
Article:
Dey, T.K., Peng, P. orcid.org/0000-0003-2700-5699, Rossi, A. et al. (1 more author) (2019) Spectral concentration and greedy k-clustering. Computational Geometry, 76. pp. 19-32. ISSN 0925-7721

https://doi.org/10.1016/j.comgeo.2018.09.001

Spectral Concentration and Greedy k-Clustering

Tamal K. Dey[a,b], Pan Peng[c], Alfred Rossi[a], Anastasios Sidiropoulos[a,b]

[a]*Dept. of Computer Science and Engineering, The Ohio State University. Columbus, OH, 43201.*
[b]*Dept. of Mathematics, The Ohio State University. Columbus, OH, 43201.*
[c]*Department of Computer Science, University of Sheffield*

Abstract

A popular graph clustering method is to consider the embedding of an input graph into $\mathbb{R}^k$ induced by the first k eigenvectors of its Laplacian, and to partition the graph via geometric manipulations on the resulting metric space. Despite the practical success of this methodology, there is limited understanding of several heuristics that follow this framework. We provide theoretical justification for one such natural and computationally efficient variant.

Our result can be summarized as follows. A partition of a graph is called *strong* if each cluster has small external conductance, and large internal conductance. We present a simple greedy spectral clustering algorithm which returns a partition that is provably close to a suitably strong partition, provided that such a partition exists. A recent result shows that strong partitions exist for graphs with a sufficiently large spectral gap between the k-th and $(k+1)$-st eigenvalues. Taking this together with our main theorem gives a spectral algorithm which finds a partition close to a strong one for graphs with large enough spectral gap. We also show how this simple greedy algorithm can be implemented in near-linear time for any fixed k and error guarantee. Finally, we evaluate our algorithm on some real-world and synthetic inputs.

Keywords: Clustering, Greedy Algorithms, Graph Partitioning, Spectral Graph Theory

1. Introduction

Spectral clustering of graphs is a fundamental technique in data analysis that has enjoyed broad practical usage because of its efficacy and simplicity. The technique maps the vertex set of a graph into a Euclidean space $\mathbb{R}^k$ where a classical clustering algorithm (such as k-means, k-center) is applied to the resulting embedding [33]. The coordinates of the vertices in the embedding are computed from k eigenvectors of a matrix associated with the graph. The exact choice of matrix depends on the specific application but is typically some weighted variant of $D - A$, for a graph with degree matrix D and adjacency matrix A. Despite widespread usage, theoretical understanding of the technique remains limited. For example, it is generally not clear for which classes of graphs spectral clustering works well, or what

Email addresses: dey.8@osu.edu (Tamal K. Dey), p.peng@sheffield.ac.uk (Pan Peng), rossi.49@osu.edu (Alfred Rossi), sidiropoulos.1@osu.edu (Anastasios Sidiropoulos)

the structure of the subgraph induced by vertices that correspond to embedded points from the same cluster is. Although the case for $k = 2$ (two clusters) is well understood, the case of general k is not yet settled and a growing body of work seeks to address the practical success of spectral clustering methods [7, 8, 15, 20, 27, 33].

In this paper we present a simple greedy spectral clustering algorithm which is guaranteed to return a high quality partition, provided that one of sufficient quality *exists*. It first chooses k clusters along with their centers greedily from the vertices spectrally embedded in an Euclidean space. Any left over vertex is assigned to one of the computed clusters whose center it is closest to. The resulting partition is close in symmetric difference to the high quality one. Our results can be viewed as providing further theoretical justification for popular clustering algorithms such as in [7] and [20].

Measuring partition quality. Intuitively, a good k-clustering of a graph is one where there are few edges between vertices residing in different clusters and where each cluster is well-connected as an induced subgraph. Such a qualitative definition of clusters can be appropriately characterized by vertex sets with small external conductance and large internal conductance, which has been first formalized by Oveis Gharan and Trevisan [13].

Let $G = (V, E)$ be an undirected unweighted graph. Let $\deg(v)$ be the degree of a vertex $v \in V$. For a subset $S \subset V$, the *external conductance* and *internal conductance* are defined to be

$$\phi_{\mathsf{out}}(S; G) := \frac{|E(S, V(G) \setminus S)|}{\mathrm{vol}(S)}, \quad \phi_{\mathsf{in}}(S) := \min_{S' \subseteq S, \mathrm{vol}(S') \leq \frac{\mathrm{vol}(S)}{2}} \phi_{\mathsf{out}}(S'; G[S])$$

respectively, where $\mathrm{vol}(S) = \sum_{v \in S} \deg(v)$ (called the *volume* of S), $E(X, Y)$ denotes the set of edges between X and Y, and $G[S]$ denotes the subgraph of G induced on S. For an *isolated* vertex v in G, we assume $\phi_{\mathsf{out}}(v, G) = 0$ and $\phi_{\mathsf{in}}(v) = 1$ by definition. Let $\phi_{\mathsf{in}}(G) := \phi_{\mathsf{in}}(V)$. It follows that if $\phi_{\mathsf{in}}(G) > 0$, then G cannot have any isolated vertex. When G is understood from context we sometimes write $\phi_{\mathsf{out}}(S)$ in place of $\phi_{\mathsf{out}}(S; G)$.

We define a *k-partition* of a graph G to be a partition $\mathcal{A} = \{A_1, \ldots, A_k\}$ of $V(G)$ into k disjoint subsets. We say that $\mathcal{A}$ is $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-*strong*, for some $\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}} \geq 0$, if for all $i \in \{1, \ldots, k\}$, we have

$$\phi_{\mathsf{in}}(A_i) \geq \alpha_{\mathsf{in}} \qquad \text{and} \qquad \phi_{\mathsf{out}}(A_i) \leq \alpha_{\mathsf{out}}.$$

Thus a high quality partition is one where α_{in} is large and α_{out} is small.

Our contribution. We present a simple spectral algorithm which computes a partition provably close to *any* $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong k-partition if there is large gap between α_{in} and α_{out} (see Theorem 2.1 for formal statement). We emphasize the fact that the algorithm's output approximates any good existing clustering in the input graph. The algorithm consists of a simple greedy clustering procedure performed on the embedding into $\mathbb{R}^k$ induced by the first k eigenvectors. We further show how to implement this algorithm in near-linear time for any fixed k and error guarantee (see Theorem 5.2).

In the analysis of our algorithm, we show some interesting spectral properties of graphs that admit strong k-partitions: each of the (rescaled) first k eigenvectors of the Laplacian matrix of the graph is close to some vector that is constant on each cluster; the image of each cluster concentrates around some point in the spectral embedding, and all these points are well separated.

Related work. The discrete version of Cheeger's inequality asserts that a graph admits a bipartition into two sets of small external conductance if and only if the second eigenvalue is small [2, 3, 9, 19, 24]. In fact, such a bipartition can be efficiently computed via a simple algorithm that examines the second eigenvector. Generalizations of Cheeger's inequality have been obtained by Lee, Oveis Gharan, and Trevisan [17], and Louis *et al.* [18]. They showed that spectral algorithms can be used to find k disjoint subsets, each with small external conductance, provided that the k-th eigenvalue is small. An improved version of Cheeger's inequality has been obtained by Kwok *et al.* [16] for graphs with large k-th eigenvalue.

Even though the clusters given by the above spectral partitioning methods have small external conductance, they are not guaranteed to have large internal conductance. In other words, for a resulting cluster C, the induced graph $G[C]$ might admit further partitioning into sub-clusters of small conductance. Kannan, Vempala and Vetta proposed quantifying the quality of a partition by measuring the internal conductance of clusters [14]. Allen Zhu, Lattanzi and Mirrokni [1] and Orecchia and Allen Zhu [21] studied local algorithms for extracting subsets with small external conductance under the assumption that subsets with small external conductance and high (internal) connectivity exist.

One may wonder under what conditions a graph admits a partition which provides guarantees on both internal and external conductance. Oveis Gharan and Trevisan, improving on a result of Tanaka [30], showed that graphs which have a sufficiently large spectral gap between the k-th and $(k+1)$-st eigenvalues (denoted as λ_k and λ_{k+1}, respectively) of its Laplacian admit a strong k-partition [13] (see Theorem 2.2). Czumaj et al. [11] recently proposed a sublinear algorithm for testing if a graph with bounded maximum degree has an $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong partition in the framework of property testing, assuming there is some gap between $\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}}$.

Follow-up work. Subsequent to the original ArXiv submission [12] of this paper, Peng, Sun, and Zanetti [23], Awasthi et al. [6] and Sinop [25] have derived spectral algorithms with weaker assumption on the gap between α_{in} and α_{out} (or some related gap, e.g., λ_{k+1} and α_{out}) to cluster the vertices of the graph. The clustering algorithm analyzed in this paper remains distinct from this body of work. For example, in [23] the authors applied k-means clustering to the spectral embedding by the first k eigenvectors; and show that the resulting algorithm is able to find k sets each of which is close to one cluster of a strong k-partition and has bounded small external conductance, under the assumption that $\lambda_{k+1}/\alpha_{\mathsf{out}} = \Omega(k^3)$. (In contrast, our assumption is $\alpha_{\mathsf{in}}^2/\lambda_k = \Omega(k^3 d_{\mathsf{max}}^3)$ for graphs with maximum degree at most d_{max}; see Theorem 2.1.) Their error guarantee ultimately depends on the approximation factor afforded by a k-means algorithm. Unfortunately, k-means is very sensitive to the initial choice of k centers and it is NP-hard to approximate to within some constant

factor [5]. They also gave a heat-kernel based algorithm that runs in near-linear time, which seems to be unappealing for implementation. In [6], the authors proposed an algorithm that iteratively applies the k-means clustering on the resistive embedding projected onto the first k eigenvectors, and outputs a k-partition such that each part is close to one set in a target partition, under the assumption that the ratio between the algebraic expansion of the clusters and α_{out} is $\Omega(k)$. Sinop [25] gave another spectral algorithm assuming that $\lambda_{k+1}/\alpha_{\mathsf{out}} = \Omega(1)$. In each case, [6, 23, 25] use either a different measure of the difference of the output partition and the target partition, or a different definition of conductance (see our remark below Theorem 2.1).

Outline. Section 2 contains a description of the greedy k-clustering algorithm and the statement of our main theorem. In Section 3 we show a spectral concentration property for any graph that admits a high quality partition. Building on this property, we argue that the image of each cluster concentrates around some point in the spectral embedding and these points are well separated. The complete proof of the main theorem is then given in Section 4. In Section 5, we give a randomized version of the algorithm which runs in time $\tilde{O}(mk+\varepsilon^{-1}k^3n)$ for any error parameter $\varepsilon > 0$. Finally, we present some experimental results in Appendix C.

