Spectral Clustering Oracles in Sublinear Time

Grzegorz Gluch
EPFL

Michael Kapralov
EPFL

Silvio Lattanzi
Google Research

Aida Mousavifar
EPFL

Christian Sohler
University of Cologne *

Abstract

Given a graph G that can be partitioned into k disjoint expanders with outer conductance upper bounded by $\epsilon \ll 1$, can we efficiently construct a small space data structure that allows quickly classifying vertices of G according to the expander (cluster) they belong to? Formally, we would like an efficient local computation algorithm that misclassifies at most an $O(\epsilon)$ fraction of vertices in every expander. We refer to such a data structure as a *spectral clustering oracle*.

Our main result is a spectral clustering oracle with query time $O^*(n^{1/2+O(\epsilon)})$ and preprocessing time $2^{O(\frac{1}{\epsilon}k^4 \log^2(k))} n^{1/2+O(\epsilon)}$ that provides misclassification error $O(\epsilon \log k)$ per cluster for any $\epsilon \ll 1/\log k$. More generally, query time can be reduced at the expense of increasing the preprocessing time appropriately (as long as the product is about $n^{1+O(\epsilon)}$) – this in particular gives a nearly linear time spectral clustering primitive.

The main technical contribution is a sublinear time oracle that provides dot product access to the spectral embedding of G by estimating distributions of short random walks from vertices in G. The distributions themselves provide a poor approximation to the spectral embedding, but we show that an appropriate linear transformation can be used to achieve high precision dot product access. We give an estimator for this linear transformation and analyze it using spectral perturbation bounds and a novel upper bound on the leverage scores of the spectral embedding matrix of a k-clusterable graph. We then show that dot product access to the spectral embedding is sufficient to design a clustering oracle. At a high level our approach amounts to hyperplane partitioning in the spectral embedding of G, but crucially operates on a nested sequence of carefully defined subspaces in the spectral embedding to achieve per cluster recovery guarantees.

*Work was partially done while author was visiting researcher at Google Research, Switzerland

Contents

1 Introduction

As a central problem in unsupervised learning, graph clustering has been extensively studied in the past decades. Several formalizations of the problem have been considered in the literature. In this paper, we focus on the following (informal) variant of graph clustering: Given a graph G and an integer k, we are interested in finding k nonoverlapping sets $C_1, C_2, \ldots, C_k$ that are internally well-connected and that have a sparse cut to the outside. A popular approach to this problem is spectral clustering [KVV04, NJW02, SM00, VL07]: One embeds vertices of the graph into k dimensional Euclidean space using the bottom k eigenvectors of the Laplacian, and clusters the points in Euclidean space using the k-means algorithm (in practice), or using a more careful space partitioning approach (in theory). Spectral clustering has been applied in the context of a wide variety of problems, for example, image segmentation [SM00], speech separation [BJ06], clustering of protein sequences [PCS06], and predicting landslides in geophysics [BMD$^+$15]. Spectral clustering usually requires to process the graph in two steps. First one computes the spectral embedding and then one clusters the resulting point set. This two stage approach seems to be highly non-local and it seems to be hard to obtain faster methods, if one only has to determine the cluster membership for a small subset of the vertices. However, such a sublinear time access is desirable in some applications. As a basic step towards such a sublinear time clustering algorithm, we need a way to quickly access the spectral embedding in some way. Therefore, we ask the following question, where we use $f_x \in \mathbb{R}^k$ to denote the spectral embedding of vertex x:

> Is it possible to obtain dot product access to the spectral embedding of a graph in sublinear time? In other words, given a pair of vertices $x, y \in V$, can we quickly approximate the dot product $\langle f_x, f_y \rangle$ in $o(n)$ time?

If such access is possible, it appears plausible that one can design a *sublinear spectral clustering oracle*, a small space data structure that provides fast query access to a good clustering of the graph. Our main result in this paper is **(a)** a small space data structure that provides query access to dot products in the spectral embedding, as above, and **(b)** a sublinear time spectral clustering oracle that uses this data structure.

We study a popular version of the spectral clustering problem where one assumes the existence of a planted solution, namely that the input graph can be partitioned into clusters $C_1, \ldots, C_k$ whose internal connectivity is nontrivially higher than the external connectivity. The goal is to recover the clusters approximately. An average case version of this problem, where the clusters induce Erdős-Rényi graphs (or random regular graphs), and the edges across clusters are similarly random, has been studied extensively in the literature on the stochastic block model (SBM) [Abb18] for its close relationship to the community detection problem. In this work we study a worst-case version of this problem:

> Given a graph $G = (V, E)$ that admits a partitioning into a disjoint union of k induced expanders $C_1, \ldots, C_k$ with outer conductance bounded by $\epsilon \ll 1$, output an approximation to $C_1, \ldots, C_k$ that is correct up to a $O(\epsilon)$ error **on every cluster**.

We define a *spectral clustering oracle with per cluster error* $\delta \in (0, 1)$ as a small space data structure that implicitly defines disjoint subsets $\widehat{C}_1, \ldots, \widehat{C}_k$ of V such that for some permutation π on k elements one has $|C_i \Delta \widehat{C}_{\pi(i)}| \leq \delta |C_i|$ for every $i = 1, \ldots, k$. The oracle must provide fast query access to such a clustering. The focus of this paper is:

> Design a sublinear time spectral clustering oracle with per cluster error $\approx O(\epsilon)$.

Our main result is a spectral clustering oracle as above, with a slight loss in error parameter. Specifically, our spectral clustering oracle is correct up to $O(\epsilon \log k)$ error on every cluster:

Theorem 1 (Informal)**.** *There exists a spectral clustering oracle that for every graph $G = (V, E)$ that admits a partitioning into a disjoint union of k induced expanders $C_1, \ldots, C_k$ with outer conductance* *bounded by $\epsilon \ll \frac{1}{\log k}$ achieves error $O(\epsilon \log k)$ per cluster, query time $\approx n^{1/2+O(\epsilon)}$, preprocessing time $\approx 2^{O(\frac{1}{\epsilon}k^4 \log^2(k))} n^{1/2+O(\epsilon)}$ and space $\approx n^{1/2+O(\epsilon)}$.*

Query times can be made faster at the expense of increased space and prepropcessing time, as long as the product of query time and preprocessing time is $\approx n^{1+O(\epsilon)}$, leading in particular to a nearly linear time algorithm for spectral clustering.

As byproduct of our main result we also obtain new efficient clustering algorithms in the Local Computation Algorithms (LCA) model (see [RTVX11] for introduction of the model and [ARVX12] for LCA with limited randomness).

A very important feature of the problem above is the fact that our algorithms recovers a $1-O(\epsilon \log k)$ fraction of **every cluster** as opposed to just classifying a $1 - O(\epsilon \log k)$ fraction of vertices of the graph correctly (this latter question allows one to output fewer than k clusters, and is much easier to solve). To put this in perspective, it is instructive to apply multiway Cheeger inequalities (e.g., [LGT14], [CKCLL$^+$13]) to our setting, noting that the k-th eigenvalue λ_k of the normalized Laplacian of a graph that can be partitioned into k clusters as above is bounded by $O(\epsilon)$. This means that multiway Cheeger inequalities can be used to recover k clusters with outer conductance $k^2\sqrt{\epsilon}$ (see [LGT14]), which becomes trivial unless $\epsilon < 1/k^4$ (we note that our problem admits a much simpler solution when $\epsilon \ll 1/k$). One may note that multiway Cheeger inequalities can also recover $0.9k$ clusters with outer conductance $\log^{O(1)} k\sqrt{\epsilon}$ in our setting (e.q. [LRTV12]), but, as mentioned above, recovering most clusters is much easier that recovering each cluster to $1 \pm O(\epsilon)$ multiplicative error, and does not solve our problem. The most relevant prior result is due to Sinop [Sin16], where the author achieves error $O(\sqrt{\epsilon})$ per cluster using spectral techniques. Sinop's result improves up on previous work of [AS12], which achieved per cluster error of $O(\epsilon k)$ (or, rather, is somewhat incomparable to [AS12] due to the worst dependence on ϵ, but a lack of dependence on k). As we argue below, Sinop's techniques are hard to extend to the sublinear time regime. At the same time, one should note that our result improves on [AS12] under the assumption that cluster sizes are comparable while using only sublinear time in the size of the input graph.

Main challenges and comparison to results on testing cluster structure. This problem is related the well-studied expansion testing problem [KS08, NS10, GR11, CS10, KPS13], which corresponds to the setting of one or two clusters, as well as to the problem of testing cluster structure of graphs, where one essentially wants to determine k, the number of clusters in G. The problem of *testing* cluster structure has recently been considered in the literature [CPS15, CKK$^+$18]: given access to a graph G as above, compute the value of k (in fact, both results [CPS15] and [CKK$^+$18] apply to the harder property testing problem of distinguishing between graphs that are k-clusterable according to the definition above and graphs that are ϵ-far from k-clusterable, but a procedure for computing k is the centerpiece of both results). It is interesting to note that the work of [CPS15] also yields an algorithm for our problem, but only under very strong assumptions on the outer conductance of the clusters (one needs $\epsilon \ll \frac{1}{\text{poly}(k)\log n}$). The recent work of Peng [Pen20] considers a robust version of testing cluster structure, but requires $\epsilon \ll \frac{1}{\text{poly}(k)\log n}$, just like the work of [CPS15].

The recent work of [CKK$^+$18] on testing cluster structure yields an optimal tester, which works for any ϵ smaller than a constant and achieves essentially optimal runtime, but unfortunately their techniques do no extend to the 'learning' version of the problem. The reason is very simple: the algorithm of [CKK$^+$18] needs to distinguish between the graph G being a union of k clusters and $k+1$ clusters, and their approach amounts to verifying whether a graph can be partitioned into k clusters. To do so it suffices to check whether the spectral embedding is effectively k-dimensional, i.e. whether it spans a nontrivial $(k+1)$-dimensional volume. In order to certify this, however, it suffices to exhibit $k+1$ vertices that span a nontrivial $(k+1)$-dimensional volume. For that, one essentially only needs to locate at least one 'typical' point in every cluster, which is much easier than our task of correctly recovering almost all, i.e. a $1-O(\epsilon)$ fraction of vertices in every cluster. In other words, testing graph cluster structure requires only a rather basic access to and control of the spectral embedding. The main technical contribution of our paper is a set of tools for getting precise dot product access to this embedding, together with several new structural claims about it that enable our clustering algorithm.

Comparison to the work of Sinop [Sin16]. The work of Sinop [Sin16] gives a nearly linear time algorithm for recovering every cluster up to error of $1 \pm O(\sqrt{\epsilon})$ using spectral techniques[1], for sufficiently

[1]One must note that the work of [Sin16] does not require the bounded degree assumption, and can handle clusters of significantly different size.

small ϵ. The algorithm would be very hard to implement in sublinear time, since one of its central tools (the Round procedure, which controls propagation of error i.e., Lemma 5.4 of [Sin16]) heavily relies on the ability to have explicit access to the eigendecomposition of the Laplacian. Specifically, Sinop's algorithm first finds a crude approximation S to a cluster to be recovered, and then improves the approximation by explicitly constructing the corresponding submatrix of the spectral embedding and performing an SVD. One could plausibly envision implementing this using random walks, but that would be challenging, since one would need to consider a random walk induced on a rather unstructured subset of vertices of the graph.

Our contributions: sublinear time access to the spectral embedding. Let $G = (V, E)$ be a d-regular graph with $n = |V|$. Without loss of generality we assume that $V = \{1, \ldots, n\}$. We assume that n and d are given to the algorithm and that we have oracle access to G: We can specify a vertex $x \in V$ and a number $i, 1 \leq i \leq d$, and we will be given in constant time the i-th neighbor of x. This is also called the bounded degree graph model.

In this paper we will consider d-regular graphs that have a certain cluster structure. We parameterize this cluster structure using the internal and external conductance parameters.

Definition 1 (**Internal and external conductance**)**.** Let $G = (V, E)$ be a graph. For a set $S \subseteq C \subseteq V$, let $E(S, C \setminus S)$ be the set of edges with one endpoint in S and the other in $C \setminus S$. The *conductance of a set S within C* is $\phi^G_C(S) = \frac{|E(S,C\setminus S)|}{d|S|}$. The *external-conductance* of set C is defined to be $\phi^G_V(C) = \frac{|E(C,V\setminus C)|}{d|C|}$. The *internal-conductance* of set $C \subseteq V$, denoted by $\phi^G(C)$, is

$$\min_{S\subseteq C, 0<|S|\leq \frac{|C|}{2}} \phi^G_C(S)$$

if $|C| > 1$ and one otherwise.

Remark 1. *For simplicity we present all the proofs for d-regular graphs, even though all the proofs also work for d-bounded graphs, with the same definition of conductance as in Definition 1 (i.e., with normalization by $d|S|$ as opposed to the volume of S; the two notions of conductance can in the worst case differ by a factor of d). Note that this is equivalent to converting a d-bounded degree graph G to a d-regular graph G^{reg} by adding $d - deg(v)$ self-loops to each vertex v with degree $deg(v)$. Let L^{reg} be the normalized Laplacian of G^{reg}. Then the random walk on graph G is exactly same as a lazy random walk on graph G^{reg} and the definition of conductance is consistent.*

Based on the conductance, clusterability of graphs is defined as follows.

Definition 2 (**(k, φ, ϵ)-clustering**)**.** Let $G = (V, E)$ be a d-regular graph. A (k, φ, ϵ)-clustering of G is a partition of vertices V into disjoint subsets $C_1 \cup \ldots \cup C_k$ such that for all $i \in [k]$, $\phi^G(C_i) \geq \varphi$, $\phi^G_V(C_i) \leq \epsilon$ and for all $i, j \in [k]$ one has $\frac{|C_i|}{|C_j|} \in O(1)$. G is called (k, φ, ϵ)-clusterable if there exists a (k, φ, ϵ)-clustering for G.

We also need for formally define spectral embedding.

Definition 3 (Spectral embedding)**.** For a d-regular graph $G = (V, E)$ and integer $2 \leq k \leq n$ we define the spectral embedding of G as follows. Let $U \in \mathbb{R}^{k\times n}$ denote the matrix of the bottom k eigenvectors of the normalized Laplacian of G (this choice is not unique; fix any such matrix U). Then for every $x \in V$ the spectral embedding $f_x \in \mathbb{R}^k$ of x is the x-th column of the matrix U, which we write as $U = (f_y)_{y\in V}$.

Remark 2. *We note that the spectral embedding $f_x, x \in V$ is not uniquely defined. However, in this paper we are only interested in obtaining dot product access to this embedding, i.e. in fast algorithms for computing $\langle f_x, f_y\rangle$ for $x, y \in V$. Such dot products are in fact uniquely defined for any G that is (k, φ, ϵ)-clusterable with ϵ/φ^2 smaller than an absolute constant – see Remark 4 below.*

Our first algorithmic result is a sublinear time spectral dot product oracle:

Theorem 2. *[Spectral Dot Product Oracle] Let $\epsilon, \varphi \in (0, 1)$ with $\epsilon \leq \frac{\varphi^2}{10^5}$. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $1 > \xi > \frac{1}{n^5}$. Then* InitializeOracle$(G, 1/2, \xi)$ *(Algorithm 4) computes in time $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^3 \cdot \frac{1}{\varphi^2}/\xi^{18})$ a sublinear space data structure $\mathcal{D}$ of size $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^3/\xi^{12})$ such that with probability at least $1 - n^{-100}$ the following property is satisfied:*

For every pair of vertices $x, y \in V$, $\textsc{SpectralDotProduct}(G, x, y, 1/2, \xi, \mathcal{D})$ *(Algorithm 5) computes an output value $\langle f_x, f_y \rangle_{apx}$ such that with probability at least $1 - n^{-100}$*

$$\left| \langle f_x, f_y \rangle_{apx} - \langle f_x, f_y \rangle \right| \le \frac{\xi}{n}.$$

The running time of $\textsc{SpectralDotProduct}(G, x, y, 1/2, \xi, \mathcal{D})$ *is* $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$.

Furthermore, for any $0 \le \delta \le 1/2$, one can obtain the following trade-offs between preprocessing time and query time: Algorithm $\textsc{SpectralDotProduct}(G, x, y, \delta, \xi, \mathcal{D})$ *requires* $O(k^{O(1)} \cdot n^{\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$ *per query when the prepressing time of Algorithm* $\textsc{InitializeOracle}(G, \delta, \xi)$ *is increased to* $O(k^{O(1)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^3 \cdot \frac{1}{\varphi^2}/\xi^{18})$.

Our results: a spectral clustering oracle. Our goal is to compute a data structure that provides sublinear time access to a (k, φ, ϵ)-clustering of G. Such a data structure is called a (k, φ, ϵ)-clustering oracle. We now formally define a spectral clustering oracle in the Local Computation (LCA) model:

Definition 4 (Spectral clustering oracle)**.** A randomized algorithm $\mathcal{O}$ is a (k, φ, ϵ)-clustering oracle if, when given query access to a d-regular graph $G = (V, E)$ that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$, the algorithm $\mathcal{O}$ provides consistent query access to a partition $\widehat{P} = (\widehat{C}_1, \ldots, \widehat{C}_k)$ of V. The partition $\widehat{P}$ is determined solely by G and the algorithm's random seed. Moreover, with probability at least $9/10$ over the random bits of $\mathcal{O}$ the partition $\widehat{P}$ has the following property: for some permutation π on k elements one has for every $i \in [k]$:

$$|C_i \triangle \widehat{C}_{\pi(i)}| \le O\left(\frac{\epsilon \cdot \log(k)}{\varphi^3}\right) |C_i|.$$

Remark 3. *Note that it is crucial that $\mathcal{O}$ provides consistent answers, i.e. classifies a given $x \in V$ in the same way every time it is queried (for a fixing of its random seed).*

We are interested in clustering oracles that perform few probes per query. Our main contribution is:

Theorem 3. *For every integer $k \ge 2$, every $\varphi \in (0, 1)$, every $\epsilon \ll \frac{\varphi^3}{\log k}$, every $\delta \in (0, 1/2]$ there exists a (k, φ, ϵ)-clustering oracle that:*

- *has* $\widetilde{O}_\varphi\left(2^{O\left(\frac{\varphi^2}{\epsilon} k^4 \log^2(k)\right)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)}\right)$ *preprocessing time,*
- *has* $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{\delta+O(\epsilon/\varphi^2)}\right)$ *query time,*
- *uses* $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)}\right)$ *space,*
- *uses* $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{O(\epsilon/\varphi^2)}\right)$ *random bits,*

where O_φ suppresses dependence on φ and $\widetilde{O}$ hides all polylog(n) *factors.*

To the best of our knowledge, our algorithm is the first sublinear spectral clustering algorithm in literature. We hope that our main technique for providing sublinear time access to the spectral embedding will have further applications in sublinear time spectral graph theory. Our simple algorithm for recovering clusters using hyperplane partitioning in a carefully defined sequence of subspaces may also be of independent interest in spectral partitioning problems. We provide a detailed overview of the analysis and the main ideas are involved in Section 3.

Other related work. Besides the work on property testing and the work on clustering with labelled, data another closely related area is local clustering. In local clustering one is interested of finding the entire cluster around a node v in time proportional to the size of the cluster. Several algorithms are known for this problem [ACL08, AGPT16, OA14, ST14, ALM13] but unfortunately they cannot be applied to solve our problem because when the clusters have linear size they take linear time (in addition, the output clusters may overlap). In this paper instead we focus on solving the problem using strictly sublinear time.

2 Preliminaries

In this paper we mostly use the matrix notation to represent graphs. For a vertex $x \in V$, we say that $\mathbb{1}_x \in \mathbb{R}^n$ is the indicator of x, that is, the vector which is 1 at index x and 0 elsewhere. For a (multi) set $I_S = \{x_1, \ldots, x_s\}$ of vertices from V we abuse notation and also denote by S the $n \times s$ matrix whose i^{th} column is $\mathbb{1}_{x_i}$. For $i \in \mathbb{N}$ we use $[i]$ to denote the set $\{1, 2, \ldots, i\}$.

For a symmetric matrix A, we write $\nu_i(A)$ (resp. $\nu_{\max}(A), \nu_{\min}(A)$) to denote the i^{th} largest (resp. maximum, minimum) eigenvalue of A.

Let $m \leq n$ be integers. For any matrix $A \in \mathbb{R}^{n \times m}$ with singular value decomposition (SVD) $A = Y\Gamma Z^T$ we assume $Y \in \mathbb{R}^{n \times n}$, $\Gamma \in \mathbb{R}^{n \times n}$ is a diagonal matrix of singular values and $Z \in \mathbb{R}^{m \times n}$ (this is a slightly non-standard definition of the SVD, but having Γ be a square matrix will be convenient). Y has orthonormal columns, the first m columns of Z are orthonormal, and the rest of the columns of Z are zero. For any integer $q \in [m]$ we denote $Y_{[q]} \in \mathbb{R}^{n \times q}$ as the first q columns of Y and $Y_{-[q]}$ to denote the matrix of the remaining columns of Y. We also denote by $Z_{[q]} \in \mathbb{R}^{m \times q}$ as the first q columns of Z and $Z_{-[q]}$ to denote the matrix of the remaining $n - q$ columns of Z. Finally we denote by $\Gamma_{[q]} \in \mathbb{R}^{q \times q}$ the submatrix of Γ corresponding to the first q rows and columns of Γ and we use $\Gamma_{-[q]}$ to denote the submatrix corresponding to the last $n - q$ rows and $n - q$ columns of Γ. So for any $q \in [m]$ the span of $Y_{-[q]}$ is the orthogonal complement of the span of $Y_{[q]}$ in $\mathbb{R}^n$, also the span of the columns of $Z_{-[q]}$ is the orthogonal complement of the span of $Z_{[q]}$ in $\mathbb{R}^m$. Thus we can write $A = Y_{[q]}\Gamma_{[q]}Z_{[q]}^T + Y_{-[q]}\Gamma_{-[q]}Z_{-[q]}^T$.

We also denote with A_G the adjacency matrix of G and with L the *normalized Laplacian* of G where $L = I - \frac{A_G}{d}$. For L we denote its eigenvalues with $0 \leq \lambda_1 \leq \ldots \leq \lambda_n \leq 2$ and we write Λ to refer to the diagonal matrix of these eigenvalues in ascending order. We also denote with $(u_1, \ldots, u_n)$ an orthonormal basis of eigenvectors of L and with $U \in \mathbb{R}^{n \times n}$ the matrix whose columns are the orthonormal eigenvectors of L arranged in increasing order of eigenvalues. Therefore the eigendecomposition of L is $L = U\Lambda U^T$. We write $U_{[k]} \in \mathbb{R}^{n \times k}$ for the matrix whose columns are the first k columns of U and also define $F = U_{[k]}^T$. For every vertex x we denote the spectral embedding of vertex x on the bottom k eigenvectors of L with $f_x \in \mathbb{R}^k$, i.e. $f_x = F\mathbb{1}_x$. For pairs of vertices $x, y \in V$ we use the notation

$$\langle f_x, f_y \rangle := f_x^T f_y$$

to denote the dot product in the embedded domain.

Remark 4. *We note that if G is a (k, φ, ϵ)-clusterable graph with ϵ/φ^2 smaller than a constant, the space spanned by the bottom k eigenvectors of the normalized Laplacian of G is uniquely defined, i.e. the choice of $U_{[k]}$ is unique up to multiplication by an orthonormal matrix $R \in \mathbb{R}^{k \times k}$ on the right. Indeed, by Lemma 3 below one has $\lambda_k \leq 2\epsilon$ and by Lemma 1 below one has $\lambda_{k+1} \geq \varphi^2/2$. Thus, since we assume that ϵ/φ^2 is smaller than an absolute constant, we have $2\epsilon < \varphi^2/2$, and therefore the subspace spanned by the bottom k eigenvectors of the Laplacian, i.e. the space of $U_{[k]}$, is uniquely defined, as required. We note that while the choice of f_x for $x \in V$ is not unique, but the dot product between the spectral embedding of $x \in V$ and $y \in V$ is well defined, since for every orthonormal $R \in \mathbb{R}^{k \times k}$ one has $\langle Rf_x, Rf_y \rangle = (Rf_x)^T(Rf_y) = f_x^T(R^TR)f_y = f_x^Tf_y$.*

In this paper we also consider the transition matrix of the *random walk associated with G* $M = \frac{1}{2} \cdot \left(I + \frac{A}{d}\right)$. From any vertex v, this random walk takes every edge incident to v with probability $\frac{1}{2d}$, and stays on v with the remaining probability which is at least $\frac{1}{2}$. Note that this random walk is exactly same as a lazy random walk on G and that $M = I - \frac{L}{2}$. Observe that $\forall i$ u_i is also an eigenvector of M, with eigenvalue $1 - \frac{\lambda_i}{2}$. We denote with Σ the diagonal matrix of the eigenvalues of M in descending order. Therefore the eigendecomposition of M is $M = U\Sigma U^T$. We write $\Sigma_{[k]} \in \mathbb{R}^{k \times k}$ for the matrix whose columns are the first k rows and columns of Σ. Furthermore, for any t, M^t is a transition matrix of random walks of length t. For any vertex x, we denote the probability distribution of a t-step random walk starting from x by $m_x = M^t\mathbb{1}_x$. For a (multi) set $I_S = \{x_1, \ldots, x_s\}$ of vertices from V, let matrix $M^tS \in \mathbb{R}^{n \times s}$ is a matrix whose columns are probability distributions of t-step random walks starting from vertices in I_S. More formally the ith column of M^tS is m_{x_i}. For any vertex $x \in V$ let $\mathcal{N}(x) : \{y \in V : \{x, y\} \in E\}$ denote the set of vertices that are adjacent to the vertex x.

