.4) holds with the strict inequality. Since $\min_{1\leq i\leq k}\lambda_2(L_i) > 1$, the last inequality in (4.4) takes equal sign if and only if $m_i^{(j)}(n_i - m_i^{(j)}) = 0$ for all i, j. But this is impossible if $\left\{V^{(j)}\right\}_{j=1}^k$ is not a relabeling of $\{V_i\}_{i=1}^k$. Therefore if $\left\{V^{(j)}\right\}_{j=1}^k$ is not a relabeling of $\{V_i\}_{i=1}^k$, we have

$$A_1 + A_2 - A_3 < 0$$

which concludes the proof. ■

4.2. Proofs for Section 2.2. We need the following two lemmas to prove Theorem 2.6. For a linear transformation T on a finite dimensional vector space, $\lambda_{\max}(T)$ and $\lambda_{\min}(T)$ denote the largest and the smallest eigenvalue of T, respectively.

Lemma 4.1. *Let T be an $n\times n$ matrix such that $||T||_\infty \leq 1$. Then*

$$\|x - T^k x\|_\infty \leq k\|x - Tx\|_\infty,$$

for all $x\in\mathbb{C}^n$ and $k\in\mathbb{N}$.

Proof. We have

$$\begin{aligned}\|x - T^n x\|_\infty &\leq \|x - Tx\|_\infty + \|Tx - T^2x\|_\infty + \ldots + \|T^{k-1}x - T^k x\|_\infty \\ &= \|x - Tx\|_\infty + \|T(x - Tx)\|_\infty + \ldots + \|T^{k-1}(x - Tx)\|_\infty \leq k\|x - Tx\|_\infty,\end{aligned}$$

where the last inequality follows from $||T||_\infty \leq 1$. ■

Lemma 4.2. *Let T be a self-adjoint $n\times n$ matrix, $n\geq 3$, such that $||T||_\infty \leq 1$. Let $\mathcal{M}$ be a subspace of $\mathbb{C}^n$ such that $T\mathcal{M}\subset\mathcal{M}$. Then*

$$\|x - Tx\|_\infty \geq \frac{(1-\lambda_{\max}(T|_{\mathcal{M}}))\|x\|_\infty}{2\ln n},$$

for all $x\in\mathcal{M}$.

Proof. We may assume that T is positive semidefinite. Indeed, $\frac{I+T}{2}$ is positive semidefinite, and if the result holds with T being replaced by $\frac{I+T}{2}$, the result will hold for T.

Since T is positive semidefinite,

$$\|T^k x\|_2 \leq \lambda_{\max}(T|_{\mathcal{M}})^k\|x\|_2,$$

for all $x \in \mathcal{M}$ and $k \in \mathbb{N}$. So $\|T^k x\|_\infty \le \lambda_{\max}(T|_{\mathcal{M}})^k \sqrt{n}\|x\|_\infty$ and hence, by Lemma 4.1,

$$\|x\|_\infty - \lambda_{\max}(T|_{\mathcal{M}})^k \sqrt{n}\|x\|_\infty \le k\|x - Tx\|_\infty,$$

for all $x \in \mathcal{M}$ and $k \in \mathbb{N}$. If k is large enough so that $\lambda_{\max}(T|_{\mathcal{M}})^k\sqrt{n} \le \frac{1}{2}$, then $\|x - Tx\|_\infty \ge \frac{1}{2k}\|x\|_\infty$. Since $||T||_\infty \le 1$, we have $\lambda_{\max}(T|_M) \le 1$. Note that $\lambda \le e^{\lambda-1}$ for all $\lambda \le 1$. So $\lambda_{\max}(T|_{\mathcal{M}})^k\sqrt{n} \le e^{k(\lambda_{\max}(T|_{\mathcal{M}})-1)}\sqrt{n} \le \frac{1}{2}$ for $k \ge \frac{\ln(2\sqrt{n})}{1-\lambda_{\max}(T|_{\mathcal{M}})}$. Taking k to be the smallest integer larger than or equal to $\frac{\ln(2\sqrt{n})}{1-\lambda_{\max}(T|_{\mathcal{M}})}$, we obtain

$$\|x - Tx\|_\infty \ge \frac{1}{2k}\|x\|_\infty \ge \frac{(1-\lambda_{\max}(T|_{\mathcal{M}}))\|x\|_\infty}{2\ln n},$$