2. Greedy k-Clustering

Let G be an undirected unweighted graph with n vertices, and let $\mathcal{L}_G = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ be its normalized Laplacian, where $\mathbf{A}$ is the adjacency matrix of G and $\mathbf{D}$ is a diagonal matrix with the entries $\mathbf{D}(v,v)$ equal to the degree of vertex v. Let $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of $\mathcal{L}_G$, and $\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_n \in \mathbb{R}^n$ a corresponding collection of orthonormal left eigenvectors[1]. Note that by the variational characterization of eigenvalues, $\frac{\mathbf{v}_i \mathcal{L}_G \mathbf{v}_i^T}{\mathbf{v}_i \mathbf{v}_i^T} = \lambda_i$ for $1 \leq i \leq n$ (see [10]).

In this paper we consider a simple geometric clustering operation on the embedding $\mathbf{F}(u)$ which carries a vertex u to a point given by a rescaling of the first k eigenvectors of $\mathcal{L}_G$,

$$\mathbf{F}(u) = \deg(u)^{-1/2} \left(\mathbf{v}_1(u), \ldots, \mathbf{v}_k(u)\right). \tag{1}$$

For any $U \subseteq V(G)$, let $\mathbf{F}(U)$ denote all the embedded points corresponding to vertices in U, that is, $\mathbf{F}(U) = \{\mathbf{F}(u) : u \in U\}$. For any set $B \subseteq \mathbb{R}^k$, let $\mathbf{F}^{-1}(B) := \{v \in V(G) : \mathbf{F}(v) \in B\}$. For any point $\mathbf{x} \in \mathbb{R}^k$ and real number $R \geq 0$, let $\mathsf{ball}(\mathbf{x}, R) := \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\|_2 \leq R\}$.

Intuitive description. The algorithm takes as input a graph G, and a desired number of clusters, k. The algorithm uses the bottom k eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of $\mathcal{L}_G$ to compute the embedding $\mathcal{F} = \mathbf{F}(V(G))$ of G into $\mathbb{R}^k$. Next, it begins an iterative process of searching for regions of the embedding containing many points from $\mathcal{F}$, and removing them to form clusters. To do so, it first computes a distance threshold $R = R(k, G) = R(k, n, d_{\mathsf{max}})$,

[1] We will use $\mathbf{x}$ to denote a row vector and $\mathbf{x}^T$ to denote a column vector.

1. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be the k first eigenvectors of $\mathcal{L}_G$.
2. Let $\mathbf{F} : V(G) \to \mathbb{R}^k$, where $\mathbf{F}(u) = \deg(u)^{-1/2} (\mathbf{v}_1(u), \ldots, \mathbf{v}_k(u))$.
3. $R = \frac{1}{36kd_{\mathsf{max}}\sqrt{n}}$
4. $V_0 = V(G)$
5. for $i = 1, \ldots, k$
6. $\quad u_i = \operatorname{argmax}_{u \in V_{i-1}} |\mathsf{ball}(\mathbf{F}(u), 2R) \cap \mathbf{F}(V_{i-1})|$
   $\quad\quad = \operatorname{argmax}_{u \in V_{i-1}} |\{w \in V_{i-1} : \|\mathbf{F}(u) - \mathbf{F}(w)\|_2 \leq 2R\}|$
7. $\quad P_i = \mathbf{F}^{-1}(\mathsf{ball}(\mathbf{F}(u_i), 2R)) \cap V_{i-1}$
8. $\quad V_i = V_{i-1} \setminus P_i$
9. Let $\mathbf{g} : V_k \to \{u_1, \cdots, u_k\}$, $\mathbf{g}(v) = u_i$ if i is the smallest index
   $\quad\quad$ satisfying $\|\mathbf{F}(v) - \mathbf{F}(u_i)\| \leq \|\mathbf{F}(v) - \mathbf{F}(u_j)\|$ for all $j \neq i$.
10. Return $\{C_1, \ldots, C_k\} = \{P_1 \cup \mathbf{g}^{-1}(u_1), \ldots, P_k \cup \mathbf{g}^{-1}(u_k)\}$

Figure 1: The greedy spectral k-clustering algorithm takes an n-vertex graph G with maximum degree at most d_{max} as input, and outputs a partition $\mathcal{C} = \{C_1, \ldots, C_k\}$ of $V(G)$.

where $n = |V(G)|$ and d_{max} represent an upper bound of the maximum degree. Using this threshold, the algorithm looks for a point $p \in \mathcal{F}$ such that the number of near-by points of $\mathcal{F}$ (points of $\mathcal{F}$ which fall within a radius of $2R$ of p) is maximized. The vertices corresponding to these points (including p) are made into a cluster, and p is remembered as the location of the cluster in the embedding. Next, p and its near-by points are removed from $\mathcal{F}$. This iterative process continues either for k iterations, or until there are no points of $\mathcal{F}$ left in the embedding. Afterward, any remaining points of $\mathcal{F}$ are thought of as "outliers", and each has its corresponding vertex assigned to a nearest cluster.

A more formal description of the algorithm appears in Figure 1. In Section 5, we show how the algorithm can be implemented in time $\tilde{O}(mk + \varepsilon^{-1}k^3 n)$ for any error parameter $\varepsilon > 0$, where m denotes the number of edges of the graph and $\tilde{O}(\cdot)$ hides $\operatorname{polylog} n$ factors.

To measure the performance of our algorithm, we introduce the following notion of symmetric difference between two collections, each of k sets, that generalizes the symmetric difference between two sets.

A distance on k-partitions. For two sets Y, Z, their symmetric difference is given by $Y \triangle Z = (Y \setminus Z) \cup (Z \setminus Y)$. Let X be a finite set, $k \geq 1$, and let $\mathcal{A} = \{A_1, \ldots, A_k\}$, $\mathcal{A}' = \{A'_1, \ldots, A'_k\}$ be collections of disjoint subsets of X. Then, we define a distance function between $\mathcal{A}$, $\mathcal{A}'$, by

$$|\mathcal{A} \triangle \mathcal{A}'| = \min_{\sigma} \sum_{i=1}^{k} \left| A_i \triangle A'_{\sigma(i)} \right|,$$

where σ ranges over all permutations σ on $\{1, \ldots, k\}$.

2.1. Main Theorem

Theorem 2.1 (Spectral partitioning via greedy k-clustering)**.** *Let G be an n-vertex graph with maximum degree at most d_{max}. Let $k \geq 1$, and let $\mathcal{A} = \{A_1, \ldots, A_k\}$ be any $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong k-partition of $V(G)$ with $\alpha_{\mathsf{in}} > 10(kd_{\mathsf{max}})^{3/2} \cdot \sqrt{\lambda_k}$. Then, on input G, the algorithm in Figure 1 outputs a partition $\mathcal{C}$ such that*

$$|\mathcal{A} \triangle \mathcal{C}| = O\left(\frac{\lambda_k \cdot d_{\mathsf{max}}^3 k^5 \cdot n}{\alpha_{\mathsf{in}}^2}\right).$$

Due to the dependency on d_{max}, the above result is mainly interesting for bounded degree graphs. We remark that even for some special classes of bounded degree graphs, analyzing the performance of spectral clustering algorithms is already interesting and challenging. For example, in their seminal work Spielman and Teng gave the first rigorous analysis of the performance of spectral clustering methods which use the second eigenvector of the matrix $D - A$ on bounded degree planar graphs and finite element meshes [28]. This result was further generalized to graphs with bounded degree and bounded genus by Kelner [15] and excluded-minor graphs by Biswal, Lee and Rao [8]. Our result holds for arbitrary bounded degree graphs that admit a good quality k-partition and demonstrates the effectiveness of the spectral clustering algorithms that use only the first k eigenvectors (cf. Alpert and Yao [4]).

We further remark that while α_{out} does not appear explicitly in the error term in Theorem 2.1, it does implicitly bound the error through a higher order Cheeger inequality [17]. In particular, $\lambda_k \leq 2\alpha_{\mathsf{out}}$ and thus when $\alpha_{\mathsf{out}}/\alpha_{\mathsf{in}}^2$ is small there is strong agreement between $\mathcal{A}$ and $\mathcal{C}$. In addition, the dependency on the upper bound d_{max} of maximum vertex degree seems unavoidable since we are measuring the *size* of the symmetric difference $\mathcal{A} \triangle \mathcal{C}$ rather than its *volume* or *total degree*. (Note that the definition of conductance is volume-based, while the error is measured with respect to the size of clusters. Such an inconsistency seems to cause the dependency on d_{max}.) In contrast, the latter measurement was used in [23], which allows the authors to derive an error term that is independent of the maximum degree. On the other hand, such a dependency also does not appear in [6] and [25], as the authors are studying the size-based definition of conductance (i.e., $\phi_{\mathsf{out}}(S; G) := \frac{|E(S, V(G)\setminus S)|}{|S|}$) instead of the volume-based definition as in our paper. We also note that the algorithms in some follow-up work (e.g., [6, 23]) can output some partition with an approximate guarantee for individual clusters, while our algorithm can only have approximate guarantee over the all k clusters.

Application of main theorem. Oveis Gharan and Trevisan [13] (see also [30]) showed that, if the *gap* between λ_k and λ_{k+1} is large enough, then there exists a partition into k clusters, each having small external conductance and large internal conductance.

Theorem 2.2 ([13])**.** *There exist universal constants $c > 0$, $\alpha > 0$, and $\beta > 0$, such that for any graph G with $\lambda_{k+1} > ck^2\sqrt{\lambda_k}$, there is a $(\alpha \cdot \lambda_{k+1}/k, \beta \cdot k^3\sqrt{\lambda_k})$-strong k-partition of G.*

The same paper [13] also shows how to compute a partition with slightly worse quantitative guarantees, using an iterative combinatorial algorithm with polynomial running time. More specifically, they have shown the following theorem.

Theorem 2.3 ([13]). *There is a polynomial time algorithm that takes as input a graph G with $\lambda_{k+1} > 0$ for any $k \geq 1$, outputs an ℓ-partition that is $(\Omega(\lambda_{k+1}^2/k^4), O(k^6\sqrt{\lambda_k}))$-strong, for some $1 \leq \ell < k+1$.*

Let τ be a number satisfying $\tau \geq \tau_0 := \max\{c, 10 \cdot \alpha^{-1} \cdot \sqrt{k} d_{\mathsf{max}}^{3/2}\}$, where α, c are the constants given in Theorem 2.2. Let G be a graph with $\lambda_{k+1} > \tau k^2 \sqrt{\lambda_k}$. By applying Theorem 2.1 on G with parameters $\alpha_{\mathsf{in}} = \alpha \cdot \lambda_{k+1}/k$, which satisfies that $\alpha_{\mathsf{in}} > 10(k d_{\mathsf{max}})^{3/2} \cdot \sqrt{\lambda_k}$, we obtain the following corollary.