Definition 5 (**Cluster Centers**). Let $G = (V, E)$ be a d-regular graph. Let $C_1, \ldots, C_k$ be a (k, φ, ϵ)-clustering of G. We define the *spectral center* of cluster C_i as

$$\mu_i := \frac{1}{|C_i|} \sum_{x \in C_i} f_x.$$

For vertex $x \in V$, we define μ_x as the cluster center of the cluster which x belongs to.

In our analysis we use the following standard results on eigenvalues and matrix norms. Recall that for any $m \times n$ matrix A, the multi-sets of nonzero eigenvalues of $AA^\top$ and $A^\top A$ are equal.

Lemma 1 ([CKK+18]). *Let G be any graph which is composed of k components $C_1, \dots C_k$ such that $\phi^G(C_i) \geq \varphi$ for any $i \in [k]$. Let L be the normalized Laplacian matrix of G, and λ_{k+1} be the $(k+1)$st smallest eigenvalue of L. Then $\lambda_{k+1} \geq \frac{\varphi^2}{2}$.*

For a d-regular graph G, let $\rho_G(k)$ denote the minimum value of the maximum conductance over any possible k disjoint nonempty subsets. That is

$$\rho_G(k) \leq \min_{\text{disjoint } S_1, \dots, S_k} \max_i \phi_G(S_i)$$

Lemma 2 ([LGT14]). *For any d-regular graph G and any $k \geq 2$, it holds that*

$$\lambda_k \leq 2\rho_G(k).$$

Lemma 3. *Let $G = (V, E)$ be a d regular graph that admits a (k, φ, ϵ)-clustering $C_1, \dots, C_k$. Let L be the normalized Laplacian matrix of G. Let $\lambda_1 \leq \dots \leq \lambda_n$ be eigenvalues of L, then we have $\lambda_{k+1} \geq \frac{\varphi^2}{2}$ and $\lambda_k \leq 2\epsilon$.*

Proof. Note that G is composed of k components $C_1, \dots C_k$ such that for all $1 \leq i \leq k$ we have $\phi^G(C_i) \geq \varphi$. Hence, by Lemma 1 we get $\lambda_{k+1} \geq \frac{\varphi^2}{2}$. Moreover for all $1 \leq i \leq k$, we have $\phi_V^G(C_i) \leq \epsilon$. Thus by Lemma 2 we have $\lambda_k \leq 2\epsilon$. □

Since we assume that the maximum ratio of cluster sizes is bounded by a constant, we have

Proposition 1. *Let $G = (V, E)$ be a d regular graph that admits a (k, φ, ϵ)-clustering $C_1, \dots, C_k$. Then we have $\min_{i \in \{1, \dots, k\}} |C_i| = \Omega\left(\frac{n}{k}\right)$ and $\max_{i \in \{1, \dots, k\}} |C_i| = O\left(\frac{n}{k}\right)$.*

A symmetric $n \times n$ matrix is positive semi-definite, if and only if all its eigenvalues are non-negative. The spectral norm of matrix $A \in \mathbb{R}^{n \times n}$ is defined as $\max_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$ that equals the square root of the largest eigenvalue of the matrix $A^T A$. The Frobenius norm of a matrix A is defined as $\sqrt{\sum_{i,j} (A_{i,j})^2}$. For matrices $A, \widetilde{A} \in \mathbb{R}^{n \times n}$, we write $A \preccurlyeq \widetilde{A}$, if $\forall x \in \mathbb{R}^n$ we have $x^T A x \leq x^T \widetilde{A} x$.

3 Technical overview

In this section we give an overview of the analysis and the main technical contributions of the paper. Recall that we denote the matrix of bottom k eigenvectors of the normalized Laplacian of G by $U_{[k]}$. The spectral embedding of a vertex $x \in V$, denoted by $f_x \in \mathbb{R}^k$, is simply the x-th column of $U_{[k]}^T$. The main intuition behind spectral clustering is that the points $f_x \in \mathbb{R}^k$ are well-concentrated around cluster means $\mu_i \in \mathbb{R}^k$, defined for every $i = 1, \dots, k$ by

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} f_x. \tag{1}$$

See Fig. 1 for an illustration.

The contributions of our paper are twofold. Our first contribution is a primitive that provides dot product access to the spectral embedding of a graph in sublinear time: we show in Theorem 2 how, given any pair of vertices $x, y \in V$ one can compute

$$\langle f_x, f_y \rangle_{apx} \approx \langle f_x, f_y \rangle, \tag{2}$$

in time $\approx n^{1/2+O(\epsilon)}$ per evaluation (see Algorithm 5 in Section 5 for the formal definition of $\langle \cdot, \cdot \rangle_{apx}$ and its analysis).

Our second contribution is to show how dot product access as in (2) above allows one to solve the cluster recovery problem. Both of these contributions are based on a new property of the spectral embedding that we establish. This property allows us to quantify the intuitive statement that vertices in the embedding concentrate around cluster means defined in (2) above in a very strong formal sense.

In the rest of this section we first present our sublinear time dot product oracle (in Section 3.1) and then outline how access to such an oracle can be used to design a simple spectral clustering algorithm (in Section 3.2). We assume that the inner conductance of the clusters φ is constant for the purposes of this overview to simplify notation.

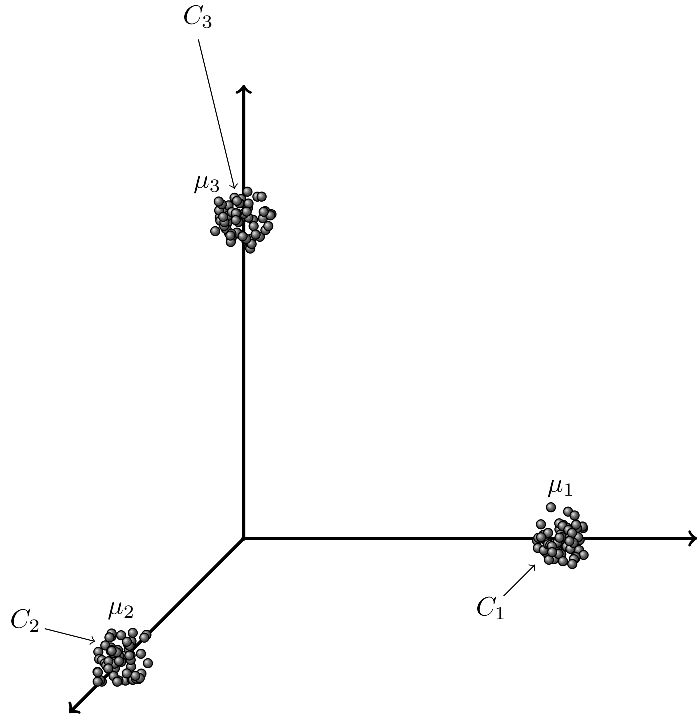

Figure 1: Example of a spectral embedding where points are concentrated around means.

3.1 Sublinear time dot product access to the spectral embedding

We start with a description of the main underlying ideas underlying the proof of Theorem 2. Our starting point from earlier work is the observation that collision statistics of random walks can be used to exhibit the structure of a (k, φ, ϵ)-clusterable graph. In particular, in (k, φ, ϵ)-clusterable graphs, there is a gap between λ_k and λ_{k+1}, and the behavior of random walks is essentially determined by the bottom k eigenvectors of the Laplacian and the corresponding eigenvalues. This suggests that we can potentially use random walks to determine the spectral embedding. The spectral embedding is of course not necessarily unique (for example, if not all of the bottom k eigenvalues are unique). However, the dot product of the embedded vertices is still well-defined as a function of the subspace spanned by the bottom k eigenvectors of the Laplacian, as the subspace itself is uniquely defined because of the aforementioned gap between λ_k and λ_{k+1}. See Remark 4 for more details. We now give an overview of our approach.

Fix two vertices $x, y \in V$. We would like to compute

$$\langle f_x, f_y \rangle = (F\mathbb{1}_x)^T (F\mathbb{1}_y) = \mathbb{1}_x^T U_{[k]} U_{[k]}^T \mathbb{1}_y.$$

The direct approach to this would amount to computing an eigendecomposition of M to obtain $U_{[k]}$, but that would take at least $\Omega(n)$ time and is too expensive for our purposes. On the other hand, it is well-known that we are able to estimate, in about $n^{1/2}$ time, the dot product

$$(M^t \mathbb{1}_x)^T (M^t \mathbb{1}_y) = \mathbb{1}_x^T M^{2t} \mathbb{1}_y.$$

Note that $\mathbb{1}_x^T M^{2t} \mathbb{1}_y = \mathbb{1}_x^T U \Sigma^{2t} U^T \mathbb{1}_y$. Thus to get $U_{[k]} U_{[k]}^T$ from $\mathbb{1}_x^T M^{2t} \mathbb{1}_y$ we need to remove the matrix Σ^{2t} from the middle. Specifically, we can estimate the quantity above as follows. For some precision parameter $\xi \in (0, 1)$ we first run $\approx n^{1/2+O(\epsilon/\varphi^2)}/\xi^2$ random walks from x, letting $\widehat{m}_x \in \mathbb{R}^n$ denote a vector whose a'th component is the fraction of random walks from x that end up at a. Similarly, we run $\approx n^{1/2+O(\epsilon/\varphi^2)}/\xi^2$ random walks from y, letting $\widehat{m}_y \in \mathbb{R}^n$ denote a vector whose a'th component is the fraction of random walks from y that end up at a. One can show[2] that with high (constant) probability we have

$$\left| \widehat{m}_x^T \widehat{m}_y - \mathbb{1}_x^T M^{2t} \mathbb{1}_y \right| \leq \xi \cdot \frac{1}{n}. \tag{3}$$

[2]This calculation is mostly amounts to a rather standard collision counting calculation that relies on the birthday paradox if one wants to establish the claim **for most vertices** $x, y \in V$ (this was done in [CPS15] and [CKK+18] for example). Our new moment bounds for the spectral embedding (see Lemmas 4 and 5 in Section 4) allow us to establish such a claim **for all vertices** $x, y \in V$ – see Lemma 22.

While (3) is not directly useful, a primitive for constructing empirical distributions $\widehat{m}_x$ and $\widehat{m}_y$ as above is a central part of our approach. We formalize it as Algorithm 1 (RunRandomWalks) below:

Algorithm 1 RunRandomWalks(G, R, t, x)

1: Run R random walks of length t starting from x
2: Let $\widehat{m}_x(y)$ be the fraction of random walks that ends at y ▷ vector $\widehat{m}_x$ has support at most R
3: **return** $\widehat{m}_x$

Even if we cannot apply (3) directly, it lets us compute a seemingly related to quantity $\mathbb{1}_x^T M^{2t} \mathbb{1}_y$ quickly by invoking Algorithm 1 and computing one dot product. In order to get from $\mathbb{1}_x^T M^{2t} \mathbb{1}_y$ to $\mathbb{1}_x^T U_{[k]} U_{[k]}^T \mathbb{1}_y$, we need to somehow apply a linear transformation on the random walk distributions **before** computing the dot product between them, i.e. we need a different dot product operation. It is easy to see that the correct linear transformation is given by the matrix $U_{[k]} \Sigma_{[k]}^{-2t} U_{[k]}^T$, where $M^t = U \Sigma^t U^T$ is the eigendecomposition of M and $U_{[k]}$ stands for the matrix of bottom k eigenvectors of the Laplacian[3]. Specifically, we have

$$(M^t \mathbb{1}_x)^T (U_{[k]} \Sigma_{[k]}^{-2t} U_{[k]}^T)(M^t \mathbb{1}_y) = \mathbb{1}_x^T U_{[k]} U_{[k]}^T \mathbb{1}_y = \langle f_x, f_y \rangle,$$

which is exactly the quantity we are interested in. Of course, there is a major problem with this approach, since $U_{[k]} \Sigma_{[k]}^{-2t} U_{[k]}^T$ is an $n \times n$ matrix! To get around this issue, we approximate $U_{[k]} \Sigma_{[k]}^{-2t} U_{[k]}^T$ by a sparse low rank matrix, as we describe below. Specifically, we let I_S be a multiset of $s \ll n$ vertices selected uniformly at random. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$ and let $\widetilde{W} \widetilde{\Sigma}^{2t} \widetilde{W}^T$ denote the eigendecomposition of $\frac{n}{s} \cdot (M^t S)^T (M^t S)$[4]. We show that with an appropriate choice of the sampling parameter $s \ll n$ one has

$$U_{[k]} \Sigma_{[k]}^{-2t} U_{[k]}^T \approx M^t S \cdot \widetilde{\Psi} \cdot S^T M^t, \tag{4}$$

where

$$\widetilde{\Psi} = \frac{n}{s} \cdot \widetilde{W}_{[k]} \widetilde{\Sigma}_{[k]}^{-4t} \widetilde{W}_{[k]}^T \tag{5}$$

is an $s \times s$ matrix that can be computed explicitly. The corresponding primitive to compute $(M^t S)^T (M^t S)$ is presented as Algorithm 2 (EstimateCollisionProbabilities) below. It basically estimates the Gram matrix of random walk distributions out of I_S (denoted by $\mathcal{G}$) by counting collisions, and taking medians of estimates to reduce failure probability appropriately. After computing the approximate Gram matrix, we derive from it the matrix $\Psi = \frac{n}{s} \cdot \widehat{W}_{[k]} \widehat{\Sigma}_{[k]}^{-2} \widehat{W}_{[k]}^T$, where $\mathcal{G} = \widehat{W} \widehat{\Sigma} \widehat{W}^T$ is the eigendecomposition of $\mathcal{G}$ (see line (8) and line (10) of Algorithm 4; note that G is a symmetric matrix, and hence an eigendecomposition exists).

Algorithm 2 EstimateCollisionProbabilities(G, I_S, R, t)

1: **for** $i = 1$ to $O(\log n)$ **do**
2: $\quad \widehat{Q}_i :=$ EstimateTransitionMatrix(G, I_S, R, t)
3: $\quad \widehat{P}_i :=$ EstimateTransitionMatrix(G, I_S, R, t)
4: $\quad \mathcal{G}_i := \frac{1}{2}\left(\widehat{P}_i^T \widehat{Q}_i + \widehat{Q}_i^T \widehat{P}_i\right)$ ▷ $\mathcal{G}_i$ is symmetric
5: Let $\mathcal{G}$ be a matrix obtained by taking the entrywise median of $\mathcal{G}_i$'s ▷ $\mathcal{G}$ is symmetric
6: **return** $\mathcal{G}$ ▷ $\mathcal{G} \in \mathbb{R}^{s \times s}$

Algorithm 2 uses an auxiliary primitive presented as

Algorithm 3 EstimateTransitionMatrix(G, I_S, R, t)

1: **for** each sample $x \in I_S$ **do**
2: $\quad \widehat{m}_x :=$ RunRandomWalks(G, R, t, x)
3: Let $\widehat{Q}$ be the matrix whose columns are $\widehat{m}_x$ for $x \in I_S$
4: **return** $\widehat{Q}$ ▷ $\widehat{Q}$ has at most Rs non-zeros

[3]Note that this matrix is not well defined in the presence of repeated eigenvectors, but any fixed choice of this matrix suffices for our purposes. It is also interesting to note that while we use a canonical choice of the eigendecomposition of M throughout the paper, all our bounds are oblivious to the choice of this basis, and hold for the *subspace* of bottom k eigenvectors, which is well defined since there is a gap between the k-th and $(k+1)$-th eigenvalues in k-clusterable graphs.

[4]We abuse notation somewhat by writing S to denote the $n \times s$ matrix whose (a, j)-th entry equals 1 if the j-th sampled vertex equals a and 0 otherwise.

The proof of (4) relies on matrix perturbation bounds (the Davis-Kahan $\sin\theta$ theorem) as well as spectral concentration inequalities, crucially coupled with our tail bounds on the spectral embedding (see Lemma 4 and Lemma 5). In particular Lemma 4 and it's consequence - Lemma 5 can be used to bound the leverage scores of $U_{[k]}$ (i.e. $||f_x||_2^2$ for $x \in V$). This part of the analysis is presented in Section 5.2.

Lemma 4. *[Tail-bound] Let $\varphi \in (0,1)$ and $\epsilon \leq \frac{\varphi^2}{100}$, and let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \dots, C_k$. Let L be the normalized Laplacian of G. Let u be a normalized eigenvector of L with $||u||_2 = 1$ and with eigenvalue at most 2ϵ. Then for any $\beta > 1$ we have*

$$\frac{1}{n} \cdot \left| \left\{ x \in V : |u(x)| \geq \beta \cdot \sqrt{\frac{10}{\min_{i\in[k]} |C_i|}} \right\} \right| \leq \left(\frac{\beta}{2}\right)^{-\varphi^2/20\cdot\epsilon}.$$

Lemma 5. *Let $\varphi \in (0,1)$ and $\epsilon \leq \frac{\varphi^2}{100}$, and let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \dots, C_k$. Let u be a normalized eigenvector of L with $||u||_2 = 1$ and with eigenvalue at most 2ϵ. Then we have*

$$||u||_\infty \leq n^{20\cdot\epsilon/\varphi^2} \cdot \sqrt{\frac{160}{\min_{i\in k} |C_i|}}.$$

We note that the number of samples s is chosen as $s \approx k^{O(1)} n^{O(\epsilon/\varphi^2)}$ (see Algorithm 4) , where the second factor is due to our upper bound on the ℓ_∞ norm of the bottom k eigenvectors of the Laplacian of a (k, φ, ϵ)-clusterable graph proved in Section 4.

Once we establish (4) in Section 5.3 (see Lemma 19), we get for every $x, y \in V$

$$(M^t \mathbb{1}_x)^T M^t S \cdot \widetilde{\Psi} \cdot S^T M^t (M^t \mathbb{1}_y) \approx \mathbb{1}_x^T U_{[k]} U_{[k]}^T \mathbb{1}_y, \tag{6}$$

which is what we would like to compute. One issue remains at this point, which is that we cannot compute $M^t \mathbb{1}_x$ or $M^t \mathbb{1}_y$ explicitly, and neither can we store and compute our approximation $M^t S \cdot \Psi \cdot S^T M^t$, since it is a dense, albeit low rank, matrix. We resolve this problem by running an appropriate number of random walks out of the sampled nodes I_S, as well as the queried nodes $x, y \in V$. Specifically, we run $\approx n^{1/2+O(\epsilon)}$ random walks from every sampled node in I_S, defining an $n \times s$ matrix Q whose (a, b)-th entry is the fraction of walks from a that ended at b and using the matrix Q as a proxy for $M^t S$ (note that the expectation of Q is exactly $M^t S$). Such a matrix Q is computed as per line (2) and line (3) of Algorithm 2 (EstimateCollisionProbabilities). We note that Algorithm 4 (InitializeOracle) performs $O(\log n)$ independent estimates that we ultimately use to boost confidence (by the median trick). The entire preprocessing is summarized in Algorithm 4 (InitializeOracle) below:

Algorithm 4 InitializeOracle(G, δ, ξ) ▷ Need: $\epsilon/\varphi^2 \leq \frac{1}{10^5}$

$t := \frac{20\cdot\log n}{\varphi^2}$
$R_{\text{init}} := O(n^{1-\delta+3\cdot10^3\cdot\epsilon/\varphi^2} \cdot k^{33}/\xi^6)$
$s := O(n^{1500\cdot\epsilon/\varphi^2} \cdot \log n \cdot k^{16}/\xi^6)$
Let I_S be the multiset of s indices chosen independently and uniformly at random from $\{1, \dots, n\}$
for $i = 1$ to $O(\log n)$ **do**
  $\widehat{Q}_i :=$ EstimateTransitionMatrix$(G, I_S, R_{\text{init}}, t)$ ▷ $\widehat{Q}_i$ has at most $R_{\text{init}} \cdot s$ non-zeros
$\mathcal{G} :=$EstimateCollisionProbabilities$(G, I_S, R_{\text{init}}, t)$
Let $\frac{n}{s} \cdot \mathcal{G} := \widehat{W}\widehat{\Sigma}\widehat{W}^T$ be the eigendecomposition of $\frac{n}{s} \cdot \mathcal{G}$ ▷ $\mathcal{G} \in \mathbb{R}^{s\times s}$
if $\widehat{\Sigma}^{-1}$ exists **then**
  $\Psi := \frac{n}{s} \cdot \widehat{W}_{[k]} \widehat{\Sigma}_{[k]}^{-2} \widehat{W}_{[k]}^T$ ▷ $\Psi \in \mathbb{R}^{s\times s}$
  return $\mathcal{D} := \{\Psi, \widehat{Q}_1, \dots, \widehat{Q}_{O(\log n)}\}$

Equipped with the primitives presented above, we can now state our final dot product estimate:

$$\widehat{m}_x^T Q \Psi Q^T \widehat{m}_y \approx \mathbb{1}_x^T U_{[k]} U_{[k]}^T \mathbb{1}_y = \langle f_x, f_y \rangle, \tag{7}$$

where $\widehat{m}_x$ and $\widehat{m}_y$ are empirical distributions of $\approx n^{1/2+O(\epsilon/\phi^2)}$ out of x and y respectively, Q is an $n \times s$ matrix with $\approx n^{1/2+O(\epsilon/\phi^2)}$ nonzeros per column, and Ψ is a possibly dense $s \times s$ matrix, where the number of sampled vertices s is ultimately chosen to be $k^{O(1)} n^{O(\epsilon/\phi^2)}$. The analysis of the error incurred in replacing (4) with (7) is presented in Section 5.4. It relies on a birthday paradox style variance computation similar to previous sublinear time algorithms for testing graph cluster structure. The actual query procedure that implements (7) is given by Algorithm 5 below.

Algorithm 5 SPECTRALDOTPRODUCTORACLE$(G, x, y, \delta, \xi, \mathcal{D})$ $\qquad \triangleright$ Need: $\epsilon/\varphi^2 \leq \frac{1}{10^5}$

$\triangleright$ $\mathcal{D} := \{\Psi, \widehat{Q}_1, \ldots, \widehat{Q}_{O(\log n)}\}$

1: $R_{\text{query}} := O(n^{\delta + 500 \cdot \epsilon/\varphi^2} \cdot k^9/\xi^2)$
2: **for** $i = 1$ to $O(\log n)$ **do**
3: $\quad \widehat{m}_x^i :=$ RUNRANDOMWALKS$(G, R_{\text{query}}, t, x)$
4: $\quad \widehat{m}_y^i :=$ RUNRANDOMWALKS$(G, R_{\text{query}}, t, y)$
5: Let α_x be a vector obtained by taking the entrywise median of $(\widehat{Q}_i)^T(\widehat{m}_x^i)$ over all runs
6: Let α_y be a vector obtained by taking the entrywise median of $(\widehat{Q}_i)^T(\widehat{m}_y^i)$ over all runs
7: **return** $\langle f_x, f_y \rangle_{apx} := \alpha_x^T \Psi \alpha_y$

Trading off preprocessing time for query time. Finally, we note that one can reduce query time (i.e., runtime of SPECTRALDOTPRODUCTORACLE) at the expense of increased preprocessing time and size of data structure. Specifically, one can run $\approx n^{\delta + O(\epsilon/\phi^2)}$ random walks from nodes x, y whose dot product is being estimated by SPECTRALDOTPRODUCTORACLE at the expense of increasing the number of random walks run to generate the matrix Q in INITIALIZEORACLE to $\approx n^{1-\delta+O(\epsilon/\phi^2)}$, for any $\delta \leq 1/2$. This in particular leads to a nearly linear time spectral clustering algorithm.

3.2 Geometry of the spectral embedding

We now describe our spectral clustering algorithm. Since we only have dot product access to the spectral embedding, the algorithm must be very simple. Indeed, our algorithm amounts to performing hyperplane partitioning in a sequence of carefully crafted subspaces of the embedding space, using (a good approximation to) cluster means μ_i.

We first present a simple hyperplane partitioning, then we give an example embedding to show why it might be hard to prove that this scheme works. After that we design a modification of the hyperplane partitioning scheme that, through the course of carving, carefully projects out some directions of the embedding. This modification is an idealized version of our final algorithm for which we can prove per cluster recovery guarantees.