if $n \ge 30$. If $3 \le n \le 29$, then

$$\|x - Tx\|_\infty \ge \frac{1}{\sqrt{n}}\|x - Tx\|_2 \ge \frac{1}{\sqrt{n}}(1-\lambda_{\max}(T|_{\mathcal{M}}))\|x\|_2 \ge \frac{(1-\lambda_{\max}(T|_{\mathcal{M}}))\|x\|_\infty}{2\ln n},$$

for all $x \in \mathcal{M}$. ■

Proof of Theorem 2.6. Without loss of generality, we may assume that $B_{i,i} \le 1$ for all $1 \le i \le n$. For every $1 \le i \le n$,

$$\sum_{j=1}^{n} |(I-B)_{i,j}| = 1 - B_{i,i} + \sum_{j\in\{1,\ldots,n\}\setminus\{i\}} |B_{i,j}| \le 1 - B_{i,i} + B_{i,i} \le 1.$$

So $||I - B||_\infty \le 1$. By Lemma 4.2, we have

$$\|Bx\|_\infty = \|x - (I-B)x\|_\infty \ge \frac{(1-\lambda_{\max}((I-B)|_{\mathcal{M}}))\|x\|_\infty}{2\ln n} = \frac{\lambda_{\min}(B|_{\mathcal{M}})\|x\|_\infty}{2\ln n},$$

for all $x \in \mathcal{M}$. ■

Proof of Corollary 2.7. Since $L\mathbb{1}_n = 0$ and L is self-adjoint, $L\{\mathbb{1}_n\}^\perp \subset \{\mathbb{1}_n\}^\perp$. By Theorem 2.6,

$$\|Lx\|_\infty \ge \frac{\lambda_{\min}\left(L|_{\{\mathbb{1}_n\}^\perp}\right)\|x\|_\infty}{2\ln n} = \frac{\lambda_2(L)\|x\|_\infty}{2\ln n},$$

for all $x \perp \mathbb{1}_n$. This proves one inequality.

To prove the other inequality, pick a vertex u_0. Let $y \in \mathbb{C}^n$ be given by $y(v) = d(u_0, v)$, for vertices v, where d is the graph distance. Then

$$(Ly)(v) = \deg(v)d(u_0, v) - \sum_{w\in N(v)} d(u_0, w),$$

for all vertex v, where $N(v)$ is the set of all neighborhood vertices of v. Since $|d(u_0, v) - d(u_0, w)| \le 1$ for all $w \in N(v)$, we have $|(Ly)(v)| \le \deg(v)$ for all vertex v. So $\|Ly\|_\infty \le M$.

Let $z = y - (\frac{1}{n}\sum_v y(v))\mathbb{1}_n \in \mathbb{C}^n$, where the sum is over all vertices v. It is easy to see that there exists a vertex w such that $d(u_0, w) \ge \frac{D}{2}$, where D is the diameter. So

$|z(w) - z(u_0)| = |y(w) - y(u_0)| = |d(u_0, w) - 0| = d(u_0, w) \geq \frac{D}{2}$. Thus, $\|z\|_\infty \geq \frac{D}{4}$. Since $L\mathbb{1}_n = 0$, we have $\|Lz\|_\infty = \|Ly\|_\infty \leq M$. Therefore,

$$\inf_{x \perp \mathbb{1}_n} \frac{\|Lx\|_\infty}{\|x\|_\infty} \leq \frac{M}{D/4} = \frac{4M}{D}.$$

The result follows. ■

With the help of Corollary 2.7, we can prove Theorem 2.4.

Proof of Theorem 2.4. We use the same notation as Lemma 2.5. Note that $||U_{\text{iso}}||_{2,\infty} = \max_{1 \leq i \leq k} 1/\sqrt{|V_i|}$ so $\mu = \sqrt{c}$. By Corollary 2.7 and (2.3) we have

$$\text{gap} \geq \min_{1 \leq i \leq k} \frac{\lambda_2(L_i)}{2 \ln |V_i|} \geq \frac{\min_{1 \leq i \leq k} \lambda_2(L_i)}{2 \ln n}.$$

If

$$r = \frac{\max_{1 \leq i \leq n} d_\delta^{(i)}}{\min_{1 \leq i \leq k} \lambda_2(L_i)} \leq \frac{1}{16(1+c) \ln n},$$

then

$$||L_\delta||_\infty = 2 \max_{1 \leq i \leq n} d_\delta^{(i)} \leq \frac{\text{gap}}{4(1+\mu^2)}$$

and

$$||L_\delta||_2 \leq \sqrt{||L_\delta||_\infty \, ||L_\delta||_1} = ||L_\delta||_\infty = 2 \max_{1 \leq i \leq n} d_\delta^{(i)} \leq \frac{\text{gap}}{5}.$$