Corollary 2.4. *Let $k \geq 1$. Let G be an n-vertex graph with maximum degree at most d_{max}, and $\lambda_{k+1} > \tau k^2 \sqrt{\lambda_k}$, where $\tau \geq \tau_0$ and τ_0 is defined as above. Let $\mathcal{A}$ be the $(\alpha \cdot \lambda_{k+1}/k, \beta \cdot k^3\sqrt{\lambda_k})$-strong partition of G guaranteed by Theorem 2.2. Then, on input G, the algorithm in Figure 1 outputs a partition $\mathcal{C}$ such that*

$$|\mathcal{A} \triangle \mathcal{C}| = O\left(\frac{d_{\mathsf{max}}^3 k^3 n}{\tau^2}\right).$$

In comparison with the algorithm from Theorem 2.3 that finds a partition that is a $(\Omega(\lambda_{k+1}^2/k^4), O(k^6\sqrt{\lambda_k}))$-strong partition, our algorithm finds a partition that is *close to* some $(\Omega(\lambda_k/k), O(k^3\sqrt{\lambda_k}))$-strong partition. It is not clear how to find in polynomial time a partition (without error) that is $(\Omega(\lambda_k/k), O(k^3\sqrt{\lambda_k}))$-strong.

3. Spectral Concentration

In this section, we prove that for any graph with some strong k-partition and any eigenvector $\mathbf{v}_i$ $(1 \leq i \leq k)$, the rescaled vector $\mathbf{x}_i := \mathbf{v}_i \mathbf{D}^{-1/2}$ is close (with respect to the ℓ_2 norm) to some vector $\widetilde{\mathbf{x}}_i$ that is constant on each cluster. We slightly abuse the notation by also using $\mathbf{F}$ to denote the $n \times k$ matrix that corresponds to our spectral embedding (i.e., with row vectors $\mathbf{F}(u)$, for all $u \in V$). It is useful to note that $\mathbf{F} = [\mathbf{x}_1^T, \cdots, \mathbf{x}_k^T]$.

Lemma 3.1. *Let G be a graph with maximum degree at most d_{max}. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^k$ denote the first k eigenvectors of $\mathcal{L}_G$. For $\alpha_{\mathsf{in}} > 0$, let $\mathcal{A} = \{A_1, \ldots, A_k\}$ be any $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong k-partition of $V(G)$. For any $i \in \{1, \ldots, k\}$, if $\mathbf{x}_i = \mathbf{v}_i \mathbf{D}^{-1/2}$, then there exists $\widetilde{\mathbf{x}}_i \in \mathbb{R}^n$, such that,*

(i) $\|\mathbf{x}_i - \widetilde{\mathbf{x}}_i\|_2^2 \leq \frac{2k\lambda_k \cdot d_{\mathsf{max}}}{\alpha_{\mathsf{in}}^2}$, *and*

(ii) $\widetilde{\mathbf{x}}_i$ *is constant on the clusters of $\mathcal{A}$, i.e. for any $A \in \mathcal{A}$, $u, v \in A$, we have $\widetilde{\mathbf{x}}_i(u) = \widetilde{\mathbf{x}}_i(v)$.*

Before laying out the proof, we provide some explanation of the statement of the theorem. First, note that the ℓ_2^2-distance between $\mathbf{x}_i$ and its uniform approximation $\widetilde{\mathbf{x}}_i$ depends linearly on the ratio $\lambda_k/\alpha_{\mathsf{in}}^2$, which, as noted above, is bounded from above by $2\alpha_{\mathsf{out}}/\alpha_{\mathsf{in}}^2$. Second, the partition-wise uniform vector $\widetilde{\mathbf{x}}_i$ which minimizes the left hand side of (i) is constructed by taking the mean values of $\mathbf{x}_i$ on each partition. This, together with the bound in (i), means that $\mathbf{x}_i$ assumes values in each partition close to their mean. In summary, if there is

a sufficiently large gap between the external conductance α_{out} and internal conductance α_{in} of the clusters, the values taken by each vector $\mathbf{x}_i$ have k prominent modes over k partitions.

We need the following result that is a slight restatement of a lemma in [11] to prove Lemma 3.1. (For completeness, a proof of Lemma 3.2 is included in Appendix A.)

Lemma 3.2. *Let $G = (V, E)$ be any undirected graph and let $C \subseteq V$ be any subset with $\phi(G[C]) \geq \phi_{\mathsf{in}} > 0$. Then for every i, $1 \leq i \leq k$, $\mathbf{x}_i = \mathbf{v}_i \mathbf{D}^{-1/2}$, the following holds:*

$$\sum_{u,v \in C} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \frac{4\lambda_k \cdot \mathrm{vol}(C)}{\phi_{\mathsf{in}}^2}.$$

We remark that in the above Lemma, there is a linear dependency on $\mathrm{vol}(C)$, which directly causes our result and the following analysis to depend on d_{max}. Now we are ready to prove Lemma 3.1.

Proof of Lemma 3.1. Let $1 \leq i \leq k$, and $1 \leq j \leq k$. By precondition of the lemma, $\mathcal{A}$ is an $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong k-partition. Now we apply $C = A_j$ and $\phi_{\mathsf{in}} = \alpha_{\mathsf{in}}$ in Lemma 3.2 to get

$$\sum_{u,v \in A_j} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \frac{4\lambda_k \cdot \mathrm{vol}(A_j)}{\alpha_{\mathsf{in}}^2} \leq \frac{4\lambda_k \cdot d_{\mathsf{max}} \cdot |A_j|}{\alpha_{\mathsf{in}}^2},$$

where the second inequality follows from our assumption that the maximum degree is d_{max}. Let $\widetilde{\mathbf{x}}_i(u) = \frac{\sum_{v \in A_j} \mathbf{x}(v)}{|A_j|}$ if $u \in A_j$. Note that $\widetilde{\mathbf{x}}_i$ is constant on each cluster. On the other hand, by the definition of $\widetilde{\mathbf{x}}_i$, we have

$$\frac{1}{|A_j|} \sum_{u,v \in A_j} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 = 2 \sum_{u \in A_j} (\mathbf{x}_i(u) - \widetilde{\mathbf{x}}_i(u))^2.$$

Therefore,

$$\begin{aligned} \|\mathbf{x}_i - \widetilde{\mathbf{x}}_i\|_2^2 &= \sum_{j=1}^{k} \sum_{u \in A_j} (\mathbf{x}_i(u) - \widetilde{\mathbf{x}}_i(u))^2 = \frac{1}{2} \sum_{j=1}^{k} \frac{1}{|A_j|} \sum_{u,v \in A_j} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \\ &\leq \frac{2k\lambda_k \cdot d_{\mathsf{max}}}{\alpha_{\mathsf{in}}^2}. \end{aligned}$$

This completes the proof of the lemma. □

4. From Spectral Concentration to Spectral Clustering

In this section we prove Theorem 2.1. We begin by showing that if there exists strong k-partition with high quality in the graph G, then in the spectral embedding defined by $\mathbf{F}(u) = \deg(u)^{-1/2}(\mathbf{v}_1(u), \ldots, \mathbf{v}_k(u))$ for any $u \in V$, one can find k well-separated center points in $\mathbb{R}^k$ such that the balls (of some appropriately chosen radius) centered at these center points are disjoint and the collection of these balls is close to any strong k-partition.

Lemma 4.1. *Let G be an n-vertex graph of maximum degree at most d_{max}. Let $\mathcal{A} = \{A_1, \ldots, A_k\}$ be any $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong k-partition of $V(G)$ with $\alpha_{\mathsf{in}} > 10(kd_{\mathsf{max}})^{3/2} \cdot \sqrt{\lambda_k}$. Let $\mathbf{F} : V(G) \to \mathbb{R}^k$ be the spectral embedding of G given by (1). Let $R = \frac{1}{36kd_{\mathsf{max}}\sqrt{n}}$. Then there exists k points $\mathbf{p}_1, \ldots, \mathbf{p}_k \in \mathbb{R}^k$ and a family, $\mathcal{A}'$, of k subsets of $V(G)$, given by $A'_i = \{u \in V : \|\mathbf{F}(u) - \mathbf{p}_i\|_2 \leq R\}$ for $1 \leq i \leq k$, such that the following conditions are satisfied:*

(i) For $1 \leq i < j \leq k$, $\|\mathbf{p}_i - \mathbf{p}_j\|_2 > 6R$.

(ii) The elements of $\mathcal{A}'$ are pairwise disjoint.

(iii) $|\mathcal{A} \triangle \mathcal{A}'| = O\left(\frac{\lambda_k \cdot d_{\mathsf{max}}^3 k^4 \cdot n}{\alpha_{\mathsf{in}}^2}\right)$.

To prove Lemma 4.1, we first give some definitions and introduce some useful tools. For any symmetric matrix $\mathbf{X}$, let $\eta_i(\mathbf{X})$ denote the ith largest eigenvalue of $\mathbf{X}$. For any (not necessarily square) matrix $\mathbf{Y}$, let $\mathbf{Y}_{\mathrm{row}(i)}$ denote the ith row vector of $\mathbf{Y}$. We will make use of the following pair of facts which are proved in Appendix B:

Fact 4.2. *For any two $p \times p$ symmetric matrices $\mathbf{X}, \mathbf{Y}$, if $\max_{i \leq p} \|\mathbf{X}_{\mathrm{row}(i)} - \mathbf{Y}_{\mathrm{row}(i)}\|_2 \leq \delta$, then for any $i \leq p$, $|\eta_i(\mathbf{X}) - \eta_i(\mathbf{Y})| \leq \sqrt{p} \cdot \delta$.*

Fact 4.3. *For any two $p \times q$ matrices $\mathbf{X}, \mathbf{Y}$, if $\max_{i \leq p} \|\mathbf{X}_{\mathrm{row}(i)}\|_2 \leq \gamma$, and $\max_{i \leq p} \|\mathbf{X}_{\mathrm{row}(i)} - \mathbf{Y}_{\mathrm{row}(i)}\|_2 \leq \delta$, then $\max_{i \leq p} \|(\mathbf{X} \cdot \mathbf{X}^T)_{\mathrm{row}(i)} - (\mathbf{Y} \cdot \mathbf{Y}^T)_{\mathrm{row}(i)}\|_2 \leq \sqrt{p}(\delta^2 + 2\gamma\delta)$.*

Now we prove Lemma 4.1.