First we assume that the cluster means (1) are known. In that case we define, for every $i = 1, \ldots, k$, the sets

$$\widetilde{C}_i := \{x \in V : \langle f_x, \mu_i \rangle \geq 0.9||\mu_i||^2\}$$

of points that are nontrivially correlated with the i-th cluster mean μ_i. Note that $\widetilde{C}_i = C_{\mu_i, 0.9}$ in terms of Definition 8, but since μ_i's are fixed in this overview, we use the simpler notation. We next define, for every $i = 1, \ldots, k$,

$$\widehat{C}_i := \widetilde{C}_i \setminus \bigcup_{j=1}^{i-1} \widetilde{C}_j. \tag{8}$$

In other words, this is a natural 'hyperplane-carving' approach: points that belong to the first hyperplane $\widetilde{C}_1$ are taken as the first cluster, points in the second hyperplane $\widetilde{C}_2$ that were not captured by the first hyperplane are taken as the second cluster, etc. This is a natural high dimensional analog of the Cheeger cut that has been used in many results on spectral partitioning. The hope here would be to show that there exists a permutation π on $[k]$ such that

$$|\widehat{C}_i \Delta C_{\pi(i)}| \leq O(\epsilon) \cdot |C_{\pi(i)}|, \tag{9}$$

for every $i = 1, \ldots, k$, where we assume that the inner conductance ϕ of the clusters is constant. Here Δ stands for the symmetric difference operation.

One natural approach to establishing (9) would be to prove that for every $i = 1, \ldots, k$ vertices $x \in C_i$ concentrate well around cluster means μ_i (see Fig. 1). This would seem to suggest that $\widetilde{C}_i$'s are close to the C_i's, and so are the $\widehat{C}_i$'s. This property of the spectral embedding is quite natural to expect, and versions of this property have been used in the literature. For example, one can show that for every $\alpha \in \mathbb{R}^k, ||\alpha||_2 = 1$,

$$\sum_{i=1}^{k} \sum_{x \in C_i} \langle f_x - \mu_i, \alpha \rangle^2 \leq O(\epsilon). \tag{10}$$

The bound in (10) follows using rather standard techniques – see Section 4.1 for this and related claims. One can check that (10) suffices to show that $\widetilde{C}_i$'s are very close to C_i's, namely that for every $i = 1, \ldots, k$ there exists $j \in [k]$ such that

$$|\widetilde{C}_i \Delta C_j| = O(\epsilon) \cdot |C_j|. \tag{11}$$

The formal proof is given in Section 6.2. The result in (11) is encouraging and suggests that the clusters $\widehat{C}_i$ defined by the simple hyperplane partitioning process approximate the C_i's, but this is not the case! The problem lies in the fact that while $\widetilde{C}_i$'s approximate the C_i's well as per (11), the bound in (11) does not preclude nontrivial overlaps in the $\widetilde{C}_i$'s – we give an example in below.

3.2.1 Hard instance for natural hyperplane partitioning

We now give an example configuration of vertices in Euclidean space such that **(a)** the configuration does not contradict (10) and **(b)** the natural hyperplane partitioning algorithm (8) fails for this configuration. This shows why we develop a different algorithm that can deal with configurations like the one presented in this subsection.

Consider the following configuration of C_i's and μ_i's. Suppose that all cluster sizes are equal $\frac{n}{k}$, and let $k = \frac{1}{\epsilon}$. Let μ_i's form an orthogonal system and for each $i \in [k]$ let $||\mu_i||_2 = \sqrt{\frac{k}{n}}$. For all $i < k = 1/\epsilon$ for all $x \in C_i$ we set $f_x = \mu_i$, that is points from all clusters except for $1/\epsilon$'th one are tightly concentrated around cluster means – see Fig. 2 for an illustration with $k = 3$. Then for cluster $C_{1/\epsilon}$ we distribute points as follows. For every $i = 1, \ldots, 1/\epsilon - 1$ we move $\epsilon/2$ fraction of its points to $\mu_{1/\epsilon} + \mu_i$, and another $\epsilon/2$ fraction of the points to $\mu_{1/\epsilon} - \mu_i$. The remaining ϵ fraction of $C_{1/\epsilon}$ stays at $\mu_{1/\epsilon}$. Now observe that all cluster means are where they should be, since we applied symmetric perturbations. Secondly notice that (10) is satisfied for every direction α. Intuitively it is the case because we moved $1/\epsilon - 1$ disjoint subsets of $C_{1/\epsilon}$ of size $\epsilon \frac{n}{k}$ in $1/\epsilon - 1$ **orthogonal** directions. Lastly observe what happens to $\widetilde{C}_i$'s. For all $i = 1, \ldots, 1/\epsilon - 1$ set $\widetilde{C}_i$ contains C_i and $\epsilon/2$ fraction of $C_{1/\epsilon}$ that was moved in direction μ_i. One can verify that this is perfectly consistent with (10), and in particular with (11). The problem is that many clusters have large overlap with one particular cluster, namely $C_{1/\epsilon}$. Indeed notice that the ball carving process returns $\widehat{C}_{1/\epsilon}$ such that $|\widehat{C}_{1/\epsilon} \cap C_{1/\epsilon}| = (\frac{1+\epsilon}{2}) \frac{n}{k}$. That means that constant (almost 1/2) fraction of cluster $C_{1/\epsilon}$ is not recovered!

3.2.2 Our hyperplane partitioning scheme

The example in Section 3.2.1 suggests that we need to develop a diffferent algorithm. Our main contribution here is an algorithm that more carefully deals with the overlaps of $\widetilde{C}_i$'s. The high level idea for the algorithm is to recover clusters in stages and after every stage project out the directions corresponding to recovered clusters.

First we observe the following property of (k, φ, ϵ)-clusterable graphs (see Lemma 16). Any collection of pairwise disjoint sets with small outer-conductance matches the original clusters well. More precisely for every collection $\{\widehat{C}_1, \ldots, \widehat{C}_k\}$ of pairwise disjoint sets satisfying for every $i \in [k]$ $\phi(\widehat{C}_i) \leq O(\epsilon \log(k))$ there exists a permutation π on $[k]$ such that

$$|\widehat{C}_i \Delta C_{\pi(i)}| \leq O(\epsilon \log(k)) \cdot |C_{\pi(i)}|, \tag{12}$$

In the algorithm we will test many candidate clusters and the property above allows us to test if a particular candidate $\widehat{C}$ is good by only computing its outer-conductance.

Now we describe our algorithm more formally. The algorithm proceeds in $O(\log(k))$ stages. In the first stage it considers k candidate clusters $\widehat{C}_i$, where $x \in \widehat{C}_i$ if it has big correlation with μ_i but small correlation with all other μ_j's. More formally

$$\widehat{C}_i := \widetilde{C}_i \setminus \bigcup_{j \neq i} \widetilde{C}_j, \tag{13}$$

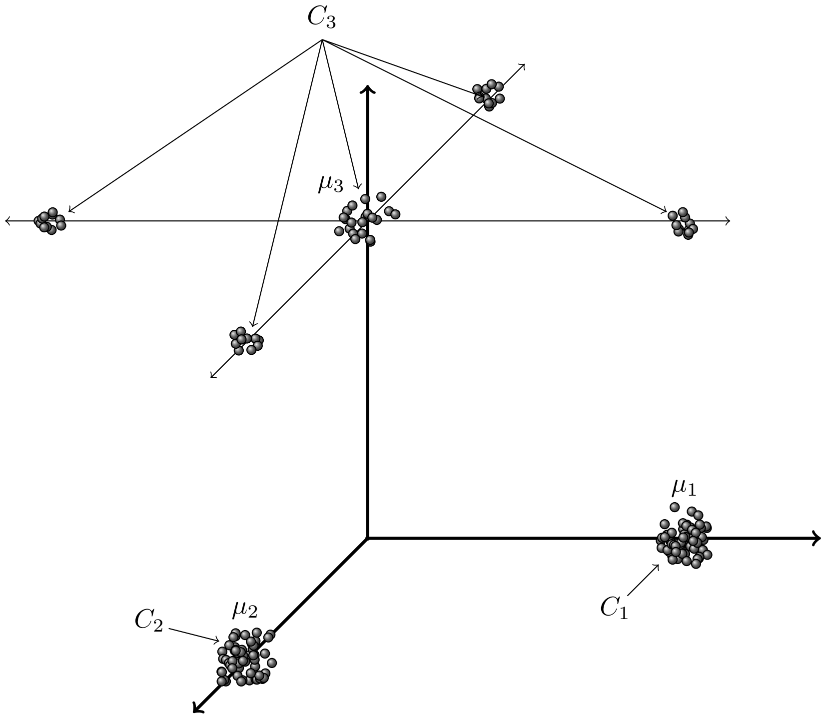

Figure 2: Example of a spectral embedding that is consistent with (10) and (11) but for which the natural hyperplane partitioning would not work.

which is equivalent to:

$$\langle f_x, \mu_i \rangle \geq 0.9||\mu_i||^2 \text{ and for all } j \neq i \; \langle f_x, \mu_j \rangle < 0.9||\mu_j||^2.$$

Note that by definition all these clusters are disjoint. At this point we return all candidate clusters $\widehat{C}_i$ for which $\phi(\widehat{C}_i) \leq O(\epsilon)$, remove the corresponding vertices from the graph, remove the corresponding μ's from the set $\{\mu_1, \ldots, \mu_k\}$ of centers and proceed to the next stage.

In the next stage we restrict our attention to a lower dimensional subspace Π of $\mathbb{R}^k$. Intuitively we want to project out all the directions corresponding to the removed cluster centers. Formally we define Π to be the subspace orthogonal to all μ's removed up to this point (we overload notation by also using Π for the orthogonal projection onto this subspace). We will see that μ's are close to being orthogonal (see Lemma 7). This fact means that $\Pi \approx \text{span}(\{\mu_1, \ldots, \mu_b\})$, where $\{\mu_1, \ldots, \mu_b\}$ is the set of μ's that were not removed in the first step. Now the algorithm considers b candidate clusters where the condition for x being in a cluster i changes to:

$$\langle f_x, \Pi\mu_i \rangle \geq 0.9||\Pi\mu_i||^2 \text{ and for all } j \in [b], j \neq i \; \langle f_x, \Pi\mu_j \rangle < 0.9||\Pi\mu_j||^2.$$

Now we return all candidate clusters that satisfy $\phi(\widehat{C}_i) \leq O(\epsilon)$ but this time the constant hidden in the O notation is bigger than in the first stage. In general at any stage t we change the test to $O(\epsilon \cdot t)$. At the end of the stage we proceed in a similar fashion by returning the clusters, removing the corresponding vertices and μ's and considering a lower dimensional subspace of Π in the next stage.

The algorithm continues in such a fashion for $O(\log(k))$ stages. Thus for all returned clusters $\widehat{C}_i$ it is true that there exists j such that[5]:

$$|\widehat{C}_i \triangle C_j| \leq O\left(\epsilon \log(k)\right) \cdot |C_j|.$$

Let's analyze how this algorithm works for the configuration presented in Section 3.2.1. In the first stage we have that, for all $i \neq \frac{1}{\epsilon}$, $\widehat{C}_i = C_i$ and moreover $|\widehat{C}_{1/\epsilon} \cap C_{1/\epsilon}| = (\frac{1+\epsilon}{2})\frac{n}{k}$. So all candidate cluster

[5]Note that this algorithm may not return a partition of the graph but only a collection of disjoint clusters. Later, in Section 6.6 in Proposition 3, we present a simple reduction that shows that an algorithm that guarantees (12) is enough to construct a clustering oracle that, as required by Definition 4, returns a partition. The high level idea is to assign the remaining vertices to clusters randomly.

$\widehat{C}_i$ for $i \neq 1/\epsilon$ are returned but crucially this time (in contrast with the natural hyperplane partitioning) cluster $C_{1/\epsilon}$ is left untouched. Then directions $\{\mu_1, \ldots, \mu_{1/\epsilon-1}\}$ are projected out. In the second stage the algorithm considers only vertices from $C_{1/\epsilon}$ projected onto one dimensional subspace span$(\mu_{1/\epsilon})$ and recovers this cluster up to $O(\epsilon)$ error.

Because of the robustness property (12), to show that this algorithm works we only need to argue that at the end of $O(\log(k))$ stages k sets are returned. We do that by showing that in every stage at least half of the remaining clusters is recovered. It is done in Lemma 37 and crucially relies on the following fact. When the algorithm considers a subspace Π then the number of points in the union of sets:

$$\{x \in V : \langle f_x, \Pi\mu_i \rangle \geq 0.9||\Pi\mu_i||^2\} \cap \{x \in V : \langle f_x, \Pi\mu_j \rangle \geq 0.9||\Pi\mu_j||^2\},$$

for all $i, j \in [b], i \neq j$ is bounded by $O(\epsilon \cdot b \cdot \frac{n}{k})$ (see Lemma 36 and Remark 7). To prove that we observe that every point x in this intersections has big projection onto some two μ_i, μ_j from $\{\mu_1, \ldots, \mu_b\}$. Then using the fact that μ's are close to being orthogonal we deduce that $\Pi \approx$ span$(\{\mu_1, \ldots, \mu_b\})$ this in particular means that $\Pi\mu_i \approx \mu_i$, $\Pi\mu_j \approx \mu_j$. Because of that f_x is abnormally far (further by a factor of $1/\epsilon$ with respect to the average) from it's center μ_x. Now applying (10) for an orthonormal basis of Π and summing the inequalities we get that that the number of points in the intersections is bounded by $O(\epsilon \cdot b \cdot \frac{n}{k})$. Having this bound we can argue that at least half of the remaining clusters is recovered as on average only $O(\epsilon \cdot \frac{n}{k})$ points from each cluster belong to the intersections. The formal argument is given in Section 6.3.

The use of subspaces is crucial for our approach. If we relied solely on the bounds on norms (i.e. bounds on $||f_x||$) we could only claim a recovery guarantee of $O(\epsilon k)$ per cluster. One of the reasons is that there can be $\Theta(\epsilon n)$ vertices of abnormally big norm and all of them can belong to one cluster (as it happens in the example from Section 3.2.1). The use of carefully crafted sequence of subspaces solves this issue as it allows to derive better bounds for the number of abnormal vertices in each stage. It is possible as we can show that the "variance of the distribution" of f_x's cannot concentrate on subspaces. This leads to an $O(\epsilon \log(k))$ error guarantee per cluster.

What remains is to remove the assumption that the cluster means μ_i are known to the algorithm. We show, using our tail bounds from Lemma 4, that a random sample of $O(1/\epsilon \cdot k^3 \log k)$ points in every cluster is likely to concentrate around the mean. This allows us to take a $O(1/\epsilon \cdot k^4 \log k)$ size sample of points, guess in exponential (in $1/\epsilon \cdot k^4 \log^2 k$) time which points belong to which cluster, and ultimately find surrogates $\widehat{\mu}_i$ that are sufficiently close to the actual μ_i's for the analysis to go through. This part of the analysis is presented in Section 6.4.1. We also need a mechanism for testing if a set of approximate $\widehat{\mu}$'s induces (via our partitioning algorithm) a good clustering. We accomplish this goal by designing a simple sampling based tester that determines whether or not the clusters induced by a particular collection of candidate cluster means have the right size and outer conductance properties. See Section 6.5 for this part of the analysis.

To design our spectral clustering algorithm we need to perform tests like $\langle f_x, \Pi\mu \rangle \overset{?}{\geq} 0.9||\Pi\mu||_2^2$ for a given vertex x, a candidate cluster mean μ, and the projection matrix Π. Hence, we need tools to approximate $\langle f_x, \Pi\mu \rangle$ and $||\Pi\mu||_2^2$. As explained above, instead of exact cluster means i.e. μ we will perform the test for approximate cluster means i.e, $\widehat{\mu} = \frac{1}{|S|}\sum_{y\in S} f_y$, where S is a small subset S of sampled nodes. First observe that for any vertex x one can estimate $\langle f_x, \widehat{\mu} \rangle_{apx}$ as follows:

$$\langle f_x, \widehat{\mu} \rangle_{apx} = \frac{1}{|S|}\sum_{y\in S} \langle f_x, f_y \rangle_{apx}$$

where $\langle f_x, f_y \rangle_{apx}$ can be computed using (SpectralDotProductOracle) Algorithm 5. Next we will explain how to compute $\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx}$ for $x, y \in V$. Recall that $\widehat{\Pi}$ is the subspace orthogonal to all $\widehat{\mu}$'s removed so far. Let $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_r\}$ denote the set of removed cluster means, and let $X \in \mathbb{R}^{k\times r}$ denote a matrix whose columns are $\widehat{\mu}_i$'s. Therefore the projection matrix onto the span of $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_r\}$ is given by $X(X^TX)^{-1}X$. Hence, we have $\widehat{\Pi} = I - X(X^TX)^{-1}X$ and we can compute $\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx}$ as follows:

$$\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} = \langle f_x, f_y \rangle_{apx} - (f_x^T X)(X^TX)^{-1}(Xf_y).$$

Note that the i-th column of matrix X is $\widehat{\mu}_i$, thus $f_x^T X \in \mathbb{R}^r$ is a vector whose i-th entry can be computed by $\langle f_x, \widehat{\mu}_i \rangle_{apx}$. Moreover notice that $X^TX \in \mathbb{R}^{r\times r}$ is matrix such that its (i, j)-th entry can be computed by $\langle \widehat{\mu}_i, \widehat{\mu}_j \rangle_{apx}$. Therefore $(f_x^T X)$, $(X f_y)$ and $(X^T X)^{-1}$ all can be computed explicitly which let us compute $\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx}$. Given the primitive to compute $\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx}$ we are able to estimate $\langle f_x, \Pi(\mu) \rangle$ and $||\Pi(\mu)||_2^2$ as follows:

$$\left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx} := \frac{1}{|B|} \cdot \sum_{y \in B} \left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx},$$

$$\left\| \widehat{\Pi}\widehat{\mu} \right\|_{apx}^2 := \frac{1}{|B|} \cdot \sum_{x \in B} \left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx}.$$

This part of the analysis is presented in Section 5.6.

4 Properties of the spectral embedding of (k, φ, ϵ)-clusterable graphs

In this section we study the spectral embedding of (k, φ, ϵ)-clusterable graphs. Recall that the spectral embedding maps every vertex $x \in V$ to a k-dimensional vector f_x. We are interested in understanding the geometric properties of this embedding. We start by recalling some standard properties of the embedding: We show that the cluster means

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} f_x$$

are almost orthogonal and of length roughly $1/\sqrt{|C_i|}$ (Lemma 7 below). Then we give a bound on the directional variance, by which we mean the sum of squared distances of points f_x to their corresponding cluster centers when projected on direction α. We show in Lemma 6 below that the directional variance is bounded by $O(\epsilon/\varphi^2)$ for every direction $\alpha \in \mathbb{R}^k$, $\|\alpha\| = 1$. This in particular implies (see Lemma 9 below) that 'rounding' the spectral embedding by mapping each vertex to its corresponding cluster center results in a matrix U that spectrally approximates the matrix of bottom k eigenvectors of the Laplacian. These bounds are rather standard, and their proofs are provided for completeness. The main shortcoming of the standard bounds is that they can only allow us to apply averaging arguments, and are thus unable to rule out that some of the embedded points are quite far away from their corresponding cluster center. For example, they do not rule out the possibility of an $\Omega(1/k)$ fraction of the points being $\approx \sqrt{k}$ further away from their corresponding centers. Since we would like to recover every cluster to up an $O(\epsilon)$ error, such bounds are not sufficient on their own.

For this reason we consider the distribution of the projection of the embedded points on the direction of any of the first k eigenvectors and we give stronger tail bounds for these distributions (in Lemma 4) than what follows from variance calculations only. Basically, we give a strong bound on the $O(\varphi^2/\epsilon)$-th moment of the spectral embedding as opposed to just on the second moment, as above. These higher moment bounds are then crucially used to achieve sublinear time access to dot products in the embedded space in Section 5 (we need them to establish spectral concentration of a small number of random samples in Section 5.2) as well as to argue that a small sample of vertices contains a good approximation to the true cluster means $\mu_i, i = 1, \ldots, k$ in its span in Section 6.4.1.

4.1 Standard bounds on cluster means and directional variance

The lemma below bounds the variance of the spectral embedding in any direction.

Lemma 6. *(Variance bounds) Let $k \geq 2$ be an integer, $\varphi \in (0, 1)$, and $\epsilon \in (0, 1)$. Let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Then for all $\alpha \in \mathbb{R}^k$, with $\|\alpha\| = 1$ we have*

$$\sum_{i=1}^{k} \sum_{x \in C_i} \langle f_x - \mu_i, \alpha \rangle^2 \leq \frac{4\epsilon}{\varphi^2}.$$

Proof. For each $i \in [k]$, and any vertex $x \in C_i$, let $d_i(x)$ denote the degree of vertex x in the subgraph C_i. Let H_i be a graph obtained by adding $d - d_i(x)$ self-loops to each vertex $x \in C_i$. Let L denote the normalized Laplacian of graph G. For each $i \in [k]$ and let L_i denote the normalized Laplacian of H_i, and let $\lambda_2(H_i)$ be the second smallest eigenvalue of L_i.

Let $z = U_{[k]}\alpha$. Note that $||z||_2 = 1$. By Lemma 3 we have $\lambda_1 \le \ldots \le \lambda_k \le 2\epsilon$, where λ_i is the i^{th} smallest eigenvalue of of L. Therefore we have

$$\langle z, Lz\rangle \le \lambda_k \le 2\epsilon \tag{14}$$

Fix some $i \in [k]$, let $z' \in \mathbb{R}^n$ be a vector such that $z'(x) := z(x) - \langle \mu_i, \alpha\rangle$. For any $S \subseteq V$, we define $z'_S \in \mathbb{R}^n$ to be a vector such that for all $x \in V$ $z'_S(x) = z'(x)$ if $x \in S$ and $z'_S(x) = 0$ otherwise. Note that $z(x) = \langle f_x, \alpha\rangle$, thus we have

$$\sum_{x\in V} z'_{C_i}(x) = \sum_{x\in C_i} z'(x) = \sum_{x\in C_i} z(x) - \langle \mu_i, \alpha\rangle = \sum_{x\in C_i} \langle f_x - \mu_i, \alpha\rangle = 0$$

Thus we have $z'_{C_i} \perp \mathbb{1}$, so by properties of Rayleigh quotient we get

$$\frac{\langle z'_{C_i}, L_i z'_{C_i}\rangle}{\langle z'_{C_i}, z'_{C_i}\rangle} = \frac{1}{d}\frac{\sum_{x,y\in C_i,(x,y)\in E}(z'(x) - z'(y))^2}{\sum_{x\in C_i}(z'(x))^2} = \frac{1}{d}\frac{\sum_{x,y\in C_i,(x,y)\in E}(z(x) - z(y))^2}{\sum_{x\in C_i}(z(x) - \langle\mu_i,\alpha\rangle)^2} \ge \lambda_2(H_i) \tag{15}$$

Furthermore, by Cheeger's inequality for any $i \in [k]$ we have $\lambda_2(H_i) \ge \frac{\varphi^2}{2}$. Hence, for any $i \in [k]$ we have

$$\frac{\sum_{x,y\in C_i,(x,y)\in E}(z(x) - z(y))^2}{d\sum_{x\in C_i}(z(x) - \langle\mu_i,\alpha\rangle)^2} \ge \lambda_2(H_i) \ge \frac{\varphi^2}{2}$$

Now observe the following:

$$\begin{aligned}
2\epsilon &\ge \langle z, Lz\rangle && \text{By (14)}\\
&= \frac{1}{d}\cdot \sum_{(x,y)\in E} (z(x) - z(y))^2\\
&\ge \frac{1}{d}\cdot \sum_{i=1}^{k} \sum_{x,y\in C_i,(x,y)\in E} (z(x) - z(y))^2\\
&\ge \frac{\varphi^2}{2}\cdot \sum_{i=1}^{k}\sum_{x\in C_i} (z(x) - \langle\mu_i,\alpha\rangle)^2 && \text{By (15)}
\end{aligned}$$

Recall that for all $x \in V$, $z(x) = \langle f_x, \alpha\rangle$. Therefore for for any $\alpha \in \mathbb{R}^k$ with $\|\alpha\| = 1$ we have

$$\sum_{i=1}^{k}\sum_{x\in C_i} \langle f_x - \mu_i, \alpha\rangle^2 \le \frac{4\epsilon}{\varphi^2}$$

□

The following lemma shows that the length of the cluster mean of cluster C_i is roughly $1/\sqrt{|C_i|}$ and that cluster means are almost orthogonal.

Lemma 7. *(Cluster means) Let $k \ge 2$ be an integer, $\varphi \in (0, 1)$, and $\epsilon \in (0, 1)$. Let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Then we have*

1. *for all $i \in [k]$, $\left| ||\mu_i||_2^2 - \frac{1}{|C_i|}\right| \le \frac{4\sqrt{\epsilon}}{\varphi}\frac{1}{|C_i|}$*
2. *for all $i \ne j \in [k]$, $|\langle\mu_i, \mu_j\rangle| \le \frac{8\sqrt{\epsilon}}{\varphi}\frac{1}{\sqrt{|C_i||C_j|}}$*

To prove Lemma 7 we need Lemma 9 in which we will use the following result from [HJ90] (Theorem 1.3.20 on page 53).