Therefore by Lemma 2.5 we have

$$\begin{aligned}
\min_{V \in \mathbf{O}^k} ||UV - U_{\text{iso}}||_{2,\infty} &\leq 8 \, ||U_{\text{iso}}||_{2,\infty} \left(\frac{||L_\delta||_2}{\text{sep}_2(\Lambda_1, \Lambda_2)} \right)^2 + 4 \frac{\left|\left|U_2 U_2^T L_\delta U_{\text{iso}}\right|\right|_{2,\infty}}{\text{gap}} \\
&\leq 8 \sqrt{\frac{c}{n}} \left(\frac{2 \max_{1 \leq i \leq n} d_\delta^{(i)}}{\min_{1 \leq i \leq k} \lambda_2(L_i)} \right)^2 + \frac{8 \ln n \left|\left|U_2 U_2^T L_\delta U_{\text{iso}}\right|\right|_{2,\infty}}{\min_{1 \leq i \leq k} \lambda_2(L_i)} \\
&= 32 \sqrt{c} r^2 \frac{1}{\sqrt{n}} + \frac{8 \ln n \left|\left|U_2 U_2^T L_\delta U_{\text{iso}}\right|\right|_{2,\infty}}{\min_{1 \leq i \leq k} \lambda_2(L_i)},
\end{aligned}$$

where we have used (2.2) in the second step. Finally

$$\begin{aligned}
\left|\left|U_2 U_2^T L_\delta U_{\text{iso}}\right|\right|_{2,\infty} &= \left|\left|(I - U_{\text{iso}} U_{\text{iso}}^T) L_\delta U_{\text{iso}}\right|\right|_{2,\infty} \\
&= \left|\left|L_\delta U_{\text{iso}} - U_{\text{iso}} U_{\text{iso}}^T L_\delta U_{\text{iso}}\right|\right|_{2,\infty} \\
&\leq \left(||L_\delta||_\infty + \left|\left|U_{\text{iso}}^T L_\delta U_{\text{iso}}\right|\right|_2 \right) ||U_{\text{iso}}||_{2,\infty} \\
&\leq \left(||L_\delta||_\infty + ||L_\delta||_2 \right) ||U_{\text{iso}}||_{2,\infty} \\
&\leq 4 \sqrt{\frac{c}{n}} \max_{1 \leq i \leq n} d_\delta^{(i)}.
\end{aligned}$$

Hence

$$\min_{V \in \mathbf{O}^k} ||UV - U_{\text{iso}}||_{2,\infty} \leq 32 \sqrt{c} \left(r^2 + r \ln n \right) \frac{1}{\sqrt{n}}.$$

■

4.3. A lemma for Section 3.

Lemma 4.3. *Let $c_1, c_2 \in \mathbb{R}^n$ such that $||c_1 - c_2||_2 = d$. For any $x \in B_{c_1}(r)$ and $y \in B_{c_2}(r)$ let M be the $(n-1)$-dimensional bisecting hyperplane that passes through $\frac{x+y}{2}$ and is perpendicular to the line segment that joins x and y. Then*

$$\operatorname{dist}\{M, B_{c_1}(r) \cup B_{c_2}(r)\} \geq \frac{1}{2}d - 3r.$$

Proof. By symmetry, it suffices to show $\operatorname{dist}\{M, B_{c_1}(r)\} \geq \frac{1}{2}d - 3r$. We may suppose $c_1 = 0$ and $d > 6r$. Then it suffices to show $\operatorname{dist}\{M, 0\} \geq \frac{1}{2}d - 2r$. For any $z \in M$ we have

$$\sum_{i=1}^{n}(x_i - y_i)\left(z_i - \frac{x_i + y_i}{2}\right) = 0.$$

By the point-plane distance formula

$$\begin{aligned} \operatorname{dist}\{M, 0\} = \frac{\left|\sum_{i=1}^{n}(x_i^2 - y_i^2)\right|}{2\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}} &= \frac{||y||_2^2 - ||x||_2^2}{2\,||x - y||_2} \\ &\geq \frac{(d-r)^2 - r^2}{2(d+2r)} \\ &= \frac{1}{2}d - \frac{2dr}{d+2r} \\ &\geq \frac{1}{2}d - 2r. \end{aligned}$$

∎

REFERENCES

[1] E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank. *arXiv:1709.09565 [math, stat]*, May 2019. arXiv: 1709.09565.