Proof of Lemma 4.1. Recall that for any i, $1 \leq i \leq k$, $\mathbf{x}_i = \mathbf{v}_i \mathbf{D}^{-1/2}$ and $\widetilde{\mathbf{x}}_i$ denotes the vector that $\widetilde{\mathbf{x}}_i(u) = \frac{1}{|A_j|} \sum_{v \in A_j} \mathbf{x}_i(v)$ if $u \in A_j$. Now for each $1 \leq j \leq k$, define

$$\mathbf{p}_j := (\widetilde{\mathbf{x}}_1(u), \cdots, \widetilde{\mathbf{x}}_k(u)), \quad \text{for any } u \in A_j.$$

Further recall that $\mathbf{F} \in \mathbb{R}^{n \times k}$ is the matrix corresponding to our spectral embedding and that $\mathbf{F} = [\mathbf{x}_1^T, \cdots, \mathbf{x}_k^T]$. Let $\mathbf{P} := [\widetilde{\mathbf{x}}_1^T, \cdots, \widetilde{\mathbf{x}}_k^T]$. Note that for any $u \in A_j$, the row vector corresponding to u of $\mathbf{P}$ is $\mathbf{p}_j$. Let $\zeta = \frac{2k\lambda_k \cdot d_{\mathsf{max}}}{\alpha_{\mathsf{in}}^2}$. We have the following claim about the eigenvalues of matrix $\mathbf{P}^T \cdot \mathbf{P}$.

Claim 4.4. *All eigenvalues of $\mathbf{P}^T \cdot \mathbf{P}$ are at least $\frac{1}{d_{\mathsf{max}}} - k(\zeta + 2\sqrt{\zeta})$.*

Proof. Let $\mathbf{T}$ be the $n \times k$ matrix with ith column $\mathbf{v}_i^T$, for each $i \leq k$. Thus, $\mathbf{F} = \mathbf{D}^{-1/2}\mathbf{T}$. Now we note that all the eigenvalues of $\mathbf{F}^T \cdot \mathbf{F}$ are at least $1/d_{\mathsf{max}}$. This is true since for any $\mathbf{y} \in \mathbb{R}^k$,

$$\mathbf{y}(\mathbf{F}^T \cdot \mathbf{F})\mathbf{y}^T = \|\mathbf{y}\mathbf{F}^T\|_2^2 = \|\mathbf{y}\mathbf{T}^T\mathbf{D}^{-1/2}\|_2^2 \geq \|\mathbf{y}\mathbf{T}^T\|_2^2/d_{\mathsf{max}} = \mathbf{y} \cdot \mathbf{y}^T/d_{\mathsf{max}},$$

where the second to last inequality follows from the fact that $\|\mathbf{z}\mathbf{M}\|_2^2 \geq (\min_i \mathbf{M}_{i,i})^2 \|\mathbf{z}\|_2^2$ for any vector $\mathbf{z} \in \mathbb{R}^n$ and diagonal matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$, which is true since $\|\mathbf{z}\mathbf{M}\|_2^2 =$

$\sum_i (\mathbf{z}_i \mathbf{M}_{i,i})^2 \geq (\min_i \mathbf{M}_{i,i})^2 \sum_i (\mathbf{z}_i)^2 = (\min_i \mathbf{M}_{i,i})^2 \|\mathbf{z}\|_2^2$; and the last equation follows from the observation that $\mathbf{T}^T \cdot \mathbf{T} = \mathbf{I}_{k\times k}$.

Now note that for each $i \leq k$, since $\mathbf{x}_i = \mathbf{v}_i \mathbf{D}^{-1/2}$, and $\|\mathbf{v}_i\|_2 = 1$, it holds that $\|\mathbf{x}_i\|_2 \leq 1$. On the other hand, by Lemma 3.1, it holds that

$$\|\mathbf{F}^T_{\mathrm{row}(i)} - \mathbf{P}^T_{\mathrm{row}(i)}\|_2 = \|\mathbf{x}_i - \widetilde{\mathbf{x}}_i\|_2 \leq \sqrt{\zeta}.$$

By Fact 4.3, we have that $\max_{i\leq p} \|(\mathbf{F}^T \cdot \mathbf{F})_{\mathrm{row}(i)} - (\mathbf{P}^T \cdot \mathbf{P})_{\mathrm{row}(i)}\|_2 \leq \sqrt{k}(\zeta + 2\sqrt{\zeta})$. Then by Fact 4.2, for each $i \leq k$, $|\eta_i(\mathbf{F}^T \cdot \mathbf{F}) - \eta_i(\mathbf{P}^T \cdot \mathbf{P})| \leq k(\zeta + 2\sqrt{\zeta})$. Since all the eigenvalues of $\mathbf{F}^T \cdot \mathbf{F}$ are at least $\frac{1}{d_{\mathsf{max}}}$, it follows that that all eigenvalues of $\mathbf{P}^T \cdot \mathbf{P}$ are at least $\frac{1}{d_{\mathsf{max}}} - k(\zeta + 2\sqrt{\zeta})$. □

On the other hand, we prove in the following claim that if there exists two vectors $\mathbf{p}_{i_0}, \mathbf{p}_{j_0}$ that are close, then $\mathbf{P}^T \cdot \mathbf{P}$ has at least one small eigenvalue.

Claim 4.5. *If there exists $i_0, j_0 \leq k$ such that $\|\mathbf{p}_{i_0} - \mathbf{p}_{j_0}\|_2 \leq 6R$, then $\mathbf{P}^T \cdot \mathbf{P}$ has an eigenvalue at most $k(36R^2 n + 12R\sqrt{n})$.*

Proof. Let $\mathbf{Q}$ denote the $n \times k$ matrix obtained from $\mathbf{P}$ by replacing each row vector that equals $\mathbf{p}_{i_0}$ by vector $\mathbf{p}_{j_0}$. Note that $\mathbf{Q}^T \cdot \mathbf{Q}$ is singular, and thus has eigenvalue 0.

Now note that $\max_{i\leq k} \|\mathbf{P}^T_{\mathrm{row}(i)} - \mathbf{Q}^T_{\mathrm{row}(i)}\|_2 \leq 6R\sqrt{n}$ since the absolute value of each entry in $\mathbf{P} - \mathbf{Q}$ is at most $6R$, and also note that for any $i \leq k$,

$$\begin{aligned} \|\mathbf{P}^T_{\mathrm{row}(i)}\|_2 &= \sqrt{\sum_{j=1}^{k} |A_j| \left(\frac{\sum_{u\in A_j} \mathbf{x}_i(u)}{|A_j|}\right)^2} \leq \sqrt{\sum_{j=1}^{k} |A_j| \frac{\sum_{u\in A_j} \mathbf{x}_i^2(u)}{|A_j|}} \\ &= \|\mathbf{x}_i\|_2 \leq 1. \end{aligned}$$

Then by Fact 4.3, $\max_{i\leq k} \|(\mathbf{P}^T \cdot \mathbf{P})_{\mathrm{row}(i)} - (\mathbf{Q}^T \cdot \mathbf{Q})_{\mathrm{row}(i)}\|_2 \leq \sqrt{k}\,(36R^2 n + 2 \cdot 6R\sqrt{n})$. Now by Fact 4.2 and the fact that $\mathbf{Q}^T \cdot \mathbf{Q}$ has eigenvalue 0, we know that at least one eigenvalue of $\mathbf{P}^T \cdot \mathbf{P}$ is at most $k\,(36R^2 n + 12R\sqrt{n})$. □

Now we prove Item (i) of the lemma. By the assumption $\alpha_{\mathsf{in}} > 10(k d_{\mathsf{max}})^{3/2}\sqrt{\lambda_k}$, it holds that $\zeta = \frac{2k\lambda_k \cdot d_{\mathsf{max}}}{\alpha_{\mathsf{in}}^2} < \frac{1}{50k^2 d_{\mathsf{max}}^2}$, which implies that $\frac{1}{d_{\mathsf{max}}} - k(\zeta + 2\sqrt{\zeta}) > \frac{1}{2d_{\mathsf{max}}}$. On the other hand, since $R = \frac{1}{36kd_{\mathsf{max}}\sqrt{n}}$, we have that $k(36R^2 n + 12R\sqrt{n}) \leq \frac{1}{2d_{\mathsf{max}}}$. Therefore, by Claim 4.4 and Claim 4.5, we have reached a contradiction regarding the minimum eigenvalue of $\mathbf{P}^T \cdot \mathbf{P}$. This implies Item (i) of the lemma, that is, for all $1 \leq i < j \leq k$, $\|\mathbf{p}_i - \mathbf{p}_j\|_2 > 6R$. Now for each $1 \leq j \leq k$, define $A'_j := \{u \in V : \|\mathbf{F}(u) - \mathbf{p}_j\|_2 \leq R\}$ as required by Item (ii). Let $\mathcal{A}' = \{A'_1, \ldots, A'_k\}$. By Item (i) $A'_1, \cdots, A'_k$ are disjoint, proving Item (ii).

Finally, we prove Item (iii) of the lemma. By Lemma 3.1,

$$\sum_{i=1}^{k} \|\mathbf{x}_i - \widetilde{\mathbf{x}}_i\|_2^2 \leq k \cdot \zeta.$$

On the other hand, if we let $A_{\mathrm{bad}} = \{u : u \in A_j, \|\mathbf{F}(u) - \mathbf{p}_j\| > R, 1 \le j \le k\}_2$, then

$$\sum_{i=1}^{k} \|\mathbf{x}_i - \widetilde{\mathbf{x}}_i\|_2^2 = \sum_{j=1}^{k} \sum_{u \in A_j} \sum_{i=1}^{k} (\mathbf{x}_i(u) - \widetilde{\mathbf{x}}_i(u))^2 \;=\; \sum_{j=1}^{k} \sum_{u \in A_j} \|\mathbf{F}(u) - \mathbf{p}_j\|_2^2$$
$$\geq \sum_{u \in A_{\mathrm{bad}}} R^2 = |A_{\mathrm{bad}}| \cdot R^2.$$

Therefore, $|A_{\mathrm{bad}}| \le \frac{k\zeta}{R^2} \le \frac{2592 \cdot \lambda_k \cdot d_{\mathsf{max}}^3 k^4 \cdot n}{\alpha_{\mathsf{in}}^2}$.