Lemma 8 ([HJ90])**.** *Let h, m, n be integers such that $1 \le h \le m \le n$. For any matrix $A \in \mathbb{R}^{m\times n}$ and matrix $B \in \mathbb{R}^{n\times m}$, the multisets of nonzero eigenvalues of AB and BA are equal. In particular, if one of AB and BA is positive semidefinite, then $\nu_h(AB) = \nu_h(BA)$.*

Lemma 9. *Let $k \ge 2$ be an integer, $\varphi \in (0, 1)$, and $\epsilon \in (0, 1)$. Let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $H \in \mathbb{R}^{k\times k}$ be a matrix whose i-th column is μ_i. Let $W \in \mathbb{R}^{k\times k}$ be a diagonal matrix such that $W(i, i) = \sqrt{|C_i|}$. Then for any $\alpha \in \mathbb{R}^k$, $\|\alpha\| = 1$, we have*

1. $|\alpha^T \left((HW)(HW)^T - I\right) \alpha| \le \frac{4\sqrt{\epsilon}}{\varphi}$

2. $|\alpha^T \left((HW)^T(HW) - I\right) \alpha| \le \frac{4\sqrt{\epsilon}}{\varphi}$

Proof. **Proof of item** (1)**:** Let $Y \in \mathbb{R}^{k\times n}$ denote a matrix whose x-th column is μ_x for any $x \in V$. Note that

$$YY^T = \sum_{i=1}^{k} |C_i|\mu_i\mu_i^T = (HW)(HW)^T.$$

We define $\widetilde{z} := Y^T\alpha$, and $z := U_{[k]}\alpha$. Note that $U_{[k]}^T U_{[k]} = I$. Therefore we have

$$\begin{aligned}
|\alpha^T \left((HW)(HW)^T - I\right) \alpha| &= |\alpha^T(YY^T - U_{[k]}^T U_{[k]})\alpha| \\
&= \left|\sum_{x\in V} \widetilde{z}(x)^2 - z(x)^2\right| && \text{From definition of } z(x) \text{ and } \widetilde{z}(x) \\
&= \left|\sum_{x\in V} \left(z(x) - \widetilde{z}(x)\right)\left(z(x) + \widetilde{z}(x)\right)\right| \\
&\le \sqrt{\sum_{x\in V}(z(x) - \widetilde{z}(x))^2 \sum_{x\in V}(\widetilde{z}(x) + z(x))^2} && \text{By Cauchy-Schwarz inequality}
\end{aligned} \tag{16}$$

Note that for any $x \in V$, we have $z(x) = \langle f_x, \alpha\rangle$ and $\widetilde{z}(x) = \langle \mu_x, \alpha\rangle$. Therefore by Lemma 6 we have

$$\sqrt{\sum_{x\in V}(z(x) - \widetilde{z}(x))^2} = \sqrt{\sum_{x\in V} \langle f_x - \mu_x, \alpha\rangle^2} \le \frac{2\sqrt{\epsilon}}{\varphi} \tag{17}$$

To complete the proof it suffices to show that $\sum_{x\in V}(\tilde{z}(x) + z(x))^2 \le 4$. Note that

$$\begin{aligned}
\sum_{x\in V} \tilde{z}(x)^2 &= \sum_{x\in V} \langle \alpha, \mu_x\rangle^2 \\
&= \sum_i |C_i| \left\langle \alpha, \frac{\sum_{x\in C_i} f_x}{|C_i|}\right\rangle^2 \\
&= \sum_i |C_i| \left(\frac{\sum_{x\in C_i} \langle \alpha, f_x\rangle}{|C_i|}\right)^2 \\
&\le \sum_i \sum_{x\in C_i} \langle \alpha, f_x\rangle^2 && \text{By Jensen's inequality} \\
&= \sum_{x\in V} z(x)^2
\end{aligned}$$

Thus we have

$$\sum_{x\in V}(\tilde{z}(x) + z(x))^2 \le \sum_{x\in V} 2(\tilde{z}(x)^2 + z(x)^2) \le 2 + 2\sum_{x\in V} \tilde{z}(x)^2 \le 4 \tag{18}$$

In the first inequality we used the fact that $(\tilde{z}(x) - z(x))^2 \ge 0$ and for the second inequality we used the fact that $||z||_2^2 = ||U_{[k]}\alpha||_2^2 = 1$. Putting (18), (17), and (16) together we get

$$|\alpha^T \left((HW)(HW)^T - I\right) \alpha| \le \frac{4\sqrt{\epsilon}}{\varphi}.$$

Proof of item (2)**:** Note that by item (2) for any vector α with $||\alpha||_2 = 1$ we have

$$1 - \frac{4\sqrt{\epsilon}}{\varphi} \le \alpha^T \left((HW)(HW)^T\right) \alpha \le 1 + \frac{4\sqrt{\epsilon}}{\varphi}$$

Thus by Lemma 8 we have that the set of eigenvalues of $(HW)(HW)^T$ and $(HW)^T(HW)$ are the same, and all of the eigenvalues lie in the interval $[1 - \frac{4\sqrt{\epsilon}}{\varphi}, 1 + \frac{4\sqrt{\epsilon}}{\varphi}]$. Thus for any vector α with $||\alpha||_2 = 1$ we have

$$1 - \frac{4\sqrt{\epsilon}}{\varphi} \le \alpha^T \left((HW)^T(HW)\right) \alpha \le 1 + \frac{4\sqrt{\epsilon}}{\varphi}.$$

□

Now we are able to prove Lemma 7.

Lemma 7. *(Cluster means) Let $k \geq 2$ be an integer, $\varphi \in (0,1)$, and $\epsilon \in (0,1)$. Let $G = (V,E)$ be a d-regular graph that admits (k,φ,ϵ)-clustering $C_1,\ldots,C_k$. Then we have*

1. *for all $i \in [k]$, $\left| ||\mu_i||_2^2 - \frac{1}{|C_i|} \right| \leq \frac{4\sqrt{\epsilon}}{\varphi}\frac{1}{|C_i|}$*
2. *for all $i \neq j \in [k]$, $|\langle \mu_i, \mu_j\rangle| \leq \frac{8\sqrt{\epsilon}}{\varphi}\frac{1}{\sqrt{|C_i||C_j|}}$*

Proof. **Proof of item** (1)**:** Let $H \in \mathbb{R}^{k\times k}$ be a matrix whose i-th column is μ_i. Let $W \in \mathbb{R}^{k\times k}$ be a diagonal matrix whose such that $W(i,i) = \sqrt{|C_i|}$. Thus by Lemma 9 item (2) for any $\alpha \in \mathbb{R}^k$ with $\|\alpha\| = 1$, we have

$$|\alpha^T\left((HW)^T(HW) - I\right)\alpha| \leq \frac{4\sqrt{\epsilon}}{\varphi}$$

Let $\alpha = \mathbb{1}_i$. Thus we have

$$|((HW)^T(HW))(i,i) - 1| \leq \frac{4\sqrt{\epsilon}}{\varphi} \tag{19}$$

Note that $((HW)^T(HW))(i,i) = (WH^THW)(i,i) = ||\mu_i||_2^2|C_i|$.Therefore we get

$$\left| ||\mu_i||_2^2 - \frac{1}{|C_i|}\right| \leq \frac{4\sqrt{\epsilon}}{\varphi}\cdot\frac{1}{|C_i|}$$

Proof of item (2)**:** Let $\alpha = \frac{1}{\sqrt{2}}(\mathbb{1}_i + \mathbb{1}_j)$. Note that $||\alpha||_2 = 1$. Thus by Lemma 9 item (2) we have

$$|\alpha^T\left((HW)^T(HW) - I\right)\alpha| \leq \frac{4\sqrt{\epsilon}}{\varphi}$$

Note that

$$\left|\alpha^T\left((HW)^T(HW) - I\right)\alpha\right| = \left|\frac{1}{2}\left(||\mu_i||_2^2|C_i| + ||\mu_j||_2^2|C_j| + 2\langle\mu_i,\mu_j\rangle\sqrt{|C_i||C_j|} - 2\right)\right|$$

Therefore we get

$$\left| ||\mu_i||_2^2|C_i| + ||\mu_j||_2^2|C_j| + 2\langle\mu_i,\mu_j\rangle\sqrt{|C_i||C_j|} - 2\right| \leq \frac{8\sqrt{\epsilon}}{\varphi}$$

Thus

$$\begin{aligned}\left|\langle\mu_i,\mu_j\rangle\sqrt{|C_i||C_j|}\right| &\leq \left|\frac{1}{2}\left(1 - ||\mu_i||_2^2|C_i|\right) + \frac{1}{2}\left(1 - ||\mu_j||_2^2|C_j|\right)\right| + \frac{4\sqrt{\epsilon}}{\varphi} \\ &\leq \frac{1}{2}\cdot\frac{4\sqrt{\epsilon}}{\varphi} + \frac{1}{2}\cdot\frac{4\sqrt{\epsilon}}{\varphi} + \frac{4\sqrt{\epsilon}}{\varphi} && \text{By item (1)} \\ &\leq \frac{8\sqrt{\epsilon}}{\varphi}\end{aligned}$$

Therefore we get

$$|\langle\mu_i,\mu_j\rangle| \leq \frac{8\sqrt{\epsilon}}{\varphi}\cdot\frac{1}{\sqrt{|C_i||C_j|}}.$$

□

4.2 Strong Tail Bounds on the Spectral Embedding

The main results of this section are the following two lemmas. The first lemma gives an upper bound on the length of the projection of any point f_x on an arbitrary direction $\alpha \in \mathbb{R}^k$. The second lemma considers the distribution of the lengths of projected f_x and we get tail bounds that show that the fraction of points whose projected length exceeds the 'expectation' (which is about $1/\sqrt{|C_i|}$ for the smallest cluster C_i) by a factor of β is bounded by $\beta^{-\varphi^2/10\epsilon}$. In other words, we bound the $O(\varphi^2/\epsilon)$-th moment as opposed to the second moment, which gives us tight control over the embedding when $\epsilon/\varphi^2 \ll 1/\log k$.

Lemma 5. *Let $\varphi \in (0,1)$ and $\epsilon \le \frac{\varphi^2}{100}$, and let $G=(V,E)$ be a d-regular graph that admits (k,φ,ϵ)-clustering $C_1,\dots,C_k$. Let u be a normalized eigenvector of L with $||u||_2 = 1$ and with eigenvalue at most 2ϵ. Then we have*

$$||u||_\infty \le n^{20\cdot\epsilon/\varphi^2} \cdot \sqrt{\frac{160}{\min_{i\in k}|C_i|}}.$$

Lemma 4. *[Tail-bound] Let $\varphi \in (0,1)$ and $\epsilon \le \frac{\varphi^2}{100}$, and let $G=(V,E)$ be a d-regular graph that admits (k,φ,ϵ)-clustering $C_1,\dots,C_k$. Let L be the normalized Laplacian of G. Let u be a normalized eigenvector of L with $||u||_2 = 1$ and with eigenvalue at most 2ϵ. Then for any $\beta > 1$ we have*

$$\frac{1}{n}\cdot\left|\left\{x\in V: |u(x)| \ge \beta\cdot\sqrt{\frac{10}{\min_{i\in[k]}|C_i|}}\right\}\right| \le \left(\frac{\beta}{2}\right)^{-\varphi^2/20\cdot\epsilon}.$$

We are interested in deriving moment bounds for the distribution of the entries of the first k eigenvectors u of L (i.e., eigenvectors with eigenvalue smaller than 2ϵ), and specifically in the distribution of the absolute values of the entries of u. In order to be able to analyze this distribution, we define the sets of all entries in u that are bigger than a threshold θ:

Definition 6 (Threshold sets)**.** Let $G=(V,E)$ be a graph with normalized Laplacian L. Let u be a normalized eigenvector of L with $||u||_2 = 1$. Then for the vector u and a threshold $\theta \in \mathbb{R}^+$ we define the threshold set $S(\theta)$ with respect to the eigenvector u and threshold θ as

$$S(\theta) := \{x \in V : u(x) \ge \theta\}.$$

Our arguments will use that for every vertex x, we have $u(x) \approx \frac{1}{d}\sum_{\{x,y\}\in E} u(y)$. So nodes neighboring other nodes with large $u(\cdot)$ values are likely to have large $u(\cdot)$ values as well. This motivates the following definition of the potential of a threshold set.

Definition 7 (Potential of a threshold set)**.** Let $G=(V,E)$ be a graph with normalized Laplacian L. Let u be a normalized eigenvector of L with $||u||_2 = 1$. Then for vector u and a threshold $\theta \in \mathbb{R}^+$ we define the potential of a threshold set $S(\theta)$ as

$$p(\theta) = \sum_{x\in S(\theta)} u(x).$$

We start by proving a core bound on the threshold sets (Lemma 10 below) that forms the basis of our approach: the main technical results of this section (Lemma 5 and Lemma 4) essentially follow by repeated application of Lemma 10. Specifically, we now argue that if a threshold set $S(\theta)$ expands in the graph G and the relative potential of the set (i.e., $p(\theta)/|S(\theta)|$) is at most 2θ, then we can slightly decrease θ to obtain a new θ' such that the corresponding threshold set is a constant factor larger that $S(\theta)$ and the relative potential is bounded by $2\theta'$.

Lemma 10 (Threshold shift for expanding threshold sets)**.** *Let $G=(V,E)$ be a d-regular graph with normalized Laplacian L. Let u be a normalized eigenvector of L with $||u||_2 = 1$ and with eigenvalue $\lambda \le 2\epsilon$. Let $\theta \in \mathbb{R}^+$ be a threshold. Suppose that $S(\theta)$ is the threshold set with respect to u and θ such that $S(\theta)$ is non-empty, $\phi^G(S(\theta)) \ge \varphi$ and $\frac{p(\theta)}{|S(\theta)|} \le 2\theta$. Then the following holds for $\theta' = \theta\left(1-\frac{8\epsilon}{\varphi}\right)$:*

1. *$|S(\theta')| \ge (1+\varphi/2)|S(\theta)|$, and*
2. *$\frac{p(\theta')}{|S(\theta')|} \le 2\theta'$.*

Proof. **Proof of item** (1)**:** Note that $\lambda u = Lu = (I - \frac{A}{d})u$. Thus for any $x \in V$ we have $(Lu)(x) = u(x) - \frac{1}{d}\sum_{\{x,y\}\in E} u(y)$. Thus we have,

$$u(x) - \frac{1}{d}\sum_{\{x,y\}\in E} u(y) = \lambda\cdot u(x).$$

We write the above as

$$\sum_{y\in\mathcal{N}(x)} (u(x)-u(y)) = d\cdot\lambda\cdot u(x), \tag{20}$$

where $\mathcal{N}(x) = \{y \in V : \exists\{x,y\} \in E\}$. Summing (20) over all $x \in S(\theta)$ we get

$$\sum_{x\in S(\theta)} \sum_{y\in\mathcal{N}(x)} (u(x) - u(y)) = \sum_{x\in S(\theta)} \lambda \cdot d \cdot u(x) = \lambda \cdot d \cdot p(\theta), \tag{21}$$

and note that

$$\sum_{x\in S(\theta)} \sum_{y\in\mathcal{N}(x)} (u(x) - u(y)) = \sum_{\substack{\{x,y\}\in E \\ x\in S(\theta), y\notin S(\theta)}} (u(x) - u(y)). \tag{22}$$

For any edge $e = \{x, y\} \in E$, we define $\Delta(e) = |u(x) - u(y)|$. Note that for any $e = \{x, y\}$ such that $x \in S(\theta)$ and $y \notin S(\theta)$ we have $u(x) \geq \theta > u(y)$, hence $\Delta(e) = u(x) - u(y)$. Therefore, putting (22) and (21) together we get

$$\sum_{e\in E(S(\theta),V\setminus S(\theta))} \Delta(e) = \lambda \cdot d \cdot p(\theta).$$

By an averaging argument there exists a set $E_L \subseteq E(S_\theta, V \setminus S_\theta)$ such that $|E_L| \geq \frac{|E(S(\theta),V\setminus S(\theta))|}{2}$ and all edges $e \in E_L$ satisfy $\Delta(e) \leq \frac{2\cdot\lambda\cdot d\cdot p(\theta)}{|E(S(\theta),V\setminus S(\theta))|}$. We define V_L as a subset of vertices of $V \setminus S(\theta)$ that are connected to vertices of $S(\theta)$ by edges in E_L, i.e.

$$V_L = \{y \in V \setminus S(\theta) : \exists \; \{x, y\} \in E_L, x \in S(\theta)\}.$$

Note that

$$|V_L| \geq \frac{|E_L|}{d} \geq \frac{|E(S(\theta), V \setminus S(\theta))|}{2d}. \tag{23}$$

Using the assumption of the lemma that $\phi^G(S(\theta)) \geq \varphi$ we obtain

$$|E(S(\theta), V \setminus S(\theta))| \geq \varphi \cdot d \cdot |S(\theta)|. \tag{24}$$

Putting (24) and (23) together we get

$$|V_L| \geq \frac{\varphi|S(\theta)|}{2}. \tag{25}$$

Recall that for all $e \in E_L$ we have $\Delta(e) \leq \frac{2\cdot\lambda\cdot d\cdot p(\theta)}{|E(S(\theta),V\setminus S(\theta))|}$. We have $\lambda \leq 2\epsilon$, therefore for all $e \in E_L$ we have $\Delta(e) \leq \frac{4\cdot\epsilon\cdot d\cdot p(\theta)}{|E(S(\theta),V\setminus S(\theta))|}$. Thus for all $y \in V_L$ we get

$$u(y) \geq \theta - \frac{4 \cdot \epsilon \cdot d \cdot p(\theta)}{|E(S(\theta), V \setminus S(\theta))|}. \tag{26}$$

By the assumption of the lemma we have $\frac{p(\theta)}{|S(\theta)|} \leq 2\theta$, hence, by inequality (24) we get

$$\theta - \frac{4 \cdot \epsilon \cdot d \cdot p(\theta)}{|E(S(\theta), V \setminus S(\theta))|} \geq \theta - \frac{4 \cdot \epsilon \cdot d \cdot p(\theta)}{\varphi \cdot d \cdot |S(\theta)|} = \theta - \frac{4\epsilon}{\varphi} \cdot \frac{p(\theta)}{|S(\theta)|} \geq \theta\left(1 - \frac{8\epsilon}{\varphi}\right). \tag{27}$$

Putting (27) and (26) together we get for all $y \in V_L$, $u(y) \geq \theta\left(1 - \frac{8\epsilon}{\varphi}\right)$. Let $\theta' := \theta(1 - \frac{8\epsilon}{\varphi})$. Thus

$$S(\theta) \cup V_L \subseteq S(\theta').$$

By definition of V_L we have $V_L \cap S(\theta) = \emptyset$. Therefore, $|S(\theta')| \geq |S(\theta)| + |V_L|$. Thus by inequality (25) we get

$$|S(\theta')| \geq |S(\theta)|\left(1 + \frac{\varphi}{2}\right). \tag{28}$$

This concludes the proof of the first part of the lemma.

Proof of item (2)**:** Now using that for all $x \notin S(\theta)$ we have $u(x) < \theta$ and that $p(\theta) \leq 2\theta|S(\theta)|$ by assumption of the lemma we obtain

$$\begin{aligned}
p(\theta') &= \sum_{u\in S(\theta')} u(x) \\
&= \sum_{x\in S(\theta)} u(x) + \sum_{x\in S(\theta')\setminus S(\theta)} u(x) \\
&\leq p(\theta) + \theta|S(\theta') \setminus S(\theta)| \\
&\leq 2\theta|S(\theta)| + \theta|S(\theta') \setminus S(\theta)|. && \text{Since } p(\theta) \leq 2\theta|S(\theta)|
\end{aligned}$$

By (28) we have $|S(\theta') \setminus S(\theta)| \geq \frac{\varphi}{2}|S(\theta)|$. Therefore, using $\epsilon \leq \frac{\varphi^2}{100}$ we get

$$\frac{p(\theta')}{|S(\theta')|} \quad \leq \quad \frac{2\theta|S(\theta)| + \theta|S(\theta') \setminus S(\theta)|}{|S(\theta)| + |S(\theta') \setminus S(\theta)|} = \theta \cdot \frac{2 + \frac{|S(\theta')\setminus S(\theta)|}{|S(\theta)|}}{1 + \frac{|S(\theta')\setminus S(\theta)|}{|S(\theta)|}} \leq \theta \cdot \frac{2 + \frac{\varphi}{2}}{1 + \frac{\varphi}{2}} \leq \theta \cdot 2\left(1 - \frac{8\epsilon}{\varphi}\right) \leq 2\theta'$$

□

We would like to apply Lemma 10 iteratively, but there is one hurdle: while the first condition on the threshold set $S(\theta)$ naturally follows as long as $S(\theta)$ is not too large (by Proposition 2), the second condition needs to be established at the beginning of the iterative process. Lemma 11 accomplishes exactly that: we prove that for any value θ_1 with threshold set $S(\theta_1)$ not empty or not too large, there exists a close value θ that meets the conditions of previous lemma.

Proposition 2. *Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. For any set $S \subseteq V$ with size $|S| \leq \frac{1}{2} \cdot \min_{i \in k} |C_i|$ we have $\phi^G(S) \geq \varphi$.*

Proof. For any $1 \leq i \leq k$ we define $S_i = S \cap C_i$. Note that

$$|S_i| \leq |S| \leq \frac{1}{2} \cdot \min_{i \in k} |C_i| \leq \frac{|C_i|}{2}.$$

Therefore since $\phi^G(C_i) \geq \varphi$ we have $E(S_i, C_i \setminus S_i) \geq \varphi d|S_i|$. Thus we get

$$E(S, V \setminus S) \geq \sum_{i=1}^{k} E(S_i, C_i \setminus S_i) \geq \varphi d \sum_{i=1}^{k} |S_i| = \varphi d|S|.$$

Hence, $\phi^G(S) \geq \varphi$. □

Lemma 11. *Let $\varphi \in (0, 1)$ and $\epsilon \leq \frac{\varphi^2}{100}$, and let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let L denote the normalized Laplacian of G. Let u be a normalized eigenvector of L with $||u||_2 = 1$ and with eigenvalue $\lambda \leq 2\epsilon$. Let $\theta_1 \in \mathbb{R}^+$ be a threshold. Let $S(\theta_1)$ be the threshold set with respect to u and θ_1. Suppose that $1 \leq |S(\theta_1)| \leq \frac{1}{2} \cdot \min_{i \in \{1,\ldots,k\}} |C_i|$. Then there exists a threshold θ_2 such that the following holds:*

1. $\theta_1 \left(1 - \frac{8\epsilon}{\varphi}\right) \leq \theta_2 \leq \theta_1$, *and*
2. $\frac{p(\theta_2)}{|S(\theta_2)|} \leq 2\theta_2$

Proof. Let

$$\theta^* := \min\left\{\theta \geq \theta_1 \;\middle|\; S(\theta) \neq \emptyset \text{ and } \frac{p(\theta)}{|S(\theta)|} \leq 2\theta\right\}.$$

We can conclude that θ^* exists, as by the assumption of the lemma we have $|S(\theta_1)| \geq 1$ and for $\theta_{\max} = \max_{x \in V} u(x)$ we have $\frac{p(\theta_{\max})}{|S(\theta_{\max})|} = \theta_{\max}$. We also have $|S(\theta^*)| \leq \min_{i \in \{1,\ldots,k\}} |C_i|/2$ as $\theta^* \geq \theta_1$ and by the assumption of the lemma. So Proposition 2 implies

$$\phi^G(S(\theta^*)) \geq \varphi. \tag{29}$$

Now Lemma 10 implies

$$\frac{p(\theta^*(1 - \frac{8\epsilon}{\varphi}))}{|S(\theta^*(1 - \frac{8\epsilon}{\varphi}))|} \leq 2\theta^* \left(1 - \frac{8\epsilon}{\varphi}\right)$$

and by minimality of θ^* we have that:

$$\theta_1 \left(1 - \frac{8\epsilon}{\varphi}\right) \leq \theta^* \left(1 - \frac{8\epsilon}{\varphi}\right) \leq \theta_1.$$

So we can set $\theta_2 := \theta^* \left(1 - \frac{8\epsilon}{\varphi}\right)$. □

We are now ready to prove our tail bound. The main idea behind the proof is to use Lemma 10 and Lemma 11 to show that if a vertex has a large entry along one of the bottom k eigenvectors this implies that many other vertices also have a relatively large value along the same eigenvector. Thus, not too many f_x can have such a large value.