[2] D. Aloise, A. Deshpande, P. Hansen, and P. Popat. NP-hardness of Euclidean sum-of-squares clustering. *Machine Learning*, 75(2):245–248, May 2009.

[3] P. Awasthi and O. Sheffet. Improved Spectral-Norm Bounds for Clustering. In A. Gupta, K. Jansen, J. Rolim, and R. Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, Lecture Notes in Computer Science, pages 37–49, Berlin, Heidelberg, 2012. Springer.

[4] A. Bandeira, A. Singer, and T. Strohmer. Mathematics of Data Science, 2020. https://people.math.ethz.ch/~abandeira/BandeiraSingerStrohmer-MDS-draft.pdf.

[5] J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Annals of Statistics*, 47(5):2405–2439, Oct. 2019. Publisher: Institute of Mathematical Statistics.

[6] F. Chung. *Spectral Graph Theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, Dec. 1996. ISSN: 0160-7642, 2380-5668.

[7] A. Damle and Y. Sun. Uniform bounds for invariant subspace perturbations. *arXiv:1905.07865 [cs, math, stat]*, June 2020. arXiv: 1905.07865.

[8] C. Davis and W. M. Kahan. The Rotation of Eigenvectors by a Perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, Mar. 1970. Publisher: Society for Industrial and Applied Mathematics.

[9] S. Deng, S. Ling, and T. Strohmer. Strong Consistency, Graph Laplacians, and the Stochastic Block Model. *arXiv:2004.09780 [cs, stat]*, Apr. 2020. arXiv: 2004.09780.

[10] J. Eldridge, M. Belkin, and Y. Wang. Unperturbed: spectral analysis beyond Davis-Kahan. In *Algorithmic Learning Theory*, pages 321–358, Apr. 2018. ISSN: 1938-7228 Section: Machine Learning.

[11] J. Fan, W. Wang, and Y. Zhong. An ℓ_{∞} Eigenvector Perturbation Bound and Its Application. *Journal of Machine Learning Research*, 18(207):1–42, 2018.

[12] A. Kumar and R. Kannan. Clustering with Spectral Norm and the k-Means Algorithm. *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, 2010.

[13] J. R. Lee, S. O. Gharan, and L. Trevisan. Multiway Spectral Partitioning and Higher-Order Cheeger Inequalities. *Journal of the ACM*, 61(6):37:1–37:30, Dec. 2014.

[14] J. Lei and A. Rinaldo. Consistency of spectral clustering in stochastic block models. *Annals of Statistics*, 43(1):215–237, Feb. 2015. Publisher: Institute of Mathematical Statistics.

[15] X. Li, Y. Li, S. Ling, T. Strohmer, and K. Wei. When do birds of a feather flock together? k-Means, proximity, and conic programming. *Mathematical Programming*, 179(1):295–341, Jan. 2020.

[16] S. Ling and T. Strohmer. Certifying Global Optimality of Graph Cuts via Semidefinite Relaxation: A Performance Guarantee for Spectral Clustering. *Foundations of Computational Mathematics*, June 2019.

[17] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, Mar. 1982.

[18] M. Mahajan, P. Nimbhorkar, and K. Varadarajan. The Planar k-Means Problem is NP-Hard. In S. Das and R. Uehara, editors, *WALCOM: Algorithms and Computation*, Lecture Notes in Computer Science, pages 274–285, Berlin, Heidelberg, 2009. Springer.

[19] A. Marcus, D. A. Spielman, and N. Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. In *2013 IEEE 54th Annual Symposium on Foundations of computer science*, pages 529–537. IEEE, 2013.

[20] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Annals of Statistics*, 39(4):1878–1915, Aug. 2011. Publisher: Institute of Mathematical Statistics.

[21] L. Su, W. Wang, and Y. Zhang. Strong Consistency of Spectral Clustering for Stochastic Block Models. *IEEE Transactions on Information Theory*, 66(1):324–338, Jan. 2020.

[22] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, Dec. 2007.

[23] D. Wagner and F. Wagner. Between Min Cut and Graph Bisection. In G. Goos, J. Hartmanis, A. M. Borzyszkowski, and S. Sokołowski, editors, *Mathematical Foundations of Computer Science 1993*, volume 711, pages 744–750. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993. Series Title: Lecture Notes in Computer Science.