Now we observe that

$$|\mathcal{A} \triangle \mathcal{A}'| \le \sum_{j=1}^{k} |A_j \triangle A'_j| = \sum_{j=1}^{k} (|A_j \setminus A'_j| + |A'_j \setminus A_j|) \le |A_{\mathrm{bad}}| + \sum_{j=1}^{k} |A'_j \setminus A_j|.$$

Note that for any $u \in A'_j \setminus A_j$, it holds that $u \in A_i$ for some $i \neq j$. Since $u \in A'_j$, it holds that $\|\mathbf{F}(u) - \mathbf{p}_j\|_2 \le R$, which implies that $\|\mathbf{F}(u) - \mathbf{p}_i\|_2 \ge \|\mathbf{p}_j - \mathbf{p}_i\|_2 - \|\mathbf{F}(u) - \mathbf{p}_j\|_2 > 6R - R = 5R$. This further implies that $u \in A_{\mathrm{bad}}$. Since all $A'_1, \cdots, A'_k$ are disjoint, we have that $\sum_{j=1}^{k} |A'_j \setminus A_j| \le |A_{\mathrm{bad}}|$.

Thus, it holds that

$$|\mathcal{A} \triangle \mathcal{A}'| \le 2|A_{\mathrm{bad}}|,$$

which proves the Item (iii) of the lemma. □

We are now ready to prove our main theorem.

Proof of Theorem 2.1. Let $\mathcal{A} = \{A_1, \ldots, A_k\}$, $\mathcal{A}' = \{A'_1, \ldots, A'_k\}$, R, and $\mathbf{p}_1, \ldots, \mathbf{p}_k$ be as in Lemma 4.1. Let $\varepsilon = |\mathcal{A} \triangle \mathcal{A}'|/n = O(\frac{\lambda_k \cdot d_{\mathsf{max}}^3 k^4}{\alpha_{\mathsf{in}}^2})$. Let $\mathcal{C} = \{C_1, \ldots, C_k\}$ be the ordered collection of pairwise disjoint subsets of $V(G)$ output by the greedy spectral k-clustering algorithm in Figure 1. Let $\mathcal{P} = \{P_1, \ldots, P_k\}$ where P_i is the subset, called *group*, found by the algorithm at the ith iteration for $1 \le i \le k$. The set of vertices not covered by any of the clusters in $\mathcal{A}'$ plays a special role in our argument which we denote as $B = V(G) \setminus \left(\bigcup_{i=1}^{k} A'_i\right)$. Clearly, $|B| \le |\mathcal{A} \triangle \mathcal{A}'| \le \varepsilon n$.

We say that a cluster A'_i is *touched* if the algorithm, while computing the centers, considers a group $P_j \in \mathcal{P}$ with $P_j \cap A'_i \neq \emptyset$. For a cluster A'_i, let $P_{\rho(i)}$ be the group in $\mathcal{P}$ that touches A'_i for the *first time* in the algorithm if it is touched at all. Let $I \subseteq \{1, \ldots, k\}$ be the support of ρ, that is, $\rho(i)$ exists if and only if $i \in I$. Let $i^* = |I|$. By permuting the indices of the clusters in $\mathcal{A}'$, we may assume w.l.o.g. that $I = \{1, \ldots, i^*\}$.

First we observe that ρ is a bijection on I. This is because the group $P_{\rho(i)}$, $i \in I$, can intersect at most one cluster in $\mathcal{A}'$. The reason is that every cluster in $\mathcal{A}'$ is contained inside some ball of radius R, the distance between any two centers of such balls is more than $6R$, and each $P_{\rho(i)}$ is contained inside some ball of radius $2R$.

In case $i \in I$, we have $|P_{\rho(i)} \setminus A'_i| \le |B \cap P_{\rho(i)}|$. This is because $P_{\rho(i)}$ cannot intersect any other cluster in $\mathcal{A}'$ but B. On the other hand, it holds that $|A'_i \setminus P_{\rho(i)}| \le \varepsilon n$, since otherwise

the algorithm could have made a better choice by taking the entire A'_i while computing $P_{\rho(i)}$. Such a choice can be made by taking $P_{\rho(i)}$ to be all the yet unclustered points that are inside a ball of radius $2R$ centered at any point in A'_i; since A'_i is in a ball of radius R, it follows by the triangle inequality that A'_i will be contained inside $P_{\rho(i)}$. Therefore, for every $i \leq i^*$, $|A'_i \triangle P_{\rho(i)}| = |P_{\rho(i)} \setminus A'_i| + |A'_i \setminus P_{\rho(i)}| \leq |B \cap P_{\rho(i)}| + \varepsilon n$.

In the other case when $i \notin I$, we claim that the cluster A'_i can have at most $2\varepsilon n$ vertices. Suppose not. Since ρ is a bijection on I and I is a proper subset of $\{1, \cdots, k\}$, there is a group P_j with $j \notin \rho(I)$, which does not intersect any cluster in $\mathcal{A}'$ for the first time. Then, it has the only option of intersecting a cluster in $\mathcal{A}'$ beyond the first time and/or intersect B. Since $|A'_i \setminus P_{\rho(i)}| \leq \varepsilon n$ for all $i \in I$, P_j can have at most $\varepsilon n + |B| \leq 2\varepsilon n$ vertices. But, the algorithm could have made a better choice by selecting A'_i while computing P_j because $|A'_i| > 2\varepsilon n$ by our assumption. We reach a contradiction.

Let $T = \{1, \cdots, k\} \setminus \{\rho(1), \cdots, \rho(i^*)\}$ be the set of indices j such that P_j does not intersect any cluster in $\mathcal{A}'$ for the first time. For any $j \in T$, P_j can only intersect set B and/or set $A_i \setminus P_{\rho(i)}$ for some i such that $\rho(i) < j$. This then gives that $|\cup_{j \in T} P_j| \leq \sum_{j \in T} |B \cap P_j| + \sum_{i \leq i^*} |A'_i \setminus P_{\rho(i)}|$.

Using the bijectivity of ρ on the set $\{1, \ldots, i^*\}$, we have

$$\begin{aligned}
|\mathcal{A}' \triangle \mathcal{P}| &\leq \sum_{i \leq i^*} |A'_i \triangle P_{\rho(i)}| + |\cup_{j \in T} P_j| + |\cup_{i > i^*} A'_i| \\
&\leq k\varepsilon n + \sum_{i \leq i^*} \left|B \cap P_{\rho(i)}\right| + \sum_{j \in T} |B \cap P_j| + k\varepsilon n + |\cup_{i > i^*} A'_i| \\
&\leq 2k\varepsilon n + |B| + 2k\varepsilon n \leq 5k\varepsilon n.
\end{aligned}$$

Now since the vertices in $V_k = V(G) \setminus \cup_{i \leq k} P_i$ are distributed to the clusters in $\mathcal{P}$ to create the output clusters $\mathcal{C}$, we have that $|\mathcal{A}' \triangle \mathcal{C}| \leq |\mathcal{A}' \triangle \mathcal{P}| + |V_k|$. Observe that by the above analysis,

$$|V_k| \leq |B| + \sum_{i \leq i^*} |A'_i \setminus P_{\rho(i)}| + |\cup_{i > i^*} A'_i| \leq \varepsilon n + k\varepsilon n + 2k\varepsilon n \leq 4k\varepsilon n.$$

It follows that $|\mathcal{A}' \triangle \mathcal{C}| \leq 9k\varepsilon n$. The following concludes the proof:

$$|\mathcal{A} \triangle \mathcal{C}| \leq |\mathcal{A} \triangle \mathcal{A}'| + |\mathcal{A}' \triangle \mathcal{C}| < \varepsilon n + |\mathcal{A}' \triangle \mathcal{C}| \leq 10k\varepsilon n = O\left(\frac{\lambda_k \cdot d_{\mathsf{max}}^3 k^5 \cdot n}{\alpha_{\mathsf{in}}^2}\right).$$

□

5. Implementation in Practice

In this section, we show how to efficiently implement the greedy algorithm in Figure 1. To this end, we discuss how to quickly compute the first k eigenvectors and how to speed up the step of finding centers by random sampling.

5.1. Eigenvectors Computation

In general, the eigenvectors cannot be computed exactly in polynomial time as the entries may be irrational. However, in our application it is sufficient to have a set of vectors that well approximate the eigenvectors. To see this, we note that only the orthonormal property of eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$ and the fact that for $1 \leq i \leq k$, $\frac{\mathbf{v}_i \mathcal{L}_G \mathbf{v}_i^T}{\mathbf{v}_i \cdot \mathbf{v}_i^T} \leq \lambda_i$ are needed for all our previous results and analysis. Therefore, if we have a set of k orthonormal vectors $\mathbf{v'}_1, \cdots, \mathbf{v'}_k$ with $\frac{\mathbf{v'}_i \mathcal{L}_G \mathbf{v'}_i^T}{\mathbf{v'}_i \cdot \mathbf{v'}_i^T} \leq 2\lambda_i$ for $1 \leq i \leq k$, then our previous results still hold if we replace λ_k by $2\lambda_k$. On the other hand, such set of k orthonormal vectors can be computed efficiently as shown in the following folklore lemma, the proof of which follows from a repeated application of the near-linear time algorithm for computing the second eigenvector given by Spielman and Teng [29] and the variational characterization of eigenvalues (see e.g., Corollary 7.6.4 in [22]).

Lemma 5.1 (folklore). *There exists a procedure,* **ApproxEigen***, that takes an n-vertex, m-edge graph G, and an integer $k \leq n$, and returns k orthonormal vectors $\mathbf{v'}_1, \cdots, \mathbf{v'}_k \in \mathbb{R}^n$ such that $\frac{\mathbf{v'}_i \mathcal{L}_G \mathbf{v'}_i^T}{\mathbf{v'}_i \cdot \mathbf{v'}_i^T} \leq 2\lambda_i$, for $1 \leq i \leq k$. The running time of the procedure is $\tilde{O}((m+n)k)$.*

5.2. A Faster Algorithm

To further speed up the running time, we note that in each iteration i such that $1 \leq i \leq k$, the greedy algorithm in Figure 1 has to consider all vertices in V_{i-1} to determine the best center. This may cause the total time in these iterations to be as large as $\Omega(kn^2)$, which is slow in practice since n can be much larger than k. We show how to speed up this step via random sampling. The main observation is that we can get good center candidates by only computing the number of near-by vertices in the embedding, for vertices from a randomly chosen subset U_i of V_i, of size about $\Theta(\epsilon^{-1} k \log n)$, for any error parameter ϵ. This will reduce the computation time for finding the best centers from $\Theta(n^2k^2)$ to $O(\epsilon^{-1} nk^3 \log n)$. The procedure is summarized in Figure 2. The performance of the algorithm is given in the following theorem (that is similar to Theorem 2.1).