Lemma 4. *[Tail-bound] Let $\varphi \in (0,1)$ and $\epsilon \leq \frac{\varphi^2}{100}$, and let $G=(V,E)$ be a d-regular graph that admits (k,φ,ϵ)-clustering $C_1,\ldots,C_k$. Let L be the normalized Laplacian of G. Let u be a normalized eigenvector of L with $||u||_2 = 1$ and with eigenvalue at most 2ϵ. Then for any $\beta > 1$ we have*

$$\frac{1}{n} \cdot \left| \left\{ x \in V : |u(x)| \geq \beta \cdot \sqrt{\frac{10}{\min_{i\in[k]} |C_i|}} \right\} \right| \leq \left(\frac{\beta}{2}\right)^{-\varphi^2/20\cdot\epsilon}.$$

Proof. Let $s_{\min} = \min_{i\in\{1,\ldots,k\}} |C_i|$. We define

$$S^+ = \left\{ x \in V : u(x) \geq \beta \cdot \sqrt{\frac{10}{s_{\min}}} \right\},$$

and

$$S^- = \left\{ x \in V : -u(x) \geq \beta \cdot \sqrt{\frac{10}{s_{\min}}} \right\}$$

Note that $-u$ is also an eigenvector of L with the same eigenvalue as u, hence, without loss of generality suppose that $|S^+| \geq |S^-|$. Let $T = \left\{ x \in V : u(x)^2 \geq \frac{10}{s_{\min}} \right\}$. Since, $1 = \|u\|_2^2 = \sum_{x\in V} u(x)^2$, an averaging argument implies $|T| \leq \frac{s_{\min}}{10}$. Let

$$T^+ = \left\{ x \in V : u(x) \geq \sqrt{\frac{10}{s_{\min}}} \right\}.$$

Note that $\beta > 1$, hence, $S^+ \subseteq T^+ \subseteq T$, and so we have $|S^+| \leq |T^+| \leq |T| \leq \frac{s_{\min}}{10}$. We may assume that S^+ is non-empty as otherwise the lemma follows immediately. Let $\theta_0 = \beta \cdot \sqrt{\frac{10}{s_{\min}}}$. Note that $S^+ = S(\theta_0)$. Hence, $1 \leq |S(\theta_0)| \leq \frac{s_{\min}}{10}$. Therefore by Lemma 11 there exists a threshold θ_1 such that

$$\left(1 - \frac{8\epsilon}{\varphi}\right) \beta \cdot \sqrt{\frac{10}{s_{\min}}} \leq \theta_1 \leq \beta \cdot \sqrt{\frac{10}{s_{\min}}}, \text{ and} \tag{30}$$

$$\frac{p(\theta_1)}{|S(\theta_1)|} \leq 2\theta_1.$$

For any $t \geq 1$ we define $\theta_{t+1} = \theta_t(1 - \frac{8\epsilon}{\varphi})$. For some $t' \geq 0$ we must have $\theta_{t'+1} \leq \sqrt{\frac{10}{s_{\min}}} \leq \theta_{t'}$. Thus by (30) we have

$$\theta_{t'} = \left(1 - \frac{8\epsilon}{\varphi}\right)^{t'-1} \theta_1 \geq \left(1 - \frac{8\epsilon}{\varphi}\right)^{t'} \cdot \beta \cdot \sqrt{\frac{10}{s_{\min}}}, \tag{31}$$

and

$$\theta_{t'} \leq \frac{\theta_{t'+1}}{\left(1 - \frac{8\epsilon}{\varphi}\right)} \leq \frac{\sqrt{\frac{10}{s_{\min}}}}{\left(1 - \frac{8\epsilon}{\varphi}\right)} \tag{32}$$

Putting (31) and (32) together we get

$$\beta \leq \left(1 - \frac{8\epsilon}{\varphi}\right)^{-t'-1} \tag{33}$$

Recall that for all $t \geq 1$ we have $\theta_{t+1} = \theta_t(1 - \frac{8\epsilon}{\varphi})$, thus

$$S^+ = S(\theta_0) \subseteq S(\theta_1) \subseteq S(\theta_2) \subseteq \ldots \subseteq S(\theta_{t'}) \subseteq T^+.$$

Therefore for all $0 \leq t \leq t'$ we have

$$|S^+| \leq |S(\theta_t)| \leq |T^+| \leq \frac{s_{\min}}{10}. \tag{34}$$

Since $|S(\theta_t)| \leq \frac{\min_{i\in\{1,\ldots,k\}} |C_i|}{10} = \frac{s_{\min}}{10}$, by Lemma 10 for all $1 \leq t \leq t'$ we have

$$|S(\theta_{t+1})| \geq |S(\theta_t)| \left(1 + \frac{\varphi}{2}\right). \tag{35}$$

Therefore

$$
\begin{aligned}
t' &\leq \log_{1+\frac{\varphi}{2}}\left(\frac{|T^+|}{|S^+|}\right) && \text{By (35)} \\
&\leq \log_{1+\frac{\varphi}{2}}\left(\frac{s_{\min}}{10\cdot|S^+|}\right) && \text{By (34)} \\
&\leq \log_{1+\frac{\varphi}{2}}\left(\frac{s_{\min}}{5\cdot|S^+\cup S^-|}\right) && \text{By the assumption } |S^+|\geq|S^-| \qquad (36)
\end{aligned}
$$

Putting (33) and (36) together we get

$$
\begin{aligned}
\beta &\leq \left(1-\frac{8\epsilon}{\varphi}\right)^{-t'-1} && \text{By (33)} \\
&\leq \left(1-\frac{8\epsilon}{\varphi}\right)^{-1-\log_{1+\frac{\varphi}{2}}\left(\frac{s_{\min}}{5\cdot|S^+\cup S^-|}\right)} && \text{By (36)} \\
&\leq 2\cdot\left(\frac{s_{\min}}{5\cdot|S^+\cup S^-|}\right)^{-\log_{1+\frac{\varphi}{2}}\left(1-\frac{8\epsilon}{\varphi}\right)} && \text{Since } \frac{\epsilon}{\varphi^2}\leq\frac{1}{100} \qquad (37)
\end{aligned}
$$

Note that for any $x\in\mathbb{R}$ we have $1+x\leq e^x$, and for any $x<0.01$ we have $1-x\geq e^{-1.2x}$, thus given $\frac{\epsilon}{\varphi}<0.01$ we have

$$
\log_{1+\frac{\varphi}{2}}\left(1-\frac{8\epsilon}{\varphi}\right)=\frac{\ln\left(1-\frac{8\epsilon}{\varphi}\right)}{\ln\left(1+\frac{\varphi}{2}\right)}\geq\frac{-\frac{10\epsilon}{\varphi}}{\frac{\varphi}{2}}\geq-\frac{20\cdot\epsilon}{\varphi^2} \tag{38}
$$

Putting (37) and (38) together we get

$$
\frac{\beta}{2}\leq\left(\frac{s_{\min}}{5\cdot|S^+\cup S^-|}\right)^{(20\cdot\epsilon/\varphi^2)}
$$

Therefore we have

$$
|S^+\cup S^-|\leq s_{\min}\cdot\left(\frac{\beta}{2}\right)^{-\left(\varphi^2/20\cdot\epsilon\right)}\leq n\cdot\left(\frac{\beta}{2}\right)^{-\left(\varphi^2/20\cdot\epsilon\right)}.
$$

□

As a consequence of our tail bound we can prove a bound on ℓ_∞-norm on any unit vector in the eigenspace spanned by the bottom k eigenvectors of L, i.e. $U_{[k]}$.

Lemma 5. *Let $\varphi\in(0,1)$ and $\epsilon\leq\frac{\varphi^2}{100}$, and let $G=(V,E)$ be a d-regular graph that admits (k,φ,ϵ)-clustering $C_1,\ldots,C_k$. Let u be a normalized eigenvector of L with $||u||_2=1$ and with eigenvalue at most 2ϵ. Then we have*

$$
||u||_\infty\leq n^{20\cdot\epsilon/\varphi^2}\cdot\sqrt{\frac{160}{\min_{i\in k}|C_i|}}.
$$

Proof. We define

$$
S=\left\{x\in V:|u(x)|\geq n^{20\epsilon/\varphi^2}\cdot\sqrt{\frac{160}{\min_{i\in k}|C_i|}}\right\}
$$

Let $\beta=4\cdot n^{20\epsilon/\varphi^2}$. By Lemma 4 we have

$$
|S|\leq n\cdot\left(\frac{\beta}{2}\right)^{-\varphi^2/20\cdot\epsilon}\leq n\cdot\left(2\cdot n^{20\epsilon/\varphi^2}\right)^{-\varphi^2/20\cdot\epsilon}<1
$$

Therefore $S=\emptyset$, hence

$$
\|u\|_\infty\leq n^{20\epsilon/\varphi^2}\cdot\sqrt{\frac{160}{\min_{i\in k}|C_i|}}.
$$

□

4.3 Centers are strongly orthogonal

The main result of this section is Lemma 12 which generalizes Lemma 7 to the orthogonal projection of cluster centers into the subspace spanned by some of the centers. To prove Lemma 12 we first need to prove Lemma 13, Lemma 14 and Lemma 15.

Lemma 12. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than an absolute positive constant. Let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $S \subset \{\mu_1, \ldots, \mu_k\}$ denote a subset of cluster means. Let $\Pi \in \mathbb{R}^{k \times k}$ denote the orthogonal projection matrix onto $span(S)^\perp$. Then the following holds:*

1. *For all $\mu_i \in \{\mu_1, \ldots, \mu_k\} \setminus S$ we have $\left| \|\Pi\mu_i\|_2^2 - ||\mu_i||_2^2 \right| \leq \frac{16\sqrt{\epsilon}}{\varphi} \cdot ||\mu_i||_2^2$.*
2. *For all $\mu_i \neq \mu_j \in \{\mu_1, \ldots, \mu_k\} \setminus S$ we have $|\langle \Pi\mu_i, \Pi\mu_j \rangle| \leq \frac{40\sqrt{\epsilon}}{\varphi} \cdot \frac{1}{\sqrt{|C_i| \cdot |C_j|}}$.*

Matrix $A \in \mathbb{R}^n$ is poitive definite if $x^T Ax > 0$ for all $x \neq 0$, and it is positive semidefinite if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$. We write $A \succ 0$ to indicate that A is positive definite, and $A \succcurlyeq 0$ to indicate that it is positive semidefinite. We use the semidefinite ordering on matrices, writing $A \succcurlyeq B$ if and only if $A - B \succcurlyeq 0$.

Theorem 4 ([Tod11])**.** *Let $A, B \in \mathbb{R}^{n \times n}$ be invertible, positive definite matrices. Then $A \succcurlyeq B \implies B^{-1} \succcurlyeq A^{-1}$.*

Proof. By symmetry, we only need to show $A \succcurlyeq B \implies B^{-1} \succcurlyeq A^{-1}$. Since $B \succ 0$ for any $x, y \in \mathbb{R}^n$ we obtain

$$\begin{aligned} 0 &\leq \left\langle y - B^{-1}x, B(y - B^{-1}x) \right\rangle \\ &= \langle y, By \rangle - \langle y, x \rangle - \left\langle B^{-1}x, By \right\rangle + \left\langle x, B^{-1}x \right\rangle \\ &= \langle y, By \rangle - 2\langle x, y \rangle + \left\langle x, B^{-1}x \right\rangle \end{aligned}$$

so

$$2\langle x, y \rangle - \langle y, By \rangle \leq \left\langle x, B^{-1}x \right\rangle \tag{39}$$

Since $A \succcurlyeq B$ it follows from (39) that

$$2\langle x, y \rangle - \langle y, Ay \rangle \leq 2\langle x, y \rangle - \langle y, Ay \rangle \leq \left\langle x, B^{-1}x \right\rangle \tag{40}$$

Letting $y = A^{-1}x$ in the leftmost expression of (40) we obtain

$$\left\langle x, A^{-1}x \right\rangle \leq \left\langle x, B^{-1}x \right\rangle$$

Since $x \in \mathbb{R}^n$ is is arbitrary, we get $B^{-1} \succcurlyeq A^{-1}$. □

Lemma 13. *Let $H, \widetilde{H} \in \mathbb{R}^{n \times n}$ be invertible, positive definite matrices. Let $\delta < 1$. Suppose that for any vector $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$ we have $(1-\delta)x^T Hx \leq x^T \widetilde{H} x \leq (1+\delta)x^T Hx$. Then for any vector $y \in \mathbb{R}^n$ with $\|y\|_2 = 1$ we have $\frac{1}{1+\delta} y^T H^{-1} y \leq y^T \widetilde{H}^{-1} y \leq \frac{1}{1-\delta} y^T H^{-1} y$.*

Proof. Note that we have $(1-\delta)H \preceq \widetilde{H} \preceq (1+\delta)H$ therefore, by Theorem 4 we have

$$\frac{1}{(1-\delta)} \cdot H^{-1} \succcurlyeq \widetilde{H}^{-1} \succcurlyeq \frac{1}{(1+\delta)} \cdot H^{-1}$$

□

Lemma 14. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than an absolute positive constant. Let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $S = \{\mu_1, \ldots, \mu_k\} \setminus \{\mu_i\}$. Let $H = [\mu_1, \mu_2, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_k]$ denote a matrix such that its columns are the vectors in S. Let $W \in \mathbb{R}^{(k-1)\times(k-1)}$ denote a diagonal matrix such that for all $j < i$ we have $W(j,j) = \sqrt{|C_j|}$ and for all $j \geq i$ we have $W(j,j) = \sqrt{|C_{j+1}|}$. Let $Z = HW$. Then $Z^T Z$ is invertible, and for any vector $x \in \mathbb{R}^{k-1}$ with $||x||_2 = 1$ we have*

$$|x^T((Z^T Z)^{-1} - I)x| \leq \frac{5\sqrt{\epsilon}}{\varphi}.$$

Proof. Let $Y \in \mathbb{R}^{k\times k}$ be a matrix, whose i-th column is equal to $\sqrt{C_i} \cdot \mu_i$. By Lemma 9 item (2) for any vector $z \in \mathbb{R}^k$ with $||\alpha||_2 = 1$ we have

$$|\alpha^T (Y^T Y - I)\alpha| \leq \frac{4\sqrt{\epsilon}}{\varphi}$$

Let $x \in \mathbb{R}^{k-1}$ be a vector with $||x||_2 = 1$, and let $\alpha \in \mathbb{R}^k$ be a vector defined as follows:

$$\alpha_j = \begin{cases} x_j & j < i \\ 0 & j = i \\ x_{j+1} & j > i \end{cases}$$

Thus we have $||\alpha||_2 = ||x||_2 = 1$ and $Y\alpha = Zx$. Hence, we get

$$|x^T (Z^T Z - I)x| = |\alpha^T (Y^T Y - I)\alpha| \leq \frac{4\sqrt{\epsilon}}{\varphi}$$

Thus for any vector $x \in \mathbb{R}^{k-1}$ with $||x||_2 = 1$ we have

$$1 - \frac{4\sqrt{\epsilon}}{\varphi} \leq x^T (Z^T Z)x \leq 1 + \frac{4\sqrt{\epsilon}}{\varphi}$$

Note that $Z^T Z$ is symmetric and positive semidefinit. Also note that $Z^T Z$ is spectrally close to I, hence, $Z^T Z$ is invertible. Thus by Lemma 13 for any vector $x \in \mathbb{R}^{k-1}$ we have

$$1 - \frac{5\sqrt{\epsilon}}{\varphi} \leq x^T (Z^T Z)^{-1} x \leq 1 + \frac{5\sqrt{\epsilon}}{\varphi}$$

Therefore we get

$$|x^T ((Z^T Z)^{-1} - I)x| \leq \frac{5\sqrt{\epsilon}}{\varphi}.$$

□

Lemma 15. *Let $k \geq 2$ be an integer, $\varphi \in (0, 1)$, and $\epsilon \in (0, 1)$. Let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $S = \{\mu_1, \ldots, \mu_k\}\backslash\{\mu_i\}$. Let $H = [\mu_1, \mu_2, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_k]$ denote a matrix such that its columns are the vectors in S. Let $W \in \mathbb{R}^{(k-1)\times(k-1)}$ denote a diagonal matrix such that for all $j < i$ we have $W(j, j) = \sqrt{|C_j|}$ and for all $j \geq i$ we have $W(j, j) = \sqrt{|C_{j+1}|}$. Let $Z = HW$. Then we have*

$$\mu_i^T Z Z^T \mu_i \leq \frac{8\sqrt{\epsilon}}{\varphi} \cdot ||\mu_i||_2^2.$$

Proof. Note that $ZZ^T = (\sum_{j=1}^k |C_j|\mu_j\mu_j^T) - |C_i|\mu_i\mu_i^T$. Thus we have

$$\mu_i^T Z Z^T \mu_i = \mu_i^T \left(\sum_{j=1}^k |C_j|\mu_j\mu_j^T\right) \mu_i - |C_i| \cdot ||\mu_i||_2^4. \tag{41}$$

By Lemma 9 item (1) for any vector x with $||x||_2 = 1$ we have

$$x^T \left(\sum_{j=1}^k |C_j|\mu_j\mu_j^T - I\right) x \leq \frac{4\sqrt{\epsilon}}{\varphi}$$

Hence we can write

$$\mu_i^T \left(\sum_{j=1}^k |C_j|\mu_j\mu_j^T\right) \mu_i = \mu_i^T \left(\sum_{j=1}^k |C_j|\mu_j\mu_j^T - I\right) \mu_i + \mu_i^T \mu_i \leq \left(1 + \frac{4\sqrt{\epsilon}}{\varphi}\right) ||\mu_i||_2^2$$

Therefore by (41) we get

$$\mu_i^T Z Z^T \mu_i = \mu_i^T \left(\sum_{j=1}^{k} |C_j| \mu_j \mu_j^T \right) \mu_i - |C_i| \cdot ||\mu_i||_2^4$$
$$\leq \left(1 + \frac{4\sqrt{\epsilon}}{\varphi} - |C_i| \cdot ||\mu_i||_2^2\right) ||\mu_i||_2^2$$

By Lemma 7 we have $|C_i| \cdot ||\mu_i||_2^2 \geq \left(1 - \frac{4\sqrt{\epsilon}}{\varphi}\right)$. Thus we get

$$\mu_i^T Z Z^T \mu_i \leq \left(1 + \frac{4\sqrt{\epsilon}}{\varphi} - |C_i| \cdot ||\mu_i||_2^2\right) ||\mu_i||_2^2$$
$$\leq \left(1 + \frac{4\sqrt{\epsilon}}{\varphi} - 1 + \frac{4\sqrt{\epsilon}}{\varphi}\right) ||\mu_i||_2^2$$
$$\leq \frac{8\sqrt{\epsilon}}{\varphi} \cdot ||\mu_i||_2^2$$

□

Now we prove the main result of the subsection (Lemma 12).

Lemma 12. *Let $k \geq 2$, $\varphi \in (0, 1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than an absolute positive constant. Let $G = (V, E)$ be a d-regular graph that admits (k, φ, ϵ)-clustering $C_1, \dots, C_k$. Let $S \subset \{\mu_1, \dots, \mu_k\}$ denote a subset of cluster means. Let $\Pi \in \mathbb{R}^{k \times k}$ denote the orthogonal projection matrix onto span$(S)^\perp$. Then the following holds:*

1. *For all $\mu_i \in \{\mu_1, \dots, \mu_k\} \setminus S$ we have $\left| ||\Pi\mu_i||_2^2 - ||\mu_i||_2^2 \right| \leq \frac{16\sqrt{\epsilon}}{\varphi} \cdot ||\mu_i||_2^2$.*
2. *For all $\mu_i \neq \mu_j \in \{\mu_1, \dots, \mu_k\} \setminus S$ we have $|\langle \Pi\mu_i, \Pi\mu_j \rangle| \leq \frac{40\sqrt{\epsilon}}{\varphi} \cdot \frac{1}{\sqrt{|C_i| \cdot |C_j|}}$.*

Proof. **Proof of item (1):** Since Π is a orthogonal projection matrix we have $||\Pi||_2 = 1$. Hence, we have $||\Pi\mu_i||_2^2 \leq ||\mu_i||_2^2 \leq \left(1 + \frac{16\sqrt{\epsilon}}{\varphi}\right) ||\mu_i||_2^2$. Thus it's left to prove $||\Pi\mu_i||_2^2 \geq \left(1 - \frac{16\sqrt{\epsilon}}{\varphi}\right) ||\mu_i||_2^2$. Note that by Pythagoras' theorem $||\Pi\mu_i||_2^2 = ||\mu_i||_2^2 - ||(I - \Pi)\mu_i||_2^2$. We will prove $||(I - \Pi)\mu_i||_2^2 \leq \frac{16\sqrt{\epsilon}}{\varphi} ||\mu_i||_2^2$ which implies

$$||\Pi\mu_i||_2^2 \geq \left(1 - 16\frac{\sqrt{\epsilon}}{\varphi}\right) ||\mu_i||_2^2.$$

Let $S' = \{\mu_1, \dots, \mu_k\} \setminus \{\mu_i\}$. Let Π' denote the orthogonal projection matrix onto span$(S')^\perp$. Note that $S \subseteq S'$, hence span(S) is a subspace of span(S'), therefore we have $||(I - \Pi)\mu_i||_2^2 \leq ||(I - \Pi')\mu_i||_2^2$. Thus it suffices to prove $||(I - \Pi')\mu_i||_2^2 \leq \frac{16\sqrt{\epsilon}}{\varphi} ||\mu_i||_2^2$. Let $H = [\mu_1, \mu_2, \dots, \mu_{i-1}, \mu_{i+1}, \dots, \mu_k]$ denote a matrix such that its columns are the vectors in S'. Let $W \in \mathbb{R}^{(k-1) \times (k-1)}$ denote a diagonal matrix such that for all $j < i$ we have $W(j, j) = \sqrt{|C_j|}$ and for all $j \geq i$ we have $W(j, j) = \sqrt{|C_{j+1}|}$. Let $Z = HW$. The orthogonal projection matrix onto the span of S' is defined as $(I - \Pi') = Z(Z^T Z)^{-1} Z^T$, and using Lemma 14 we get

$$||(I - \Pi')\mu_i||_2^2 = \mu_i^T Z (Z^T Z)^{-1} Z^T \mu_i$$
$$= \mu_i^T Z ((Z^T Z)^{-1} - I) Z^T \mu_i + \mu_i^T Z Z^T \mu_i$$

By Lemma 14 $(Z^T Z)^{-1}$ is spectrally close to I, therefore we have

$$\left| \mu_i^T Z \left((Z^T Z)^{-1} - I \right) Z^T \mu_i \right| \leq \frac{5\sqrt{\epsilon}}{\varphi} ||Z^T \mu_i||_2^2$$

Thus we get

$$||(I - \Pi')\mu_i||_2^2 \leq \left(\frac{5\sqrt{\epsilon}}{\varphi} + 1 \right) ||Z^T \mu_i||_2^2 \leq 2||Z^T \mu_i||_2^2$$

By Lemma 15 we have

$$||Z^T \mu_i||_2^2 = \mu_i^T Z Z^T \mu_i \leq \frac{8\sqrt{\epsilon}}{\varphi} \cdot ||\mu_i||_2^2$$

Therefore we get

$$||(I-\Pi)\mu_i||_2^2 \le ||(I-\Pi')\mu_i||_2^2 \le 2||Z^T\mu_i||_2^2 \le \frac{16\sqrt{\epsilon}}{\varphi}||\mu_i||_2^2 \tag{42}$$

Hence,

$$||\Pi\mu_i||_2^2 \ge \left(1-16\frac{\sqrt{\epsilon}}{\varphi}\right)||\mu_i||_2^2.$$

Proof of item (2)**:** Note that

$$\langle\mu_i,\mu_j\rangle = \langle (I-\Pi)\mu_i+\Pi\mu_i,(I-\Pi)\mu_j+\Pi\mu_j\rangle = \langle (I-\Pi)\mu_i,(I-\Pi)\mu_j\rangle + \langle \Pi\mu_i,\Pi\mu_j\rangle$$

Thus by triangle inequality we have

$$|\langle\Pi\mu_i,\Pi\mu_j\rangle| \le |\langle\mu_i,\mu_j\rangle| + |\langle (I-\Pi)\mu_i,(I-\Pi)\mu_j\rangle| \tag{43}$$

By Cauchy Schwarz we have

$$\begin{aligned} |\langle (I-\Pi)\mu_i,(I-\Pi)\mu_j\rangle| &\le ||(I-\Pi)\mu_i||_2||(I-\Pi)\mu_i||_2 \\ &\le \frac{16\sqrt{\epsilon}}{\varphi}\cdot||\mu_i||_2||\mu_j||_2 && \text{By (42)} \\ &\le \frac{32\sqrt{\epsilon}}{\varphi}\cdot\frac{1}{\sqrt{|C_i||C_j|}} && \text{By Lemma 7 for small enough } \frac{\epsilon}{\varphi^2} \end{aligned} \tag{44}$$

Also by Lemma 7 we have

$$|\langle\mu_i,\mu_j\rangle| \le \frac{8\sqrt{\epsilon}}{\varphi}\cdot\frac{1}{\sqrt{|C_i||C_j|}} \tag{45}$$

Therefore by (43), (44) and (45) we get

$$|\langle\Pi\mu_i,\Pi\mu_j\rangle| \le |\langle\mu_i,\mu_j\rangle| + |\langle (I-\Pi)\mu_i,(I-\Pi)\mu_j\rangle| \le \frac{40\sqrt{\epsilon}}{\varphi}\cdot\frac{1}{\sqrt{|C_i||C_j|}}.$$

□

4.4 Robustness property of (k,φ,ϵ)-clusterable graphs

In this subsection we show a Lemma that establishes a robustness property of (k,φ,ϵ)-clusterable graphs. That is we show that any collection $\{S_1,S_2,\ldots,S_k\}$ of pairwise disjoint subsets of vertices must match clusters $\{C_1,\ldots,C_k\}$ well.