Theorem 5.2. *Let $\varepsilon > 0$. Let G be an n-vertex m-edge graph with maximum degree at most d_{max}. Let $k \geq 1$, and let $\mathcal{A} = \{A_1, \ldots, A_k\}$ be any $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong k-partition of $V(G)$ with $\alpha_{\mathsf{in}} > \max\{10(kd_{\mathsf{max}})^{3/2}, c\sqrt{k^5 d_{\mathsf{max}}^3}\} \cdot \sqrt{\frac{2\lambda_k}{\varepsilon}}$ for some sufficiently large constant c. Then, on input G, with high probability, the algorithm in Figure 2 outputs a partition $\mathcal{C}$ such that*

$$|\mathcal{A} \triangle \mathcal{C}| \leq \varepsilon n.$$

Furthermore, the running time of the algorithm is $\tilde{O}(mk + \varepsilon^{-1} k^3 n)$.

Proof sketch. Note that $\alpha_{\mathsf{in}} > 10(kd_{\mathsf{max}})^{3/2} \cdot \sqrt{2\lambda_k}$, and that by Lemma 5.1 the vectors $\mathbf{v'}_1, \cdots, \mathbf{v'}_k$ returned by **ApproxEigen** are orthonormal and satisfy $\frac{\mathbf{v'}_i \mathcal{L}_G \mathbf{v'}_i^T}{\mathbf{v'}_i \cdot \mathbf{v'}_i^T} \leq 2\lambda_i$, for $1 \leq i \leq k$. Thus, we can apply the argument of the proof of Lemma 4.1 on vectors $\mathbf{v'}_1, \cdots, \mathbf{v'}_k$ to find a collection $\mathcal{A}' = \{A'_1, \ldots, A'_k\}$ of pairwise disjoint subsets of $V(G)$, such

1. Let $\mathbf{v'}_1, \ldots, \mathbf{v'}_k$ be the returned vectors of the procedure **ApproxEigen**(G, k).
2. Let $\mathbf{F} : V(G) \to \mathbb{R}^k$, where $\mathbf{F}(u) = \deg(u)^{-1/2} (\mathbf{v'}_1(u), \ldots, \mathbf{v'}_k(u))$.
3. $R = \frac{1}{36kd_{\mathsf{max}}\sqrt{n}}$
4. $V_0 = V(G)$
5. for $i = 1, \ldots, k$
6. Sample uniformly with repetition a subset $U_{i-1} \subseteq V_{i-1}$, $|U_{i-1}| = \Theta(\varepsilon^{-1}k \log n)$.
7. $u_i = \text{argmax}_{u \in U_{i-1}} |\mathsf{ball}(\mathbf{F}(u), 2R) \cap \mathbf{F}(V_{i-1})|$
$= \text{argmax}_{u \in U_{i-1}} |\{w \in V_{i-1} : \|\mathbf{F}(u) - \mathbf{F}(w)\|_2 \leq 2R\}|$
8. $P_i = \mathbf{F}^{-1}(\mathsf{ball}(\mathbf{F}(u_i), 2R)) \cap V_{i-1}$
9. $V_i = V_{i-1} \setminus P_i$
10. Let $\mathbf{g} : V_k \to \{u_1, \cdots, u_k\}$, $\mathbf{g}(v) = u_i$ where i is the smallest index satisfying $\|\mathbf{F}(v) - \mathbf{F}(u_i)\| \leq \|\mathbf{F}(v) - \mathbf{F}(u_j)\|$ for all $j \neq i$.
11. Return $\{C_1, \ldots, C_k\} = \{P_1 \cup \mathbf{g}^{-1}(u_1), \ldots, P_k \cup \mathbf{g}^{-1}(u_k)\}$

Figure 2: The fast spectral k-clustering algorithm takes an n-vertex graph G with maximum degree at most d_{max}, and an $\varepsilon > 0$ as input, and returns a partition $\mathcal{C} = \{C_1, \ldots, C_k\}$ of $V(G)$.

that $|\mathcal{A} \triangle \mathcal{A}'|/n \leq O(\frac{\lambda_k \cdot d_{\mathsf{max}}^3 k^4}{\alpha_{\mathsf{in}}^2})$. Since c is sufficiently large and $\alpha_{\mathsf{in}} \geq c\sqrt{k^5 d_{\mathsf{max}}^3} \cdot \sqrt{\frac{2\lambda_k}{\varepsilon}}$, we can guarantee that $|\mathcal{A} \triangle \mathcal{A}'|/n \leq \frac{\varepsilon}{10k}$. Let $\varepsilon' = \frac{\varepsilon}{10k}$. Similar to the proof of Theorem 2.1, we define a function $\rho : \{1, \ldots, k\} \to \{1, \ldots, k\}$ that maps clusters in $\mathcal{A}'$ to clusters in $\mathcal{C}$, where $\mathcal{C} = \{C_1, \ldots, C_k\}$ are the ordered collection of pairwise disjoint subsets of $V(G)$ output by the greedy spectral k-clustering algorithm in Figure 2. Now note that for any $i \leq k$, a vertex from a cluster A'_i of size at least $2\varepsilon' n$ will be sampled out with probability at least $1 - \frac{1}{n^2}$. This implies that with high probability the following holds: for each $i \in \{1, \ldots, k\}$, if $|A'_i| \geq 2\varepsilon' n$, then $\rho(i)$ exists, since the vertices in the set U_{i-1} of the ith iteration of the algorithm are sampled uniformly at random. The rest of the argument for the correctness of the algorithm is identical to the proof of Theorem 2.1, and eventually, we can guarantee that $|\mathcal{A} \triangle \mathcal{C}| \leq 10k\varepsilon' n = \varepsilon n$.

For the running time of the algorithm, note that by Lemma 5.1, the procedure **Approx-Eigen** takes time $\tilde{O}((m+n)k)$. In each iteration, we need to sample $O(\varepsilon^{-1}k \log n)$ vertices, and for each sampled vertex we need to determine the number of near-by vertices in the embedding, which takes $O(nk)$ time, as each vertex corresponds to a point in $\mathbb{R}^k$. This means that the computation in all the k iterations takes time $O(\varepsilon^{-1}nk^3 \log n)$. Therefore, the total running time of the algorithm is thus $\tilde{O}((m+n)k) + O(\varepsilon^{-1}nk^3 \log n) = \tilde{O}(mk + \varepsilon^{-1}nk^3)$. □

In practice it is common to work with fixed k. We note that the fast randomized algorithm runs in near-linear time for $k = O(\text{polylog}(n))$, provided that the user specifies $\varepsilon = \Omega(1/\text{polylog}(n))$.

6. Conclusion

In this paper we have presented a very simple spectral clustering algorithm that provably approximates any good partition of an input graph, provided that one exists. Further, our preliminary experimental results given in Appendix C indicate that the algorithm gives meaningful output even when the spectral gap condition is much smaller than what our theorems require. This provides some evidence that stronger theoretical guarantees may be obtainable, possibly by weakening the gap condition. It is also natural to wonder how small of a separation is necessary for good performance of the algorithm. We currently know of no such lower bounds. We believe that these are interesting research directions.

One interesting remark is that a qualitatively similar result can be obtained for weighted graphs by introducing the input graph's minimum edge weight, w_{min}, as a lower bound for the (now weighted) degree in Lemma 3.2 and in Claim 4.4. This results in an additional factor of w_{min}^{-3} in the corresponding analog of Theorem 2.1, provided that R of Figure 1 is also scaled by a factor of $\sqrt{w_{\mathsf{min}}}$.

Unfortunately, due to the appearance of w_{min} in the denominator, the result of the previous paragraph is unstable under perturbation of the input graph by a low-weight edge. This instability appears to be an artifact of the analysis. For instance we can show that given any $(\alpha_{\mathsf{in}}, \alpha_{\mathsf{out}})$-strong partition $\mathcal{A}$, there exists a choice of R such that the resulting clustering is stable under perturbation by a low-weight edge. Such a result essentially follows by replacing w_{min} with δ_{min}, the minimum weighted vertex degree among all induced graphs $G[A]$ for $A \in \mathcal{A}$. The obstacle with turning this into an algorithm is that to obtain the corresponding error guarantee we must scale R (as it appears in Figure 1) by a factor of $\sqrt{\delta_{\mathsf{min}}}$, which is not known a priori. It remains an open problem to give a similar algorithm for weighted graphs which is stable under perturbation by a low-weight edge.

Acknowledgments

The authors wish to thank James R. Lee for bringing to their attention a result from the latest version of [17]. This work was partially supported by the NSF grants CCF 1318595 and CCF 1423230. The second author acknowledges the support of ERC grant No. 307696 and a 973 program Grant No. 2014CB340302.

References

[1] Zeyuan Allen Zhu, Silvio Lattanzi, and Vahab S. Mirrokni. A local algorithm for finding well-connected clusters. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, volume 28 of *JMLR Workshop and Conference Proceedings*, pages 396–404. JMLR.org, 2013.

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

[3] Noga Alon and V. D. Milman. lambda$_1$, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.

[4] Charles J. Alpert and So-Zen Yao. Spectral partitioning: The more eigenvectors, the better. In Bryan Preas, editor, *Proceedings of the 32st Conference on Design Automation, San Francisco, California, USA, Moscone Center, June 12-16, 1995.*, pages 195–200. ACM Press, 1995.

[5] Pranjal Awasthi, Moses Charikar, Ravishankar Krishnaswamy, and Ali Kemal Sinop. The hardness of approximation of euclidean k-means. In Lars Arge and János Pach, editors, *31st International Symposium on Computational Geometry, SoCG 2015, June 22-25, 2015, Eindhoven, The Netherlands*, volume 34 of *LIPIcs*, pages 754–767. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.

[6] Pranjal Awasthi, Moses Charikar, Ravishankar Krishnaswamy, and Ali Kemal Sinop. Spectral embedding of k-cliques, graph partitioning and k-means. In Madhu Sudan, editor, *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, Cambridge, MA, USA, January 14-16, 2016*, pages 301–310. ACM, 2016.

[7] Sivaraman Balakrishnan, Min Xu, Akshay Krishnamurthy, and Aarti Singh. Noise thresholds for spectral clustering. In John Shawe-Taylor, Richard S. Zemel, Peter L. Bartlett, Fernando C. N. Pereira, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain.*, pages 954–962, 2011.

[8] Punyashloka Biswal, James R. Lee, and Satish Rao. Eigenvalue bounds, spectral partitioning, and metrical deformations via flows. *J. ACM*, 57(3):13:1–13:23, 2010.

[9] Jeff Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in Analysis*, pages 195–199. Princeton Univ. Press, Princeton, NJ, 1970.

[10] Fan R.K. Chung. *Spectral Graph Theory*. Number 92 in CBMS Regional Conference Series. Conference Board of the Mathematical Sciences, 1997.

[11] Artur Czumaj, Pan Peng, and Christian Sohler. Testing cluster structure of graphs. In Rocco A. Servedio and Ronitt Rubinfeld, editors, *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 723–732. ACM, 2015.

[12] Tamal K. Dey, Alfred Rossi, and Anastasios Sidiropoulos. Spectral concentration, robust k-center, and simple clustering. *CoRR*, abs/1404.1008, 2014.

[13] Shayan Oveis Gharan and Luca Trevisan. Partitioning into expanders. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1256–1266. SIAM, 2014.

[14] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, 2004.

[15] Jonathan A. Kelner. Spectral partitioning, eigenvalue bounds, and circle packings for graphs of bounded genus. *SIAM J. Comput.*, 35(4):882–902, 2006.

[16] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 11–20. ACM, 2013.

[17] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1117–1130. ACM, 2012.

[18] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1131–1140. ACM, 2012.

[19] Milena Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. In *30th Annual Symposium on Foundations of Computer Science, Research Triangle Park, North Carolina, USA, 30 October - 1 November 1989*, pages 526–531. IEEE Computer Society, 1989.

[20] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani, editors, *Advances in Neural Information Processing Systems 14 [Neural Information Processing Systems: Natural and Synthetic, NIPS 2001, December 3-8, 2001, Vancouver, British Columbia, Canada]*, pages 849–856. MIT Press, 2001.

[21] Lorenzo Orecchia and Zeyuan Allen Zhu. Flow-based algorithms for local graph clustering. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1267–1286. SIAM, 2014.

[22] Shayan Oveis Gharan. *New Rounding Techniques for the Design and Analysis of Approximation Algorithms*. PhD thesis, Stanford University, 2014.

[23] Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs: Spectral clustering works! In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, volume 40 of *JMLR Workshop and Conference Proceedings*, pages 1423–1455. JMLR.org, 2015.

[24] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, 1989.

[25] Ali Kemal Sinop. How to round subspaces: A new spectral clustering algorithm. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1832–1847. SIAM, 2016.

[26] Speech and University of Eastern Finland Image Processing Unit, School of Computing. Shape sets. `http://cs.joensuu.fi/sipu/datasets/`.

[27] Daniel A. Spielman and Shang-Hua Teng. Disk packings and planar separators. In Sue Whitesides, editor, *Proceedings of the Twelfth Annual Symposium on Computational Geometry, Philadelphia, PA, USA, May 24-26, 1996*, pages 349–358. ACM, 1996.

[28] Daniel A. Spielman and Shang-Hua Teng. Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra and its Applications*, 421(2):284–305, 2007.

[29] Daniel A. Spielman and Shang-Hua Teng. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM J. Matrix Analysis Applications*, 35(3):835–885, 2014.

[30] Mamoru Tanaka. Multi-way expansion constants and partitions of a graph. *arXiv.org*, December 2011.

[31] Terence Tao. *Topics in Random Matrix Theory.* Graduate studies in mathematics. American Mathematical Soc., 2012.

[32] Alfred Ultsch. Clustering with SOM: U*C. In *Proc. Workshop on Self-Organizing Maps*, pages 75–82, Paris, France, 2005.

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

Appendix A. Missing Proofs from Section 3

Proof of Lemma 3.2. For any $i \leq k$,

$$\frac{\mathbf{v}_i \mathcal{L}_G \mathbf{v}_i^T}{\mathbf{v}_i \mathbf{v}_i^T} = \frac{\mathbf{x}_i \mathbf{D}^{1/2} \mathcal{L}_G \mathbf{D}^{1/2} \mathbf{x}_i^T}{\mathbf{x}_i \mathbf{D} \mathbf{x}_i^T} = \frac{\sum_{(u,v)\in E} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2}{\sum_{u \in V} \deg(u) \mathbf{x}_i^2(u)} = \lambda_i \leq \lambda_k \ .$$

This further gives that

$$\sum_{(u,v)\in E} (\mathbf{x}_i(u) - \mathbf{x}_i(v))^2 \leq \lambda_k, \tag{A.1}$$

since $\sum_{u\in V}\deg(u)\mathbf{x}_i^2(u)=\sum_{u\in V}\mathbf{v}_i^2(u)=1$. Let us recall a known result (see, e.g., [10, (1.5), p. 5 and (1.14), p. 13]) that for any graph $H=(V_H,E_H)$,[2]

$$\lambda_2(H)=\mathrm{vol}_H(V_H)\cdot\min_{\mathbf{f}}\left\{\frac{2\cdot\sum_{(u,v)\in E_H}(\mathbf{f}(u)-\mathbf{f}(v))^2}{\sum_{u,v\in V_H}(\mathbf{f}(u)-\mathbf{f}(v))^2\deg_H(u)\deg_H(v)}\right\}\ , \tag{A.2}$$

where $\lambda_2(H)$ denotes the second smallest eigenvalue of the normalized Laplacian of H. Let us consider the induced subgraph $H:=G[C]$ on C. Since $\phi(H)\geq\phi_{\mathsf{in}}$, Cheeger's inequality yields $\lambda_2(H)\geq\frac{\phi_{\mathsf{in}}^2}{2}$. Therefore, if we apply this bound to inequality (A.2), then,

$$\mathrm{vol}_H(V_H)\cdot\frac{2\cdot\sum_{(u,v)\in E_H}(\mathbf{x}_i(u)-\mathbf{x}_i(v))^2}{\sum_{u,v\in V_H}(\mathbf{x}_i(u)-\mathbf{x}_i(v))^2\deg_H(u)\deg_H(v)}\geq\lambda_2(H)\geq\frac{\phi_{\mathsf{in}}^2}{2}\ .$$

Combining this with the fact that

$$\sum_{(u,v)\in E_H}(\mathbf{x}_i(u)-\mathbf{x}_i(v))^2\leq\sum_{(u,v)\in E_G}(\mathbf{x}_i(u)-\mathbf{x}_i(v))^2\leq\lambda_k,$$

where the last inequality follows from inequality (A.1), we have that

$$\sum_{u,v\in V_H}(\mathbf{x}_i(u)-\mathbf{x}_i(v))^2\deg_H(u)\deg_H(v)\leq\frac{4\lambda_k\mathrm{vol}_H(V_H)}{\phi_{\mathsf{in}}^2}\ .$$

Next, since $\phi(H)\geq\phi_{\mathsf{in}}>0$ implies that $\deg_H(u)\geq1$ for all $u\in V_H$. Using the bound above we obtain:

$$\begin{aligned}\sum_{u,v\in V_H}(\mathbf{x}_i(u)-\mathbf{x}_i(v))^2 &\leq \sum_{u,v\in V_H}(\mathbf{x}_i(u)-\mathbf{x}_i(v))^2\deg_H(u)\deg_H(v)\\ &\leq \frac{4\lambda_k\mathrm{vol}_H(V_H)}{\phi_{\mathsf{in}}^2}\leq\frac{4\lambda_k\mathrm{vol}(C)}{\phi_{\mathsf{in}}^2}.\end{aligned}$$

This completes the proof of the lemma. □

Appendix B. Missing Proofs from Section 4

Proof of Fact 4.2. Since $\max_{i\leq p}\|\mathbf{X}_{\mathrm{row}(i)}-\mathbf{Y}_{\mathrm{row}(i)}\|_2\leq\delta$, then the Frobenius norm $\|\mathbf{X}-\mathbf{Y}\|_F$ of $\mathbf{X}-\mathbf{Y}$ is at most $\sqrt{p}\cdot\delta$, and therefore, the induced 2-norm $\|\mathbf{X}-\mathbf{Y}\|_2$ of $\mathbf{X}-\mathbf{Y}$ is at most $\sqrt{p}\cdot\delta$. By Weyl's inequality [31], for any i, $|\eta_i(\mathbf{X})-\eta_i(\mathbf{Y})|\leq|\eta_1(\mathbf{X}-\mathbf{Y})|\leq\|\mathbf{X}-\mathbf{Y}\|_2\leq\sqrt{p}\cdot\delta$. □

[2]We remark that in [10], the summation in the denominator is over all unordered pairs of vertices, while in our context, the summation is over all possible $|V_H|^2$ vertex pairs. Therefore, a multiplicative factor 2 appears in the numerator in Equation (A.2) compared to the form in [10, (1.5), p. 5].

Proof of Fact 4.3. For simplicity, let $\mathbf{X}_i := \mathbf{X}_{\mathrm{row}(i)}$ and $\mathbf{Y}_i := \mathbf{Y}_{\mathrm{row}(i)}$. Note that the i,jth entry of $\mathbf{X} \cdot \mathbf{X}^T$ and $\mathbf{Y} \cdot \mathbf{Y}^T$ are $\langle \mathbf{X}_i, \mathbf{X}_j \rangle$ and $\langle \mathbf{Y}_i, \mathbf{Y}_j \rangle$, respectively, and that

$$\begin{aligned} |\langle \mathbf{X}_i, \mathbf{X}_j \rangle - \langle \mathbf{Y}_i, \mathbf{Y}_j \rangle| &= |\langle \mathbf{X}_i, \mathbf{X}_j \rangle - \langle \mathbf{Y}_i - \mathbf{X}_i + \mathbf{X}_i, \mathbf{Y}_j - \mathbf{X}_j + \mathbf{X}_j \rangle| \\ &\leq |\langle \mathbf{Y}_i - \mathbf{X}_i, \mathbf{Y}_j - \mathbf{X}_j \rangle| + |\langle \mathbf{Y}_i - \mathbf{X}_i, \mathbf{X}_j \rangle| + |\langle \mathbf{X}_i, \mathbf{Y}_j - \mathbf{X}_j \rangle| \\ &\leq \|\mathbf{Y}_i - \mathbf{X}_i\|_2 \|\mathbf{Y}_j - \mathbf{X}_j\|_2 + \|\mathbf{Y}_i - \mathbf{X}_i\|_2 \|\mathbf{X}_j\|_2 + \|\mathbf{X}_i\|_2 \|\mathbf{Y}_j - \mathbf{X}_j\|_2 \\ &\leq \delta^2 + 2\gamma\delta. \end{aligned}$$

Therefore, $\max_{i \leq p} \|(\mathbf{X} \cdot \mathbf{X}^T)_{\mathrm{row}(i)} - (\mathbf{Y} \cdot \mathbf{Y}^T)_{\mathrm{row}(i)}\|_2 \leq \sqrt{p}(\delta^2 + 2\gamma\delta)$. □

Appendix C. Experimental Evaluation

Results from our greedy k-clustering implementation are shown in Figures C.3, C.4, C.6. Cluster assignments for graphs are shown as colored nodes. In the case where the graph comes from a triangulated surface, we have extended the coloring to a small surface patch in the vicinity of the node. Each experiment includes a plot of the eigenvalues of the normalized Laplacian (y-axis), by eigenvector number (x-axis). A small rectangle on each plot highlights the corresponding spectral gap between k and $k+1$.