Lemma 16. *Let $G=(V,E)$ be a d-regular graph that admits a (k,φ,ϵ)-clustering $C_1,\ldots,C_k$. Let $k\ge 2$, $\varphi\in(0,1)$ and $\frac{\epsilon}{\varphi^3}$ be smaller than an absolute positive constant. If $S_1,S_2,\ldots,S_k\subseteq V$ are k disjoint sets such that for all $i\in[k]$*

$$\phi(S_i) \le O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right)$$

then there exists a permutation π on k elements so that for all $i\in[k]$:

$$|C_{\pi(i)}\triangle S_i| \le O\left(\frac{\epsilon}{\varphi^3}\cdot\log(k)\right)|C_{\pi(i)}|$$

Proof. Fix $i\in[k]$ and let $J_i=\{j : |S_i\cap C_j|\le |C_j|/2\}$. Then observe that because the inner conductance of every C_i is at least φ we get:

$$\varphi\sum_{j\in J_i}|S_i\cap C_j| \le O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right)|S_i| \tag{46}$$

Using (46) and the assumption $\frac{\epsilon}{\varphi^3}$ is sufficiently small we get that

$$\sum_{j\in J_i}|S_i\cap C_j| \le O\left(\frac{\epsilon}{\varphi^3}\cdot\log(k)\right)|S_i| < |S_i| \tag{47}$$

(47) and $\sum_{j\in[k]} |S_i \cap C_j| = |S_i|$ gives us that

$$\text{For all } i \in [k], J_i \neq [k] \tag{48}$$

We will show that for each i: $|[k] \setminus J_i| = 1$ and that a function $i \mapsto \pi(i) \in [k] \setminus J_i$ (that is $\pi(i)$ is the only element of $[k] \setminus J_i$) is a permutation and that it satisfies the claim of the Lemma.

Assume that there exist $i_1 \neq i_2 \in [k]$ and $j \in ([k] \setminus J_{i_1}) \cap ([k] \setminus J_{i_2})$. By definition of J_i's we get that $|S_{i_1} \cap C_j|, |S_{i_2} \cap C_j| > |C_j|/2$ but S_i's are disjoint so it's impossible that two of them intersect more than half of the same C_j. That means that sets $([k] \setminus J_i)$ are pairwise disjoint for all i's. But we also know from (48) that for all i $([k] \setminus J_i) \neq \emptyset$. So we have k nonempty, pairwise disjoint subsets of $[k]$, which means that every set contains one element and all elements are different. That in turn means that we can define π as a function $i \mapsto \pi(i) \in [k] \setminus J_i$ and π is a permutation.

Now we show that π satisfies the claim of the Lemma. Observe that because for all $i \in [k]$ the set $[k] \setminus J_i$ contains only one element we get for all $i \in [k]$.

$$\sum_{j\in J_i} |S_i \cap C_j| = |S_i \setminus C_{\pi(i)}| \tag{49}$$

Note that because of (46) and (49) for all $i \in [k]$:

$$|S_i \setminus C_{\pi(i)}| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log k\right) |S_i|. \tag{50}$$

Moreover because inner conductance of every C_i is at least φ and $|C_{\pi(i)} \setminus S_i| < |C_{\pi(i)}|/2$ we get that for all $i \in [k]$

$$\varphi \cdot |C_{\pi(i)} \setminus S_i| \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right) |S_i| \tag{51}$$

Finally combining (50) and (51) we get that:

$$|C_{\pi(i)} \triangle S_i| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|$$

□

5 A spectral dot product oracle

Our goal in this section is to develop what we call a *spectral dot product oracle*. The oracle is a sublinear time and space data structure that has oracle access to a (k, φ, ϵ)-clusterable graph G and after a preprocessing step can answer dot products queries for the spectral embedding. Specifically, if $L = U\Lambda U^T$ is the normalized Laplacian of G and the x-th column of $F = U_{[k]}^T$ is called f_x for $x \in V$ then our oracle gets as input two vertices x, y and returns an approximation of $\langle f_x, f_y \rangle$. Both the preprocessing time and the time to evaluate an oracle query are $k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^{O(1)}$, that is, sublinear in n for $\epsilon \ll \varphi^2$. We now state the main theorem that we prove in this section. The algorithms mentioned in Theorem 2 can be found later in this section.

Theorem 2. *[Spectral Dot Product Oracle] Let $\epsilon, \varphi \in (0,1)$ with $\epsilon \leq \frac{\varphi^2}{10^5}$. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $1 > \xi > \frac{1}{n^5}$. Then* InitializeOracle$(G, 1/2, \xi)$ *(Algorithm 4) computes in time $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^3 \cdot \frac{1}{\varphi^2}/\xi^{18})$ a sublinear space data structure $\mathcal{D}$ of size $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^3/\xi^{12})$ such that with probability at least $1 - n^{-100}$ the following property is satisfied:*

For every pair of vertices $x, y \in V$, SpectralDotProduct$(G, x, y, 1/2, \xi, \mathcal{D})$ *(Algorithm 5) computes an output value $\langle f_x, f_y \rangle_{apx}$ such that with probability at least $1 - n^{-100}$*

$$\left| \langle f_x, f_y \rangle_{apx} - \langle f_x, f_y \rangle \right| \leq \frac{\xi}{n}.$$

The running time of SpectralDotProduct$(G, x, y, 1/2, \xi, \mathcal{D})$ *is $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$.*

Furthermore, for any $0 \leq \delta \leq 1/2$, one can obtain the following trade-offs between preprocessing time and query time: Algorithm SpectralDotProduct$(G, x, y, \delta, \xi, \mathcal{D})$ *requires $O(k^{O(1)} \cdot n^{\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$ per query when the prepressing time of Algorithm* InitializeOracle(G, δ, ξ) *is increased to $O(k^{O(1)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^3 \cdot \frac{1}{\varphi^2}/\xi^{18})$.*

5.1 The spectral dot product oracle - overview

In the following sections we provide the proof of the spectral dot product oracle. Recall from the technical overview that we are using the following algorithms (we restate them for convenience of the reader). Our main tool for accessing the spectral embedding of the graph is a primitive that runs a few short (logarithmic length) random walks from a given vertex.

Algorithm 1 RunRandomWalks(G, R, t, x)

```
Run R random walks of length t starting from x
Let m̂_x(y) be the fraction of random walks that ends at y      ▷ vector m̂_x has support at most R
return m̂_x
```

Another key primitive uses collision statistics to estimate the Gram matrix of random walk distributions started at vertices in a set S.

Algorithm 2 EstimateCollisionProbabilities(G, I_S, R, t)

```
for i = 1 to O(log n) do
    Q̂_i := EstimateTransitionMatrix(G, I_S, R, t)
    P̂_i := EstimateTransitionMatrix(G, I_S, R, t)
    𝒢_i := 1/2 (P̂_i^T Q̂_i + Q̂_i^T P̂_i)                          ▷ 𝒢_i is symmetric
Let 𝒢 be a matrix obtained by taking the entrywise median of 𝒢_i's   ▷ 𝒢 is symmetric
return 𝒢
```

We also need the following procedure.

Algorithm 3 EstimateTransitionMatrix(G, I_S, R, t)

for each sample $x \in I_S$ do
 $\widehat{m}_x :=$ RunRandomWalks(G, R, t, x)
Let $\widehat{Q}$ be the matrix whose columns are $\widehat{m}_x$ for $x \in I_S$
return $\widehat{Q}$ ▷ $\widehat{Q}$ has at most Rs non-zeros

Then we can initialize the dot product oracle.

Algorithm 4 InitializeOracle(G, δ, ξ) ▷ Need: $\epsilon/\varphi^2 \leq \frac{1}{10^5}$

$t := \frac{20 \cdot \log n}{\varphi^2}$
$R_{\text{init}} := O(n^{1-\delta+3\cdot 10^3 \cdot \epsilon/\varphi^2} \cdot k^{33}/\xi^6)$
$s := O(n^{1500\cdot\epsilon/\varphi^2} \cdot \log n \cdot k^{16}/\xi^6)$
Let I_S be the multiset of s indices chosen independently and uniformly at random from $\{1, \dots, n\}$
for $i = 1$ to $O(\log n)$ do
 $\widehat{Q}_i :=$ EstimateTransitionMatrix$(G, I_S, R_{\text{init}}, t)$ ▷ $\widehat{Q}_i$ has at most $R_{\text{init}} \cdot s$ non-zeros
$\mathcal{G} :=$EstimateCollisionProbabilities$(G, I_S, R_{\text{init}}, t)$
Let $\frac{n}{s} \cdot \mathcal{G} := \widehat{W}\widehat{\Sigma}\widehat{W}^T$ be the eigendecomposition of $\frac{n}{s} \cdot \mathcal{G}$ ▷ $\mathcal{G} \in \mathbb{R}^{s \times s}$
if $\widehat{\Sigma}^{-1}$ exists then
 $\Psi := \frac{n}{s} \cdot \widehat{W}_{[k]}\widehat{\Sigma}_{[k]}^{-2}\widehat{W}_{[k]}^T$ ▷ $\Psi \in \mathbb{R}^{s \times s}$
 return $\mathcal{D} := \{\Psi, \widehat{Q}_1, \dots, \widehat{Q}_{O(\log n)}\}$

Finally, we have the query algorithm.

Algorithm 5 SpectralDotProductOracle$(G, x, y, \delta, \xi, \mathcal{D})$ ▷ Need: $\epsilon/\varphi^2 \leq \frac{1}{10^5}$
▷ $\mathcal{D} := \{\Psi, \widehat{Q}_1, \dots, \widehat{Q}_{O(\log n)}\}$

$R_{\text{query}} := O(n^{\delta+500\cdot\epsilon/\varphi^2} \cdot k^9/\xi^2)$
for $i = 1$ to $O(\log n)$ do
 $\widehat{m}_x^i :=$ RunRandomWalks$(G, R_{\text{query}}, t, x)$
 $\widehat{m}_y^i :=$ RunRandomWalks$(G, R_{\text{query}}, t, y)$
Let α_x be a vector obtained by taking the entrywise median of $(\widehat{Q}_i)^T(\widehat{m}_x^i)$ over all runs
Let α_y be a vector obtained by taking the entrywise median of $(\widehat{Q}_i)^T(\widehat{m}_y^i)$ over all runs
return $\langle f_x, f_y \rangle_{apx} := \alpha_x^T \Psi \alpha_y$

Let $I_S = \{i_1, \dots, i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1, \dots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. As already explained in detail in the technical overview, we first prove stability bounds for the pseudoinverse. Then we show that that M^t is approximated by $M^t S$ and finally we show that algorithm RunRandomWalks approximates the $M^t \mathbb{1}_x$ sufficiently well. We conclude with the proof of Theorem 2.

5.2 Stability bounds for the low rank approximation

The main result of this section is a bound on the stability of the pseudoinverse of the rank-k approximation of two symmetric, positive semi-definite matrices $A, \widetilde{A} \in \mathbb{R}^{n \times n}$ that are spectrally close and that have an eigenvalue gap between the k-th and $(k+1)$-st eigenvalue. In order to prove this result, we use Weyl's inequality, which gives bounds on the eigenvalues of the sum of a matrix A and a perturbation matrix P. Recall that for a symmetric matrix A, we write $\nu_i(A)$ (resp. $\nu_{\max}(A), \nu_{\min}(A)$) to denote the i^{th} largest (resp. maximum, minimum) eigenvalue of A.

Lemma 17 (Weyl's Inequality)**.** *Let $A, P \in \mathbb{R}^{n \times n}$ be two symmetric matrices. Then we have for all $i \in \{1, \dots, n\}$:*

$$\nu_i(A) + \nu_{\min}(P) \leq \nu_i(A + P) \leq \nu_i(A) + \nu_{\max}(P),$$

where for a symmetric matrix $H \in \mathbb{R}^{n \times n}$ $\nu_i(H)$ denotes its i^{th} largest eigenvalue and $\nu_{\min}(H)$ and $\nu_{\max}(H)$ refer to the smallest and largest eigenvalues of H.

We will use the Davis-Kahan $\sin(\theta)$ Theorem [DK70] (the version given in the note [DK]).

Theorem 5 (Davis-Kahan $sin(\theta)$-Theorem [DK70])**.** . *Let $H = E_0 A_0 E_0^T + E_1 A_1 E_1^T$ and $\widetilde{H} = F_0 \Lambda_0 F_0^T + F_1 \Lambda_1 F_1^T$ be symmetric real-valued matrices with E_0, E_1 and F_0, F_1 orthogonal. If the eigenvalues of A_0 are contained in an interval (a, b), and the eigenvalues of Λ_1 are excluded from the interval $(a-\eta, b+\eta)$for some $\eta > 0$, then for any unitarily invariant norm $\|.\|$*

$$\|F_1^T E_0\| \leq \frac{\|F_1^T(\widetilde{H} - H)E_0\|}{\eta}.$$

Let $m \leq n$ be integers. For any matrix $A \in \mathbb{R}^{n\times m}$ with singular value decomposition (SVD) $A = Y\Gamma Z^T$ we assume $Y \in \mathbb{R}^{n\times n}$, $\Gamma \in \mathbb{R}^{n\times n}$ is a diagonal matrix of singular values and $Z \in \mathbb{R}^{m\times n}$ (this is a slightly non-standard definition of the SVD, but having Γ be a square matrix will be convenient). Y has orthonormal columns, the first m columns of Z are orthonormal, and the rest of the columns of Z are zero. For any integer $q \in [m]$ we denote $Y_{[q]} \in \mathbb{R}^{n\times q}$ as the first q columns of Y and $Y_{-[q]}$ to denote the matrix of the remaining columns of Y. We also denote by $Z_{[q]} \in \mathbb{R}^{m\times q}$ as the first q columns of Z and $Z_{-[q]}$ to denote the matrix of the remaining $n-q$ columns of Z. Finally we denote by $\Gamma_{[q]} \in \mathbb{R}^{q\times q}$ the submatrix of Γ corresponding to the first q rows and columns of Γ and we use $\Gamma_{-[q]}$ to denote the submatrix corresponding to the last $n-q$ rows and $n-q$ columns of Γ. So for any $q \in [m]$ the span of $Y_{-[q]}$ is the orthogonal complement of the span of $Y_{[q]}$ in $\mathbb{R}^n$, also the span of the columns of $Z_{-[q]}$ is the orthogonal complement of the span of $Z_{[q]}$ in $\mathbb{R}^m$. Thus we can write $A = Y_{[q]}\Gamma_{[q]}Z_{[q]}^T + Y_{-[q]}\Gamma_{-[q]}Z_{-[q]}^T$.

Claim 1. *For every symmetric matrix E and every pair of orthogonal projection matrices $P, \widetilde{P}$ one has*

$$||P \cdot E \cdot P - \widetilde{P} \cdot E \cdot \widetilde{P}||_2 \leq 2\|E\|_2 \cdot (\|P \cdot (I - \widetilde{P})\|_2 + \|\widetilde{P} \cdot (I - P)\|_2).$$

Proof. Since $\widetilde{P} + (I - \widetilde{P}) = I$ we can write

$$\begin{aligned} P \cdot E \cdot P &= (\widetilde{P} + (I - \widetilde{P}))P \cdot E \cdot P \cdot (\widetilde{P} + (I - \widetilde{P})) \\ &= P \cdot E \cdot P \cdot (I - \widetilde{P}) + \widetilde{P} \cdot P \cdot E \cdot P \cdot \widetilde{P} + (I - \widetilde{P}) \cdot P \cdot E \cdot P \cdot \widetilde{P} \end{aligned} \tag{52}$$

Since $P + (I - P) = I$ we have

$$\begin{aligned} \widetilde{P} \cdot E \cdot \widetilde{P} &= \widetilde{P}(P + (I - P)) \cdot E \cdot (P + (I - P))\, \widetilde{P}||_2 \\ &= \widetilde{P} \cdot E \cdot (I - P)\widetilde{P} + \widetilde{P} \cdot P \cdot E \cdot P \cdot \widetilde{P} + \widetilde{P} \cdot (I - P) \cdot E \cdot P \cdot \widetilde{P} \end{aligned} \tag{53}$$

Putting (52) and (53) together and by triangle inequality we get

$$\begin{aligned} &||P \cdot E \cdot P - \widetilde{P} \cdot E \cdot \widetilde{P}||_2 \\ &\leq \|P \cdot E \cdot P \cdot (I - \widetilde{P})\|_2 + \|(I - \widetilde{P}) \cdot P \cdot E \cdot P \cdot \widetilde{P}\|_2 + \|\widetilde{P} \cdot E \cdot (I - P)\widetilde{P}\|_2 + \|\widetilde{P} \cdot (I - P) \cdot E \cdot P \cdot \widetilde{P}\|_2 \end{aligned}$$

Thus by submultiplicativity of the operator norm we get

$$\begin{aligned} &||P \cdot E \cdot P - \widetilde{P} \cdot E \cdot \widetilde{P}||_2 \\ &\leq \|P\|_2\|E\|_2\|P \cdot (I - \widetilde{P})\|_2 + \|(I - \widetilde{P}) \cdot P\|_2\|E\|_2\|P\|_2\|\widetilde{P}\|_2 + \|\widetilde{P}\|_2\|E\|_2\|(I - P)\widetilde{P}\|_2 + \|\widetilde{P}\|_2\|E\|_2\|(I - P)\widetilde{P}\|_2 \\ &\leq \|E\|_2 \left(\|P \cdot (I - \widetilde{P})\|_2 + \|(I - \widetilde{P}) \cdot P\|_2 + \|(I - P)\widetilde{P}\|_2 + \|\widetilde{P}(I - P)\|_2\right) \text{ Since } ||P|| = ||\widetilde{P}||_2 = 1 \\ &= 2 \cdot \|E\|_2 \cdot (\|P \cdot (I - \widetilde{P})\|_2 + \|\widetilde{P} \cdot (I - P)\|_2), \end{aligned}$$

where the last equality holds since $\|P \cdot (I - \widetilde{P})\|_2 = \|(I - \widetilde{P})^T \cdot P^T\|_2 = \|(I - \widetilde{P}) \cdot P\|_2$ and similarly since $\|\widetilde{P} \cdot (I - P)\|_2 = \|(I - P)^T \cdot \widetilde{P}^T\|_2 = \|(I - P) \cdot \widetilde{P}\|_2$. □

Recall that for matrices $A, \widetilde{A} \in \mathbb{R}^{n\times n}$, we write $A \preccurlyeq \widetilde{A}$, if $\forall x \in \mathbb{R}^n$ we have $x^T Ax \leq x^T \widetilde{A}x$ and we write $A \prec \widetilde{A}$, if $\forall x \in \mathbb{R}^n$ we have $x^T Ax < x^T \widetilde{A}x$. Now we can state the main technical result of this section (Lemma 18), whose proof relies on matrix perturbation bounds Davis-Kahan $\sin\theta$ theorem (Theorem 5).

Lemma 18. *Let $A, \widetilde{A} \in \mathbb{R}^{n\times n}$ be symmetric matrices with eigendecompositions $A = Y\Gamma Y^T$ and $\widetilde{A} = \widetilde{Y}\widetilde{\Gamma}\widetilde{Y}^T$. Let the eigenvalues of A be $1 \geq \gamma_1 \geq \cdots \geq \gamma_n \geq 0$. Suppose that $\|A - \widetilde{A}\|_2 \leq \frac{\gamma_k}{100}$ and $\gamma_{k+1} < \gamma_k/4$. Then we have*

$$\|Y_{[k]}\Gamma_{[k]}^{-1}Y_{[k]}^T - \widetilde{Y}_{[k]}\widetilde{\Gamma}_{[k]}^{-1}\widetilde{Y}_{[k]}^T\|_2 \leq \frac{32\left(\|A-\widetilde{A}\|_2\right)^{1/3}}{\gamma_k^2}.$$

Proof. We assume without loss of generality that $\|A - \widetilde{A}\|_2 > 0$, as otherwise the claim of the lemma follows since $A = \widetilde{A}$.

Let $\eta \geq 4\|A - \widetilde{A}\|_2$ and $\delta \in (2\|A - \widetilde{A}\|_2/\eta, 1/2]$. Then we have $\widetilde{A} + \eta \cdot I$ is invertible since

$$\widetilde{A} + \eta \cdot I \succeq A + (\eta - \|A - \widetilde{A}\|_2) \cdot I \succeq A + (\eta/2) \cdot I \succ 0.$$

Note that

$$\begin{aligned}
(1+\delta)(\widetilde{A} + \eta \cdot I) - A - \eta \cdot I &= (\widetilde{A} - A + (\delta/2)\eta \cdot I) + \delta\widetilde{A} + (\delta/2)\eta \cdot I \\
&\succ (\widetilde{A} - A + \|A - \widetilde{A}\|_2 \cdot I) + \delta\widetilde{A} + (\delta/2)\eta \cdot I \quad && \text{Since } \delta > 2\|A - \widetilde{A}\|_2/\eta \\
&\succ \delta\widetilde{A} + (\delta/2)\eta \cdot I \\
&\succ -\delta\|A - \widetilde{A}\|_2 \cdot I + \|A - \widetilde{A}\|_2 \cdot I && \text{Since } \delta > 2\|A - \widetilde{A}\|_2/\eta \\
&\succ 0. && \text{Since } \delta < 1 \qquad (54)
\end{aligned}$$

Also note that

$$\begin{aligned}
A + \eta \cdot I - ((1-\delta)(\widetilde{A} + \eta \cdot I)) &= (A - \widetilde{A} + (\delta/2)\eta \cdot I) + (\delta\widetilde{A} + (\delta/2)\eta \cdot I) \\
&\succ (A - \widetilde{A} + \|A - \widetilde{A}\|_2 \cdot I) + (\delta\widetilde{A} + (\delta/2)\eta \cdot I) \quad && \text{Since } \delta > 2\|A - \widetilde{A}\|_2/\eta \\
&\succ \delta\widetilde{A} + (\delta/2)\eta \cdot I \\
&\succ -\delta\|A - \widetilde{A}\|_2 \cdot I + \|A - \widetilde{A}\|_2 \cdot I && \text{Since } \delta > 2\|A - \widetilde{A}\|_2/\eta \\
&\succ 0 && \text{Since } \delta < 1 \qquad (55)
\end{aligned}$$

Therefore, by (54) and (55) for every $\eta \geq 4\|A - \widetilde{A}\|_2$ and $\delta \in (2\|A - \widetilde{A}\|_2/\eta, 1/2]$ we have

$$(1-\delta)(\widetilde{A} + \eta \cdot I) \prec A + \eta \cdot I \prec (1+\delta)(\widetilde{A} + \eta \cdot I), \tag{56}$$

Note that since $(\widetilde{A} + \eta \cdot I)$ and $(A + \eta \cdot I)$ are symmetric, invertible and positive definite matrices, thus by Lemma 18 we get from (56) that $\frac{1}{1-\delta}(\widetilde{A} + \eta \cdot I)^{-1} \succ (A + \eta \cdot I)^{-1}$ and $(A + \eta \cdot I)^{-1} \succ \frac{1}{1+\delta}(\widetilde{A} + \eta \cdot I)^{-1}$, which implies that

$$(1-\delta)(A + \eta \cdot I)^{-1} \prec (\widetilde{A} + \eta \cdot I)^{-1} \prec (1+\delta)(A + \eta \cdot I)^{-1}. \tag{57}$$

At the same time note that if $P = Y_{[k]}Y_{[k]}^T$ and $\widetilde{P} = \widetilde{Y}_{[k]}\widetilde{Y}_{[k]}^T$, we have, since $\eta \geq 4\|A - \widetilde{A}\|_2$,

$$Y_{[k]}(\Gamma_{[k]} + \eta I_k)^{-1}Y_{[k]}^T = P \cdot (A + \eta \cdot I)^{-1}P \tag{58}$$

and

$$\widetilde{Y}_{[k]}(\widetilde{\Gamma}_{[k]} + \eta I_k)^{-1}\widetilde{Y}_{[k]}^T = \widetilde{P} \cdot (\widetilde{A} + \eta \cdot I)^{-1}\widetilde{P}, \tag{59}$$

where I_k in $(\widetilde{\Gamma}_{[k]} + \eta I_k)^{-1}$ stands for the $k \times k$ identity matrix. Indeed, the first equation above follows by noting that

$$\begin{aligned}
P(A + \eta \cdot I)^{-1}P &= PY(\Gamma + \eta \cdot I)^{-1}Y^TP \\
&= PY_{[k]}(\Gamma_{[k]} + \eta \cdot I_k)^{-1}Y_{[k]}^TP + PY_{-[k]}(\Gamma_{-[k]} + \eta \cdot I_{n-k})^{-1}Y_{-[k]}^TP \\
&= PY_{[k]}(\Gamma_{[k]} + \eta \cdot I_k)^{-1}Y_{[k]}^TP \qquad \text{Since } PY_{-[k]} = Y_{[k]}Y_{[k]}^TY_{-[k]} = 0 \\
&= Y_{[k]}Y_{[k]}^TY_{[k]}(\Gamma_{[k]} + \eta \cdot I_k)^{-1}Y_{[k]}^TY_{[k]}Y_{[k]}^T \\
&= Y_{[k]}(\Gamma_{[k]} + \eta \cdot I_k)^{-1}Y_{[k]}^T,
\end{aligned}$$

where we write I_k and I_{n-k} above to denote identity matrices of dimension k and $n-k$ respectively (we omit the subscript on I in what follows to simplify notation). The argument for $\widetilde{A}$ is analogous. Using (58) and (59)

$$\begin{aligned}\|Y_{[k]}\Gamma_{[k]}^{-1}Y_{[k]}^T - \widetilde{Y}_{[k]}\widetilde{\Gamma}_{[k]}^{-1}\widetilde{Y}_{[k]}^T\|_2 &\le \|Y_{[k]}(\Gamma_{[k]}^{-1} - (\Gamma_{[k]} + \eta\cdot I_k)^{-1})Y_{[k]}^T\|_2 + \|\widetilde{Y}_{[k]}(\widetilde{\Gamma}_{[k]}^{-1} - (\widetilde{\Gamma}_{[k]} + \eta\cdot I_k)^{-1})\widetilde{Y}_{[k]}^T\|_2\\ &+ \|P\cdot(A+\eta\cdot I)^{-1}P - \widetilde{P}\cdot(\widetilde{A}+\eta\cdot I)^{-1}\widetilde{P}\|_2\end{aligned} \tag{60}$$

The rest of the proof proceeds in three steps. We first prove an upper bound on the first two terms on the rhs of (60), and then on the third term. Finally we put everything together and choose $\eta > 0$ and $\delta > 0$ appropriately to achieve the final bound.

Step 1. We now prove an upper bound on the first two terms in (60). One has

$$\begin{aligned}\|Y_{[k]}(\Gamma_{[k]}^{-1} - (\Gamma_{[k]} + \eta\cdot I_k)^{-1})Y_{[k]}^T\|_2 &\le \max_{\xi\ge\gamma_k}\left(\frac{1}{\xi} - \frac{1}{\xi+\eta}\right)\\ &= \max_{\xi\ge\gamma_k}\frac{\eta}{\xi(\xi+\eta)}\\ &\le \frac{\eta}{\gamma_k^2}\end{aligned}$$

and similarly, since $\nu_k(\widetilde{A}) \ge \nu_k(A) - \|A-\widetilde{A}\|_2$ by Weyl's inequality (Lemma 17),

$$\begin{aligned}\|\widetilde{Y}_{[k]}(\widetilde{\Gamma}_{[k]}^{-1} - (\widetilde{\Gamma}_{[k]} + \eta\cdot I_k)^{-1})\widetilde{Y}_{[k]}^T\|_2 &\le \max_{\xi\ge\gamma_k-\|A-\widetilde{A}\|_2}\left(\frac{1}{\xi} - \frac{1}{\xi+\eta}\right)\\ &= \max_{\xi\ge\gamma_k-\|A-\widetilde{A}\|_2}\frac{\eta}{\xi(\xi+\eta)}\\ &\le \frac{4\eta}{\gamma_k^2} \qquad \text{Since } \|A-\widetilde{A}\|_2 \le \gamma_k/2 \text{ by assumption}\end{aligned}$$

Combining the two bounds, we get

$$\|Y_{[k]}(\Gamma_{[k]}^{-1} - (\Gamma_{[k]} + \eta\cdot I_k)^{-1})Y_{[k]}^T\|_2 + \|\widetilde{Y}_{[k]}(\widetilde{\Gamma}_{[k]}^{-1} - (\widetilde{\Gamma}_{[k]} + \eta\cdot I_k)^{-1})\widetilde{Y}_{[k]}^T\|_2 \le \frac{5\eta}{\gamma_k^2}. \tag{61}$$

Step 2. One has

$$\begin{aligned}\|P\cdot(A+\eta\cdot I)^{-1}P - \widetilde{P}\cdot(\widetilde{A}+\eta\cdot I)^{-1}\widetilde{P}\|_2 &\le \|P\cdot(A+\eta\cdot I)^{-1}P - \widetilde{P}\cdot(A+\eta\cdot I)^{-1}\widetilde{P}\|_2\\ &+ \|\widetilde{P}\cdot(A+\eta\cdot I)^{-1}\widetilde{P} - \widetilde{P}\cdot(\widetilde{A}+\eta\cdot I)^{-1}\widetilde{P}\|_2.\end{aligned} \tag{62}$$

For the second term, we have by (57)

$$\begin{aligned}\|\widetilde{P}\cdot(A+\eta\cdot I)^{-1}\widetilde{P} - \widetilde{P}\cdot(\widetilde{A}+\eta\cdot I)^{-1}\widetilde{P}\|_2 &= \|\widetilde{P}((A+\eta\cdot I)^{-1} - (\widetilde{A}+\eta\cdot I)^{-1})\widetilde{P}\|_2\\ &\le \delta\|(A+\eta\cdot I)^{-1}\|_2\\ &\le \frac{\delta}{\eta}. \qquad \text{Since } A\succeq 0\end{aligned} \tag{63}$$

Here the transition from the first line to the second uses the fact that by (57) one has

$$\begin{aligned}(\widetilde{A}+\eta\cdot I)^{-1} - (A+\eta\cdot I)^{-1} &= (\widetilde{A}+\eta\cdot I)^{-1} - (1+\delta)(A+\eta\cdot I)^{-1} + \delta(A+\eta\cdot I)^{-1}\\ &\preceq \delta(A+\eta\cdot I)^{-1}\end{aligned}$$

and

$$\begin{aligned}(A+\eta\cdot I)^{-1} - (\widetilde{A}+\eta\cdot I)^{-1} &= (1-\delta)(A+\eta\cdot I)^{-1} - (\widetilde{A}+\eta\cdot I)^{-1} + \delta(A+\eta\cdot I)^{-1}\\ &\preceq \delta(A+\eta\cdot I)^{-1}.\end{aligned}$$

We now bound the first term on the rhs of (62). Using Claim 1 we get for every $\eta > 0$

$$\begin{aligned}\|P \cdot (A + \eta \cdot I)^{-1} P - \widetilde{P} \cdot (A + \eta \cdot I)^{-1} \widetilde{P}\|_2 &\le 2\|(A + \eta \cdot I)^{-1}\|_2 \cdot (\|P \cdot (I - \widetilde{P})\|_2 + \|\widetilde{P} \cdot (I - P)\|_2) \\ &\le \frac{2}{\eta} \cdot (\|P \cdot (I - \widetilde{P})\|_2 + \|\widetilde{P} \cdot (I - P)\|_2).\end{aligned}$$

In order to bound $\|P \cdot (I - \widetilde{P})\|_2$ and $\|\widetilde{P} \cdot (I - P)\|_2$, we first note that by Weyl's inequality

$$\nu_{k+1}(\widetilde{A}) \le \nu_{k+1}(A) + ||A - \widetilde{A}||_2 \le \gamma_k/4 + \gamma_k/100 < (3/4)\gamma_k$$

and $\nu_k(A) = \gamma_k$ by assumption of the lemma. Hence we can apply Theorem 5 by choice of $H = A$, $E_0 = Y_{[k]}$, $E_1 = Y_{-[k]}$, $A_0 = \Gamma_{[k]}$, $A_1 = \Gamma_{-[k]}$, and $\widetilde{H} = \widetilde{A}$, $F_0 = \widetilde{Y}_{[k]}$, $F_1 = \widetilde{Y}_{-[k]}$, $\Lambda_0 = \widetilde{\Gamma}_{[k]}$, $\Lambda_1 = \widetilde{\Gamma}_{-[k]}$. Let $\eta = \frac{\gamma_k}{4}$. Note that the eigenvalues of $A_0 = \Gamma_{[k]}$ are at least γ_k and the eigenvalues of $\Lambda_1 = \widetilde{\Gamma}_{-[k]}$ are at most $(3/4)\gamma_k = \gamma_k - \eta$. Therefore, by Theorem 5 we have

$$\|\widetilde{Y}_{-[k]}^T Y_{[k]}\|_2 = \|F_1^T E_0\|_2 \le \frac{\|F_1^T(\widetilde{A} - A)E_0\|_2}{\eta} \le \frac{\|A - \widetilde{A}\|_2}{\gamma_k/4}.$$

Thus we have $\|Y_{[k]}^T \widetilde{Y}_{-[k]}\|_2 \le \frac{\|A-\widetilde{A}\|_2}{\gamma_k/4}$. Similarly, we have

$$\nu_{k+1}(A) \le \gamma_k/4$$

and $\nu_k(\widetilde{A}) \ge \nu_k(A) - \|A - \widetilde{A}\|_2 \ge \gamma_k - \gamma_k/100$. Hence we can apply Theorem 5 by choice of $H = A$, $E_0 = Y_{-[k]}$, $E_1 = Y_{[k]}$, $A_0 = \Gamma_{-[k]}$, $A_1 = \Gamma_{[k]}$, and $\widetilde{H} = \widetilde{A}$, $F_0 = \widetilde{Y}_{-[k]}$, $F_1 = \widetilde{Y}_{[k]}$, $\Lambda_0 = \widetilde{\Gamma}_{-[k]}$, $\Lambda_1 = \widetilde{\Gamma}_{[k]}$. Let $\eta = \frac{\gamma_k}{4}$. Note that the eigenvalues of $A_0 = \Gamma_{-[k]}$ are at most γ_{k+1} and the eigenvalues of $\Lambda_1 = \widetilde{\Gamma}_{[k]}$ are at least $\gamma_k - \gamma_k/100 \ge \gamma_k - \eta$. Therefore, by Theorem 5 we have

$$\|\widetilde{Y}_{[k]}^T Y_{-[k]}\|_2 = \|F_1^T E_0\| \le \frac{\|F_1^T(\widetilde{A} - A)E_0\|}{\eta} \le \frac{\|A - \widetilde{A}\|}{\gamma_k/4}.$$

Thus, we have $\|\widetilde{Y}_{[k]}^T Y_{-[k]}\|_2 \le \frac{\|A-\widetilde{A}\|_2}{\gamma_k/4}$. Putting these two bounds together, we get

$$\|P(I - \widetilde{P})\|_2 = \|Y_{[k]} Y_{[k]}^T \widetilde{Y}_{-[k]} \widetilde{Y}_{-[k]}^T\|_2 = \|Y_{[k]}^T \widetilde{Y}_{-[k]}\|_2 \le \frac{\|A - \widetilde{A}\|_2}{\gamma_k/4},$$

and similarly

$$\|\widetilde{P}(I - P)\|_2 \le \frac{\|A - \widetilde{A}\|_2}{\gamma_k/4}.$$

Substituting the above bounds together with (63) into (62), we get

$$\|P \cdot (A + \eta \cdot I)^{-1} P - \widetilde{P} \cdot (\widetilde{A} + \eta \cdot I)^{-1} \widetilde{P}\|_2 \le \frac{\delta}{\eta} + \frac{2}{\eta} \cdot \frac{8\|A - \widetilde{A}\|_2}{\gamma_k}. \tag{64}$$

Putting it together. Substituting (61) and (64) into (60) and setting $\eta = \left(\|A - \widetilde{A}\|_2\right)^{1/3}$ and $\delta = 4 \cdot \left(\|A - \widetilde{A}\|_2\right)^{2/3}$ (which satisfies $\eta \ge 4\|A - \widetilde{A}\|_2$ and $\delta \in (2\|A - \widetilde{A}\|_2/\eta, 1/2]$ as required since $\|A - \widetilde{A}\|_2 \le \frac{1}{100}$), we get

$$\begin{aligned}\|Y_{[k]} \Gamma_{[k]}^{-1} Y_{[k]}^T - \widetilde{Y}_{[k]} \widetilde{\Gamma}_{[k]}^{-1} \widetilde{Y}_{[k]}^T\|_2 &\le \frac{5\eta}{\gamma_k^2} + \frac{\delta}{\eta} + \frac{2}{\eta} \cdot \frac{8\|A - \widetilde{A}\|_2}{\gamma_k} \\ &\le \frac{5\left(\|A - \widetilde{A}\|_2\right)^{1/3}}{\gamma_k^2} + 4\left(\|A - \widetilde{A}\|_2\right)^{1/3} + \frac{16\left(\|A - \widetilde{A}\|_2\right)^{2/3}}{\gamma_k} \qquad (65) \\ &\le \frac{32\left(\|A - \widetilde{A}\|_2\right)^{1/3}}{\gamma_k^2},\end{aligned}$$

as required. In the last transition we used the assumption that $0 \le \gamma_k \le \gamma_1 \le 1$. □

5.3 Stability bounds under sampling of vertices

The main result of this section is Lemma 19, in which we give bounds for the stability of the pseudoinverse of the rank-k-approximation when we are sampling columns of the k-step random walk matrix of a (k, φ, ϵ)-clusterable graph.

Lemma 19. *Let $k \geq 2$ be an integer, $\varphi \in (0, 1)$ and $\epsilon \in (0, 1)$. Let $G = (V, E)$ be a d-regular and (k, φ, ϵ)-clusterable graph. Let M be the random walk transition matrix of G. Let $1/n^5 < \xi < 1$, $t \geq \frac{20 \log n}{\varphi^2}$. Let $c > 1$ be a large enough constant and let $s \geq c \cdot n^{(1280 \cdot \epsilon/\varphi^2)} \cdot \log n \cdot k^{16}/\xi^6$. Let $I_S = \{i_1, \dots, i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1, \dots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let $M^t = U \Sigma^t U^T$ be an eigendecomposition of M^t. Let $\sqrt{\frac{n}{s}} \cdot M^t S = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ be an SVD of $\sqrt{\frac{n}{s}} \cdot M^t S$ where $\widetilde{U} \in \mathbb{R}^{n \times n}, \widetilde{\Sigma} \in \mathbb{R}^{n \times n}, \widetilde{W} \in \mathbb{R}^{s \times n}$. If $\frac{\epsilon}{\varphi^2} \leq \frac{1}{10^5}$ then with probability at least $1 - n^{-100}$ matrix $\widetilde{\Sigma}_{[k]}^{-4}$ exists and we have*

$$\left| \mathbb{1}_x^T U_{[k]} U_{[k]}^T \mathbb{1}_y - (M^t \mathbb{1}_x)^T (M^t S) \left(\frac{n}{s} \cdot \widetilde{W}_{[k]} \widetilde{\Sigma}_{[k]}^{-4} \widetilde{W}_{[k]}^T \right) (M^t S)^T (M^t \mathbb{1}_y) \right| \leq \frac{\xi}{n}.$$

To prove Lemma 19 we require the following matrix concentration bound, which is a generalization of Bernstein's inequality to matrices.

Lemma 20 (Matrix Bernstein [Tro12])**.** *Consider a finite sequence X_i of independent, random matrices with dimensions $d_1 \times d_2$. Assume that each random matrix satisfies $\mathbb{E}[X_i] = 0$ and $\|X_i\|_2 \leq b$ almost surely. Define $\sigma^2 = \max\{\| \sum_i \mathbb{E}[X_i X_i^T]\|_2, \| \sum_i \mathbb{E}[X_i^T X_i]\|_2\}$. Then for all $t \geq 0$,*

$$\mathbb{P}\left[\| \sum_i X_i \|_2 \geq t \right] \leq (d_1 + d_2) \cdot \exp\left(\frac{-t^2/2}{\sigma^2 + bt/3} \right).$$

Equiped with the Matrix Bernstein bound, we can show that under certain spectral conditions we can approximate a matrix AA^T by $(AS)(AS)^T$, i.e. by sampling rows of M. The idea is to write $AA^T = \sum_{i=1}^n (A\mathbb{1}_i)(A\mathbb{1}_i)^T$ as a sum over the outer products of its columns and make the sample size depend on the spectral norm of the summands.

Lemma 21. *Let $A \in \mathbb{R}^{n \times n}$ be a matrix. Let $B = \max_{\ell \in \{1,\dots,n\}} \|(A\mathbb{1}_\ell)(A\mathbb{1}_\ell)^T\|_2$. Let $1 > \xi > 0$. Let $s \geq \frac{40 n^2 B^2 \log n}{\xi^2}$. Let $I_S = \{i_1, \dots, i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1, \dots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Then we have*

$$\mathbb{P}\left[\left\| AA^T - \frac{n}{s}(AS)(AS)^T \right\|_2 \geq \xi \right] \leq n^{-100}.$$

Proof. Observe that

$$AA^T = \sum_{\ell \in \{1,\dots,n\}} (A\mathbb{1}_\ell)(A\mathbb{1}_\ell)^T. \tag{66}$$

and

$$\frac{n}{s}(AS)(AS)^T = \frac{n}{s} \cdot \sum_{i_j \in I_S} (A\mathbb{1}_{i_j})(A\mathbb{1}_{i_j})^T. \tag{67}$$

For every $j = 1, 2, \dots, s$ let $X_j = \frac{n}{s} \cdot (A\mathbb{1}_{i_j})(A\mathbb{1}_{i_j})^T$. Thus we have

$$\mathbb{E}[X_j] = \frac{n}{s} \cdot \mathbb{E}[(A\mathbb{1}_{i_j})(A\mathbb{1}_{i_j})^T] = \frac{n}{s} \cdot \frac{1}{n} \sum_{\ell \in \{1,\dots,n\}} (A\mathbb{1}_\ell)(A\mathbb{1}_\ell)^T = \frac{1}{s} \cdot AA^T \tag{68}$$

By equality (67) we have $\frac{n}{s}(AS)(AS)^T = \sum_{j=1}^s X_j$. Thus by equality (68) we get

$$\left\| \frac{n}{s}(AS)(AS)^T - AA^T \right\|_2 = \left\| \sum_{j=1}^s (X_j - \mathbb{E}[X_j]) \right\|_2. \tag{69}$$

Let $Z_j = X_j - \mathbb{E}[X_j]$. We then have $\|Z_j\|_2 = \|X_j - \mathbb{E}[X_j]\|_2 \leq \|X_j\|_2 + \|\mathbb{E}[X_j]\|_2$ Now let $B = \max_{\ell \in \{1,\dots,n\}} \|(A\mathbb{1}_\ell)(A\mathbb{1}_\ell)^T\|_2$. Furthermore, by our assumption we have

$$\|X_j\|_2 = \left\| \frac{n}{s} \cdot (A\mathbb{1}_j)(A\mathbb{1}_j)^T \right\|_2 \leq \frac{n}{s} \cdot B \tag{70}$$

By subadditivity of the spectral norm and (68) we get

$$\|\mathbb{E}[X_j]\|_2 \leq \frac{n}{s} \cdot B \tag{71}$$

Putting (70) and (71) together we get

$$\|Z_j\|_2 = \|X_j - \mathbb{E}[X_j]\|_2 \leq \|X_j\|_2 + \|\mathbb{E}[X_j]\|_2 \leq 2 \cdot \frac{n}{s} \cdot B \tag{72}$$

We now bound for the variance. Since Z_j is symmetric, we have $Z_j^T Z_j = Z_j Z_j^T = Z_j^2$.

$$\left\|\sum_{j=1}^{s} \mathbb{E}[Z_j^2]\right\|_2 = s \cdot \|\mathbb{E}[Z_j^2]\|_2 = s \cdot \|\mathbb{E}[X_j^2] - \mathbb{E}[X_j]^2\|_2 \leq s \cdot \|\mathbb{E}[X_j^2]\|_2 + s \cdot \|\mathbb{E}[X_j]^2\|_2$$

By submultiplicativity of the spectral norm we get

$$\|\mathbb{E}[X_j^2]\|_2 = \left\|\frac{1}{n} \cdot \frac{n^2}{s^2} \sum_{\ell \in \{1,\ldots,n\}} ((A\mathbb{1}_\ell)(A\mathbb{1}_\ell)^T)^2\right\|_2 \leq \frac{n^2}{s^2} \cdot B^2 \tag{73}$$

Moreover by submultiplicativity of spectral norm we have $\|\mathbb{E}[X_j]^2\|_2 \leq \|\mathbb{E}[X_j]\|_2^2 \leq \frac{n^2}{s^2} \cdot B^2$. Putting things together we obtain

$$\|\sum_{j=1}^{s} \mathbb{E}[Z_j^2]\|_2 \leq \frac{2n^2 B^2}{s}$$

Now we can apply Lemma 20 and we get with $b = 2\frac{n}{s}B$ and $\sigma^2 \leq \frac{2n^2B^2}{s}$ using $s \geq \frac{40n^2B^2 \log n}{\xi^2}$

$$\mathbb{P}\left[\|\sum_{j=1}^{s} Z_j\|_2 > \xi\right] \leq 2n \cdot \exp\left(\frac{\frac{-\xi^2}{2}}{\sigma^2 + \frac{b\xi}{3}}\right) \leq n^{-100} \tag{74}$$

□

The following lemma upper bounds the collision probability from **every** vertex in a (k, φ, ϵ)-clusterable graph using our ℓ_∞ norm bounds on the bottom k eigenvectors of the Laplacian of such graphs[6]:

Lemma 22. *Let $k \geq 2$ be an integer, $\varphi \in (0,1)$ and $\epsilon \in (0,1)$. Let $G = (V, E)$ be a d-regular and that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let M be the random walk transition matrix of G. For any $t \geq \frac{20 \log n}{\varphi^2}$ and any $x \in V$ we have*

$$\|M^t \mathbb{1}_x\|_2 \leq O(k \cdot n^{-1/2+(20\epsilon/\varphi^2)}).$$

Proof. Let L be the normalized Laplacian of G. Recall that $(u_1, \ldots, u_n)$ are an orthonormal basis of eigenvectors of L with corresponding eigenvalues $0 = \lambda_1 \leq \ldots \leq \lambda_n$. Observe that each u_i is also an eigenvector of M, with eigenvalue $1 - \frac{\lambda_i}{2}$. We write $\mathbb{1}_x$ in the eigenbasis of L as $\mathbb{1}_x = \sum_{j=1}^{n} \beta_j u_j$ and note that the β_j correspond to the row of x in the matrix U. We have

$$M^t \mathbb{1}_x = M^t \left(\sum_{j=1}^{n} \beta_j u_j\right) = \sum_{j=1}^{n} \beta_j M^t u_j = \sum_{j=1}^{n} \beta_j \left(1 - \frac{\lambda_j}{2}\right)^t u_j.$$

Thus we get

$$\|M^t \mathbb{1}_x\|_2^2 = \sum_{j=1}^{n} \beta_j^2 \left(1 - \frac{\lambda_j}{2}\right)^{2t} \leq \sum_{j=1}^{k} \beta_j^2 + \left(1 - \frac{\lambda_{k+1}}{2}\right)^{2t} \cdot \sum_{j=k+1}^{n} \beta_j^2. \tag{75}$$

[6]It is interesting to note that a weaker average case version of this lemma was used in two prior works on testing graph cluster structure [CPS15] and [CKK+18]. The stronger version of the lemma presented here is important for spectral concentration bounds that we present, which are in turn crucial for sublinear time dot product access to the spectral embedding.

Note that G is (k, φ, ϵ)-clusterable, therefore by Lemma 3 we have $\lambda_{k+1} \geq \frac{\varphi^2}{2}$. Note that $t \geq \frac{20 \log n}{\varphi^2}$. Hence, we have

$$\left(1 - \frac{\lambda_{k+1}}{2}\right)^{2t} \leq n^{-10}. \tag{76}$$

Moreover since G is (k, φ, ϵ)-clusterable and $\min_i |C_i| \geq \Omega(\frac{n}{k})$ by Lemma 5 for all $j \in [k]$ we have

$$\beta_j \leq \|u_j\|_\infty \leq O(\sqrt{k} \cdot n^{-1/2+(20\epsilon/\varphi^2)}). \tag{77}$$

Thus by (75), (76) and (77) we get

$$\|M^t \mathbb{1}_x\|_2^2 \leq O(k \cdot k \cdot \frac{1}{n} \cdot n^{40\epsilon/\varphi^2}) + n \cdot n^{-10}.$$

Therefore we have

$$\|M^t \mathbb{1}_x\|_2 \leq O(k \cdot n^{-1/2+(20\epsilon/\varphi^2)}).$$

□

Combining the previous lemmas and Lemma 18 we obtain Lemma 23. We show that for (k, φ, ϵ)-clusterable graphs, the outer products of the columns of the t-step random walk transition matrix have small spectral norm. This is because the matrix power is mostly determined by the first k eigenvectors and by the fact that these eigenvectors have bounded infinity norm.