Multiple spectral gaps. Recall that graphs which have a sufficiently large spectral gap between the k-th and $(k+1)$-st eigenvalues admit a strong clustering [13]. Figure C.3 shows the result of our algorithm on a graph with two prominent spectral gaps, $k = 2$ (left) and $k = 5$ (right).

This graph is sampled from the following generative model. Let $C_1, \ldots, C_5$ be disjoint vertex sets of equal size, depicted as circles. Every edge with both endpoints in the same C_i appears with probability p_1, every edge between $C_1 \cup C_2$ and $C_3 \cup C_4 \cup C_5$ appears with probability p_3, and every other edge appears with probability p_2, for some $p_1 \gg p_2 \gg p_3$. The resulting graph admits a strong k-partition, for any $k \in \{2, 5\}$. This fact is reflected in the output of our algorithm.

We remark that to achieve a sufficiently large gap at $k = 5$ many intra-circle edges are necessary, which makes the resulting figures too dense. To make the plots readable, we have displayed only a subsampling of these edges.

Comparison with k-means. In Figure C.5 we compare the greedy approach to k-means on the spectral embedding. Our experiments build a graph on two and three-dimensional euclidean point-cloud data by selecting a threshold value, D, and connecting any two points which are no further than D. Additionally, any singletons are removed. A key feature of the chosen point-cloud data is that it comes with ground truth labeling which we lift to the corresponding graph.

[3] Only the largest component was used for clustering.

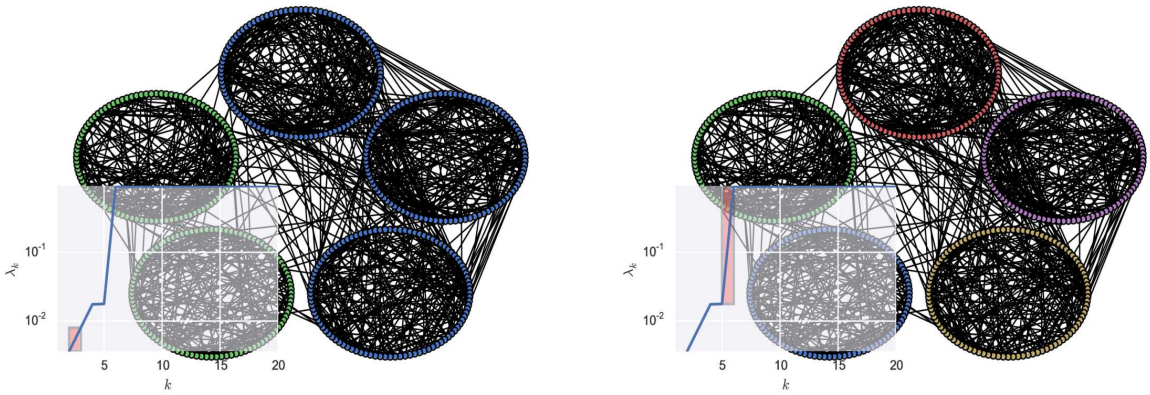

Figure C.3: Clustering a graph with multiple spectral gaps. The graph has $n = 500$ nodes, with edges chosen according to our model below. Clustering is performed with $R = 18/(kd_{\max}\sqrt{n})$ for $k = 2$ and $k = 5$.

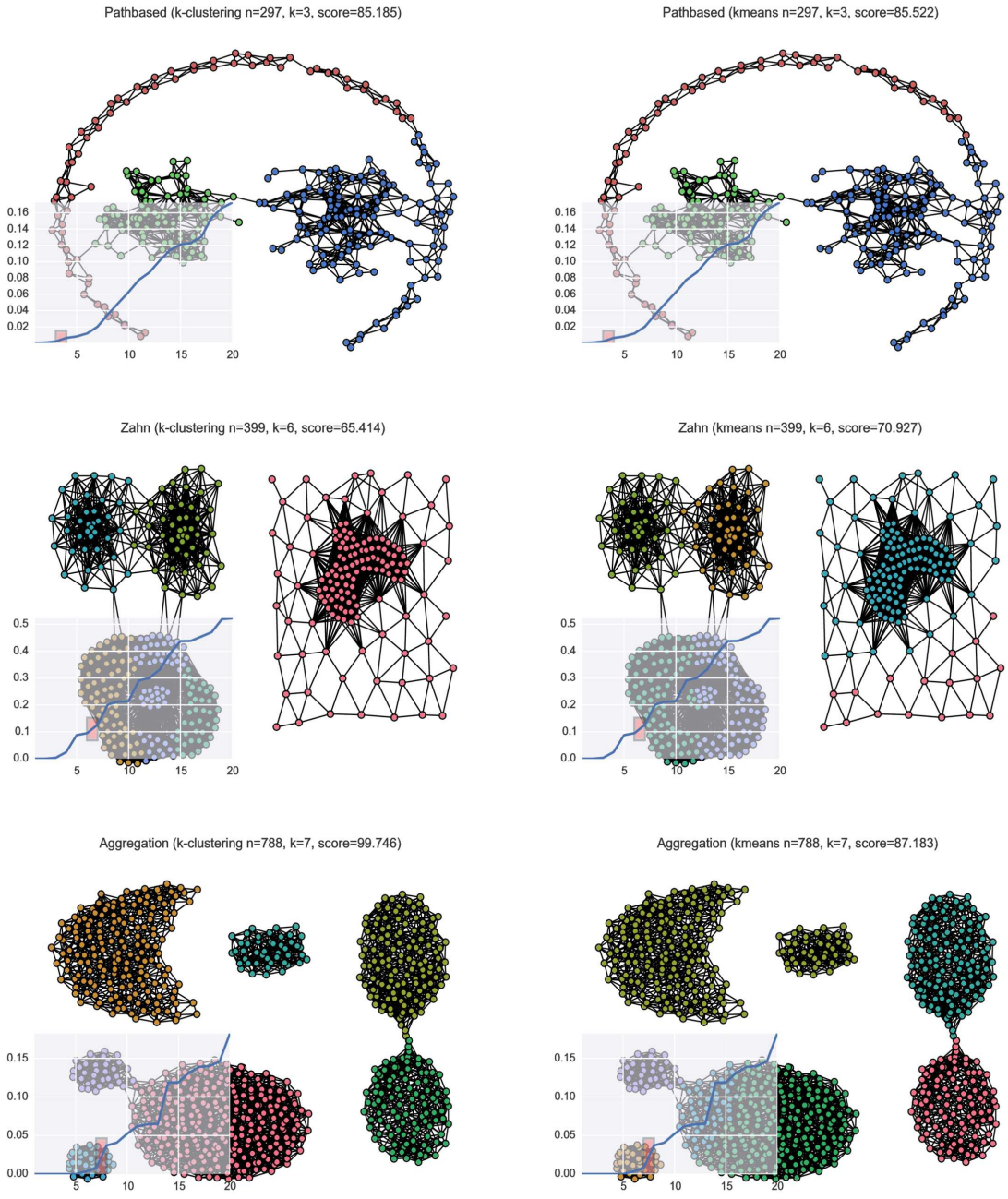

Figure C.4: Visual comparison of cluster assignments between Greedy k-Center and k-means. Clustering was performed on graphs with radius $R = 18/(kd_{\max}\sqrt{n})$. The scores in the titles indicate percent correct classification with respect to the ground truth.

Data set	n	k	D	Greedy k-C	k-means
pathbased	297	3	2	85.185	85.522
jain	373	2	3	100	100
zahn	399	6	3	65.414	71.471
aggregation	788	7	2	99.746	96.063
LSun	400	3	1	93.250	93.250
Tetra	400	4	1	100	100
Hepta	212	7	5	44.811	57.849
Chainlink	1000	2	1	80.600	83.196
EngyTime[3]	4082	2	0.5	96.546	96.423
TwoDiamonds	800	2	1	100	100

Figure C.5: Percentage agreement to ground truth for Greedy k-Clustering and k-means. The results reported in the k-means column are the mean percent agreement over 100 trials where k-means was initialized by random selection of k points in the embedding. The R parameter for Greedy k-Clustering was given by the formula $R = 18/(kd_{\mathsf{max}}\sqrt{n})$. The first four input graphs are based on data sets of the same name from [26], the rest are based on similarly named point-cloud data from FCPS. [32].

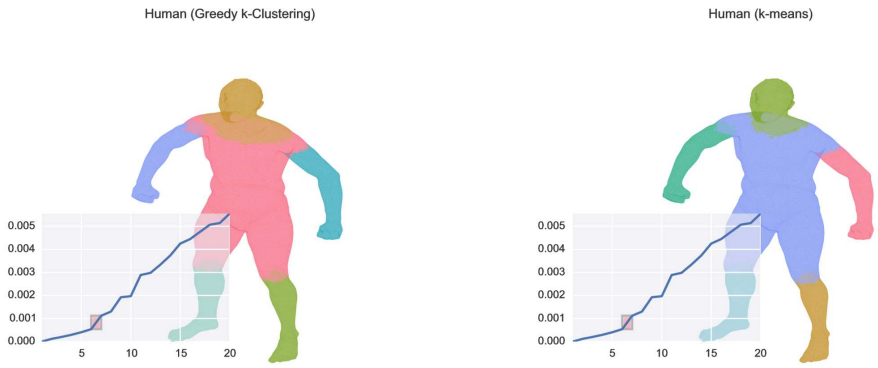

Figure C.6: A $k = 6$ clustering performed on the 1-skeleton of triangulated mesh ($n = 12500$), using $R = 30/(kd_{\mathsf{max}}\sqrt{n})$. Here the two algorithms agree on 97.176% of the vertices.