Lemma 23. *Let $k \geq 2$ be an integer, $\varphi \in (0, 1)$ and $\epsilon \in (0, 1)$. Let $G = (V, E)$ be a d-regular and (k, φ, ϵ)-clusterable graph. Let M be the random walk transition matrix of G. Let $1 > \xi > 1/n^5$, $t \geq \frac{20 \log n}{\varphi^2}$. Let $c > 1$ be a large enough constant and let $s \geq c \cdot k^4 \cdot n^{(1040 \cdot \epsilon/\varphi^2)} \log n/\xi^6$. Let $I_S = \{i_1, \ldots, i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1, \ldots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let $M^t = U\Sigma^t U^T$ be an eigendecomposition of M^t. Let $\sqrt{\frac{n}{s}} \cdot M^t S = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ be an SVD of $\sqrt{\frac{n}{s}} \cdot M^t S$ where $\widetilde{U} \in \mathbb{R}^{n \times n}, \widetilde{\Sigma} \in \mathbb{R}^{n \times n}, \widetilde{W} \in \mathbb{R}^{s \times n}$. If $\frac{\epsilon}{\varphi^2} \leq \frac{1}{10^5}$ then with probability at least $1 - n^{-100}$ matrix $\widetilde{\Sigma}_{[k]}^{-2}$ exists and we have*

$$\left|\left|U_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^T - \widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^T\right|\right|_2 < \xi$$

Proof. Let

$$A = (M^t)(M^t)^T = U\Sigma^{2t}U^T,$$

and

$$\widetilde{A} = \frac{n}{s}\left(M^t S\right)\left(M^t S\right)^T = \widetilde{U}\widetilde{\Sigma}^2\widetilde{U}^T.$$

Let γ_k and γ_{k+1} denote the k-th and $(k+1)$-th largest eigenvalues of A. Let U be an orthonormal basis of eigenvectors of L with corresponding eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$. Observe that each u_i is also an eigenvector of M, with eigenvalue $1 - \frac{\lambda_i}{2}$. Note that G is (k, φ, ϵ)-clusterable, therefore by Lemma 3 we have $\lambda_k \leq 2\epsilon$ and $\lambda_{k+1} \geq \frac{\varphi^2}{2}$. Note that $t \geq \frac{20 \log n}{\varphi^2}$. Hence, we have

$$\gamma_{k+1} = \left(1 - \frac{\lambda_{k+1}}{2}\right)^{2t} \leq n^{-10} \tag{78}$$

and

$$\gamma_k = \left(1 - \frac{\lambda_k}{2}\right)^{2t} \geq n^{(-80\epsilon/\varphi^2)}. \tag{79}$$

In order to apply Lemma 21 we need to derive an upper bound on the spectral norm of $(M^t\mathbb{1}_x)(M^t\mathbb{1}_x)^T$ for any column of A corresponding to vertex x. By Lemma 22 we have

$$B = \|(M^t\mathbb{1}_x)(M^t\mathbb{1}_x)^T\|_2 = \|M^t\mathbb{1}_x\|_2^2 \leq O(k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}).$$

Thus, with $1 \geq \xi > 1/n^5$ and for large enough c we have $s \geq c \cdot k^4 n^{(1040 \cdot \epsilon/\varphi^2)} \log n/\xi^6 \geq \frac{40n^2B^2 \log n}{1/32^6 \cdot \xi^6 n^{-960\epsilon/\varphi}}$. Thus by Lemma 21 we obtain that with probability at least $1 - n^{-100}$ that

$$\|A - \widetilde{A}\|_2 \leq \left(\frac{1}{32} \cdot \xi \cdot n^{-160\epsilon/\varphi^2}\right)^3. \tag{80}$$

We observe that equation 80 together with our bound on γ_k (79) and the positive semi-definiteness of $\widetilde{A}$ imply that the k largest eigenvalues of $\widetilde{A}$ are non-zero and so $\widetilde{\Sigma}_{[k]}^{-2}$ is exists with high probability.

Now observe that A is positive semi-definite, we have $\gamma_k/4 > \gamma_{k+1}$ and $\|A-\widetilde{A}\| \le \gamma_k/100$, so the preconditions of Lemma 18 are met and we have with probability $1-n^{-100}$

$$\left|\left|U_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^T - \widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^T\right|\right|_2 \le \frac{32\cdot\left(\left|\left|A-\widetilde{A}\right|\right|_2\right)^{1/3}}{\gamma_k^2} \le \xi.$$

□

Now we are ready to prove Lemma 19.

Lemma 19. *Let $k \ge 2$ be an integer, $\varphi \in (0,1)$ and $\epsilon \in (0,1)$. Let $G=(V,E)$ be a d-regular and (k,φ,ϵ)-clusterable graph. Let M be the random walk transition matrix of G. Let $1/n^5 < \xi < 1$, $t \ge \frac{20\log n}{\varphi^2}$. Let $c > 1$ be a large enough constant and let $s \ge c\cdot n^{(1280\cdot\epsilon/\varphi^2)}\cdot\log n\cdot k^{16}/\xi^6$. Let $I_S=\{i_1,\ldots,i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1,\ldots,n\}$. Let S be the $n\times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let $M^t = U\Sigma^tU^T$ be an eigendecomposition of M^t. Let $\sqrt{\frac{n}{s}}\cdot M^tS = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ be an SVD of $\sqrt{\frac{n}{s}}\cdot M^tS$ where $\widetilde{U}\in\mathbb{R}^{n\times n}, \widetilde{\Sigma}\in\mathbb{R}^{n\times n}, \widetilde{W}\in\mathbb{R}^{s\times n}$. If $\frac{\epsilon}{\varphi^2}\le\frac{1}{10^5}$ then with probability at least $1-n^{-100}$ matrix $\widetilde{\Sigma}_{[k]}^{-4}$ exists and we have*

$$\left|\mathbb{1}_x^TU_{[k]}U_{[k]}^T\mathbb{1}_y - (M^t\mathbb{1}_x)^T(M^tS)\left(\frac{n}{s}\cdot\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(M^tS)^T(M^t\mathbb{1}_y)\right| \le \frac{\xi}{n}.$$

Proof. Let $m_x = M^t\mathbb{1}_x$ and $m_y = M^t\mathbb{1}_y$. We first prove $m_x^T(U_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^T)m_y = \mathbb{1}_x^TU_{[k]}U_{[k]}^T\mathbb{1}_y$ and $m_x^T(M^tS)(\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T)(M^tS)^Tm_y = m_x^T\widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^Tm_y$. Then we upper bound

$$\left|m_x^TU_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^Tm_y - m_x^T\widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^Tm_y\right|.$$

Step 1: Note that $M^t = U\Sigma^tU^T$. Therefore we get $M^t\mathbb{1}_x = U\Sigma^tU^T\mathbb{1}_x$, and $M^t\mathbb{1}_y = U\Sigma^tU^T\mathbb{1}_y$. Thus we have

$$m_x^TU_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^Tm_y = \mathbb{1}_x^T\left(\left(U\Sigma^tU^T\right)\left(U_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^T\right)\left(U\Sigma^tU^T\right)\right)\mathbb{1}_y \tag{81}$$

Note that $U^TU_{[k]}$ is an $n\times k$ matrix such that the top $k\times k$ matrix is $I_{k\times k}$ and the rest is zero. Also $U_{[k]}^TU$ is a $k\times n$ matrix such that the left $k\times k$ matrix is $I_{k\times k}$ and the rest is zero. Therefore we have

$$U\Sigma^t\left(U^TU_{[k]}\right)\Sigma_{[k]}^{-2t}\left(U_{[k]}^TU\right)\Sigma^tU^T = UHU^T,$$

where H is an $n\times n$ matrix such that the top left $k\times k$ matrix is $I_{k\times k}$ and the rest is zero. Hence, we have

$$UHU^T = U_{[k]}U_{[k]}^T.$$

Thus we have

$$m_x^T(U_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^T)m_y = \mathbb{1}_x^TU_{[k]}U_{[k]}^T\mathbb{1}_y \tag{82}$$

Step 2: We have $\sqrt{\frac{n}{s}}\cdot M^tS = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ where $\widetilde{U}\in\mathbb{R}^{n\times n}$, $\widetilde{\Sigma}\in\mathbb{R}^{n\times n}$ and $\widetilde{W}\in\mathbb{R}^{s\times n}$. Therefore,

$$\begin{aligned}
&(m_x)^T(M^tS)\left(\frac{n}{s}\cdot\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(M^tS)^T(m_y)\\
&= m_x^T\left(\sqrt{\frac{s}{n}}\cdot\widetilde{U}\widetilde{\Sigma}\widetilde{W}^T\right)\left(\frac{n}{s}\cdot\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)\left(\sqrt{\frac{s}{n}}\cdot\widetilde{W}\widetilde{\Sigma}\widetilde{U}^T\right)m_y\\
&= m_x^T\left(\widetilde{U}\widetilde{\Sigma}\widetilde{W}^T\right)\left(\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)\left(\widetilde{W}\widetilde{\Sigma}\widetilde{U}^T\right)m_y
\end{aligned} \tag{83}$$

Note that $\widetilde{W}^T\widetilde{W}_{[k]}$ is an $n\times k$ matrix such that the top $k\times k$ matrix is $I_{k\times k}$ and the rest is zero. Also $\widetilde{W}_{[k]}^T\widetilde{W}$ is a $k\times n$ matrix such that the left $k\times k$ matrix is $I_{k\times k}$ and the rest is zero. Therefore we have

$$\widetilde{\Sigma}\left(\widetilde{W}^T\widetilde{W}_{[k]}\right)\widetilde{\Sigma}_{[k]}^{-4}\left(\widetilde{W}_{[k]}^T\widetilde{W}\right)\widetilde{\Sigma} = \widetilde{H},$$

where $\widetilde{H}$ is an $n \times n$ matrix such that the top left $k \times k$ matrix is $\widetilde{\Sigma}_{[k]}^{-2}$ and the rest is zero. Hence, we have

$$(\widetilde{U}\widetilde{\Sigma}\widetilde{W}^T)\left(\frac{n}{s} \cdot \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(\widetilde{W}\widetilde{\Sigma}\widetilde{U}^T) = \widetilde{U}\widetilde{H}\widetilde{U}^T = \widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^T \tag{84}$$

Putting (84) and (83) together we get

$$m_x^T(M^tS)(\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T)(M^tS)^T m_y = m_x^T\widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^T m_y \tag{85}$$

Put together: Let $c' > 1$ be a large enough constant we will set later. Let $\xi' = \frac{\xi}{c' \cdot k^2 \cdot n^{40\epsilon/\varphi^2}}$. Let c_1 be a constant in front of s in Lemma 23. Thus for large enough c we have $s \geq c \cdot n^{(1280 \cdot \epsilon/\varphi^2)} \cdot \log n \cdot k^{16}/\xi^6 \geq c_1 \cdot k^4 \cdot n^{(1040 \cdot \epsilon/\varphi^2)} \log n/\xi'^6$, hence, by Lemma 23 applied with ξ', with probability at least $1 - n^{-100}$ we have

$$\left|\left|U_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^T - \widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^T\right|\right|_2 \leq \xi'$$

Therefore by submultiplicativity of norm we have

$$\begin{aligned}\left|m_x^TU_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^Tm_y - m_x^T\widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^Tm_y\right| &\leq \left|\left|U_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^T - \widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^T\right|\right|_2 \|m_x\|_2\|m_y\|_2 \\ &\leq \xi'\|m_x\|_2\|m_y\|_2 \end{aligned} \tag{86}$$

Therefore we have

$$\begin{aligned}&\left|m_x^T(M^tS)\left(\frac{n}{s} \cdot \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(M^tS)^Tm_y - \mathbb{1}_x^TU_{[k]}U_{[k]}^T\mathbb{1}_y\right| \\ &= \left|m_x^T\widetilde{U}_{[k]}\widetilde{\Sigma}_{[k]}^{-2}\widetilde{U}_{[k]}^Tm_y - m_x^TU_{[k]}\Sigma_{[k]}^{-2t}U_{[k]}^Tm_y\right| && \text{By (82) and (85)} \\ &\leq \xi' \cdot \|m_x\|_2\|m_y\|_2 && \text{By (86)}\end{aligned} \tag{87}$$

By Lemma 22 for any vertex $x \in V$ we have

$$\|m_x\|_2^2 = \|M^t\mathbb{1}_x\|_2^2 \leq O(k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}). \tag{88}$$

Therefore by choice of c' as a large enough constant and choosing $\xi' = \frac{\xi}{c' \cdot k^2 \cdot n^{40\epsilon/\varphi^2}}$ we have

$$\left|m_x^T(M^tS)\left(\frac{n}{s} \cdot \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(M^tS)^Tm_y - \mathbb{1}_x^TU_{[k]}U_{[k]}^T\mathbb{1}_y\right| \leq O\left(\xi' \cdot k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right) \leq \frac{\xi}{n}. \tag{89}$$

□

5.4 Stability bounds under approximations of columns by random walks

The main result of this section is Lemma 24, which shows that if a graph is (k, φ, ϵ)-clusterable, then the pseudoinverseve of the low rank approximation of a random walk matrix are stable when it is empirically approximated by running random walks from sample vertices.

Lemma 24. *Let $k \geq 2$ be an integer, $\varphi \in (0,1)$ and $\epsilon \in (0,1)$. Let $G = (V,E)$ be a d-regular and (k,φ,ϵ)-clusterable graph. Let $1/n^5 < \xi < 1$ and $t \geq \frac{20\log n}{\varphi^2}$. Let $c_1 > 1$ and $c_2 > 1$ be a large enough constants. Let $s \geq c_1 \cdot n^{240\epsilon/\varphi^2} \cdot \log n \cdot k^4$ and $R \geq \frac{c_2 \cdot k^9 \cdot n^{(1/2+2100 \cdot \epsilon/\varphi^2)}}{\xi^6}$. Let $I_S = \{i_1, \ldots, i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1, \ldots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let $\mathcal{G} \in \mathbb{R}^{s \times s}$ be the output of* EstimateCollisionProbabilities(G, I_S, R, t) *(Algorithm 2). Let M be the random walk transition matrix of G. Let $\sqrt{\frac{n}{s}} \cdot M^tS = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ be an SVD of $\sqrt{\frac{n}{s}} \cdot M^tS$ where $\widetilde{U} \in \mathbb{R}^{n \times n}, \widetilde{\Sigma} \in \mathbb{R}^{n \times n}, \widetilde{W} \in \mathbb{R}^{s \times n}$. Let $\frac{n}{s} \cdot \mathcal{G} = \widehat{W}\widehat{\Sigma}\widehat{W}^T$ be an eigendecomposition of $\frac{n}{s} \cdot \mathcal{G}$. If $\frac{\epsilon}{\varphi^2} \leq \frac{1}{10^5}$ then with probability at least $1 - 2 \cdot n^{-100}$ matrices $\widehat{\Sigma}_{[k]}^{-2}$ and $\widetilde{\Sigma}_{[k]}^{-4}$ exist and we have*

$$\left|\left|\widehat{W}_{[k]}\widehat{\Sigma}_{[k]}^{-2}\widehat{W}_{[k]}^T - \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right|\right|_2 < \xi$$

To prove Lemma 24 we need the following lemma.

Lemma 25. *Let $k \geq 2$ be an integer, $\varphi \in (0,1)$ and $\epsilon \in (0,1)$. Let $G = (V,E)$ be a d-regular and (k,φ,ϵ)-clusterable graph. Let L and M be the normalized Laplacian and transition matrix of G respectively. For any $t \geq \frac{10\log n}{\varphi^2}$ and any r and any $x \in V$ we have*

$$\|M^t \mathbb{1}_x\|_r \leq O\left(k^2 \cdot n^{-1+1/r+(40\epsilon/\varphi^2)}\right).$$

Proof. Let L be the normalized Laplacian of G with eigenvectors $u_1, \ldots, u_n$ and corresponding eigenvalues $\lambda_1 \leq \ldots \leq \lambda_n$. Observe that each u_i is also an eigenvector of M, with eigenvalue $1 - \frac{\lambda_i}{2}$. Note that G is (k,φ,ϵ)-clusterable. Therefore by Lemma 3 we have

$$\lambda_{k+1} \geq \frac{\varphi^2}{2}. \tag{90}$$

We write $\mathbb{1}_x$ in the eigenbasis of L as $\mathbb{1}_x = \sum_{j=1}^n \beta_j u_j$ where $\beta_j = u_j \cdot \mathbb{1}_x = u_j(x)$. Thus for any vertex u we have

$$M^t \mathbb{1}_x = M^t \left(\sum_{j=1}^n \beta_j u_j\right) = \sum_{j=1}^n \beta_j M^t u_j = \sum_{j=1}^n \beta_j \left(1 - \frac{\lambda_j}{2}\right)^t u_j.$$

Let $m_x = M^t \mathbb{1}_x$. Therefore for any vertex $y \in V$ we have

$$\begin{aligned} m_x(y) &= \sum_{j=1}^n \beta_j \left(1 - \frac{\lambda_j}{2}\right)^t u_j(y) \\ &= \sum_{j=1}^k \beta_j \left(1 - \frac{\lambda_j}{2}\right)^t u_j(y) + \sum_{j=k+1}^n \beta_j \left(1 - \frac{\lambda_j}{2}\right)^t u_j(y) \end{aligned}$$

Therefore,

$$|m_x(y)| \leq \left(1 - \frac{\lambda_1}{2}\right)^t \sum_{j=1}^k |\beta_j| \cdot |u_j(y)| + \left(1 - \frac{\lambda_{k+1}}{2}\right)^t \sum_{j=k+1}^n |\beta_j| \cdot |u_j(y)| \tag{91}$$

By (90) we have $\lambda_{k+1} \geq \frac{\varphi^2}{2}$, and $t \geq \frac{8\log n}{\varphi^2}$. Thus we have

$$\left(1 - \frac{\lambda_{k+1}}{2}\right)^t \leq n^{-2}$$

Note that for any $j \in [n]$

$$|\beta_j| \leq \sqrt{\sum_{j=1}^n \beta_j^2} = \|\mathbb{1}_x\|_2 = 1. \tag{92}$$

Morover for any $j \in [n]$ and any $y \in V$

$$|u_j(y)| \leq \|u_j\|_2 = 1 \tag{93}$$

Putting (92), (93) and (91) together we get

$$\begin{aligned} |m_x(y)| &\leq \sum_{j=1}^k |\beta_j| \cdot |u_j(y)| + \left(1 - \frac{\lambda_{k+1}}{2}\right)^t \sum_{j=k+1}^n |\beta_j| \cdot |u_j(y)| \\ &\leq \sum_{j=1}^k |\beta_j| \cdot |u_j(y)| + n^{-2} \cdot n \end{aligned} \tag{94}$$

Note that G is (k,φ,ϵ)-clusterable and $\min_i |C_i| \geq \Omega(\frac{n}{k})$. Therefore by Lemma 5 for all $j \leq k$ we have

$$\beta_j = u_j(x) \leq \|u_j\|_\infty \leq O\left(\sqrt{k} \cdot n^{-1/2+(20\epsilon/\varphi^2)}\right).$$

Moreover

$$u_j(y) \leq \|u_j\|_\infty \leq O\left(\sqrt{k} \cdot n^{-1/2+(20\epsilon/\varphi^2)}\right)$$

Thus, we get

$$\sum_{j=1}^{k} |\beta_j| \cdot |u_j(y)| \le O\left(k \cdot k \cdot n^{-1+(40\epsilon/\varphi^2)}\right). \tag{95}$$

Therefore by (94) and (95) we get

$$\begin{aligned} |m_x(y)| &\le O\left(k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right) + n^{-1} \\ &\le O\left(k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right). \end{aligned} \tag{96}$$

Therefore we have

$$\|m_x\|_r \le \left(n \cdot O\left(k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right)^r\right)^{1/r} = O\left(k^2 \cdot n^{-1+1/r+(40\epsilon/\varphi^2)}\right).$$

□

Lemma 26. *Let $k \ge 2$ be an integer, $\varphi \in (0,1)$ and $\epsilon \in (0,1)$. Let $G = (V,E)$ be a d-regular and (k,φ,ϵ)-clusterable graph. Let M be the random walk transition matrix of G. Let $\sigma_{err} > 0$. Let t, R_1 and R_2 be integers. Let $a, b \in V$. Suppose that we run R_1 random walks of length t from vertex a and R_2 random walks of length t from vertex b. For any $x \in V$, let $\widehat{m}_a(x)$ (resp. $\widehat{m}_b(x)$) be a random variable which denotes the fraction out of the R_1 (resp. R_2) random walks starting from a (resp. b), which end in x. Let $c > 1$ be a large enough constant. If*

$$\min(R_1, R_2) \ge \frac{c \cdot k^5 \cdot n^{-2+(100\epsilon/\varphi^2)}}{\sigma_{err}^2}, \text{ and } R_1 R_2 \ge \frac{c \cdot k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}}{\sigma_{err}^2}$$

then with probability at least 0.99 we have

$$|\widehat{m}_a^T \widehat{m}_b - (M^t \mathbb{1}_a)^T (M^t \mathbb{1}_b)| \le \sigma_{err}.$$

Remark 5. *The success probability of Lemma 26 can be boosted up to $1-n^{-100}$ using standard techniques (taking the median of $O(\log n)$ independent runs).*

Proof. Let $m_a = M^t \mathbb{1}_a$ and $m_b = M^t \mathbb{1}_b$. Let $X^i_{a,r}$ be a random variable which is 1 if the r^{th} random walk starting from a, ends at vertex i, and 0 otherwise. Let $Y^i_{b,r}$ be a random variable which is 1 if the r^{th} random walk starting from b, ends at vertex i, and 0 otherwise. Thus, $\mathbb{E}[X^i_{a,r}] = m_a(i)$ and $\mathbb{E}[Y^i_{b,r}] = m_b(i)$. For any two vertices $a, b \in S$, let $Z_{a,b} = \widehat{m}_a^T \widehat{m}_b$ be a random variable given by

$$Z_{a,b} = \frac{1}{R_1 R_2} \sum_{i \in V} (\sum_{r_1=1}^{R_1} X^i_{a,r_1})(\sum_{r_2=1}^{R_2} Y^i_{b,r_2}).$$

Thus,

$$\begin{aligned} \mathbb{E}[Z_{a,b}] &= \frac{1}{R_1 R_2} \sum_{i \in V} (\sum_{r_1=1}^{R_1} \mathbb{E}[X^i_{a,r_1}])(\sum_{r_2=1}^{R_2} \mathbb{E}[Y^i_{b,r_2}]) \\ &= \frac{1}{R_1 R_2} \sum_{i \in V} (R_1 \cdot m_a(i))(R_2 \cdot m_b(i)) \\ &= \sum_{i \in V} m_a(i) \cdot m_b(i) = (m_a)^T (m_b). \end{aligned} \tag{97}$$

We know that $\mathrm{Var}(Z_{a,b}) = \mathbb{E}[Z_{a,b}^2] - \mathbb{E}[Z_{a,b}]^2$. Let us first compute $\mathbb{E}[Z_{a,b}^2]$.

$$\begin{aligned} \mathbb{E}[Z_{a,b}^2] &= \mathbb{E}\left[\frac{1}{(R_1 R_2)^2} \sum_{i \in V} \sum_{j \in V} \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_1'=1}^{R_1} \sum_{r_2'=1}^{R_2} X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r_1'} Y^j_{b,r_2'}\right] \\ &= \frac{1}{(R_1 R_2)^2} \sum_{i \in V} \sum_{j \in V} \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_1'=1}^{R_1} \sum_{r_2'=1}^{R_2} \mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r_1'} Y^j_{b,r_2'}] \end{aligned}$$

To compute $\mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r_1'} Y^j_{b,r_2'}]$, we need to consider the following cases.

1. $i \neq j$: $\mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r'_1} Y^j_{b,r'_2}] \leq m_a(i) \cdot m_b(i) \cdot m_a(j) \cdot m_b(j)$. (This is an equality if $r_1 \neq r'_1$ and $r_2 \neq r'_2$. Otherwise, the expectation is zero.)
2. $i = j, \quad r_1 = r'_1, \quad r_2 = r'_2$: $\mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r'_1} Y^j_{b,r'_2}] = m_a(i) \cdot m_b(i)$.
3. $i = j, \quad r_1 = r'_1, \quad r_2 \neq r'_2$: $\mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r'_1} Y^j_{b,r'_2}] = m_a(i) \cdot m_b(i) \cdot \cdot m_b(i)$.
4. $i = j, \quad r_1 \neq r'_1, \quad r_2 = r'_2$: $\mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r'_1} Y^j_{b,r'_2}] = m_a(i) \cdot m_b(i) \cdot m_a(i)$.
5. $i = j, \quad r_1 \neq r'_1, \quad r_2 \neq r'_2$: $\mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r'_1} Y^j_{b,r'_2}] = m_a(i) \cdot m_b(i) \cdot m_a(i) \cdot m_b(i)$.

Thus we have,

$$\begin{aligned}
\mathbb{E}[Z^2_{a,b}] &= \frac{1}{(R_1R_2)^2} \sum_{i \in V} \sum_{j \in V} \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r'_1=1}^{R_1} \sum_{r'_2=1}^{R_2} \mathbb{E}[X^i_{a,r_1} Y^i_{b,r_2} X^j_{a,r'_1} Y^j_{b,r'_2}] \\
&\leq \sum_{i \in V} \sum_{j \in V \setminus \{i\}} m_a(i) \cdot m_a(j) \cdot m_b(i) \cdot m_b(j) + \sum_{i \in V} m_a(i)^2 \cdot m_b(i)^2 \\
&+ \frac{1}{R_1R_2} \sum_{i \in V} m_a(i) \cdot m_b(i) + \frac{1}{R_1} \sum_{i \in V} m_a(i) \cdot m_b(i)^2 + \frac{1}{R_2} \sum_{i \in V} m_a(i)^2 \cdot m_b(i) \\
&= \sum_{i,j \in V} m_a(i) \cdot m_a(j) \cdot m_b(i) \cdot m_b(j) + \frac{1}{R_1R_2} \sum_{i \in V} m_a(i) \cdot m_b(i) \\
&+ \frac{1}{R_1} \sum_{i \in V} m_a(i) \cdot m_b(i)^2 + \frac{1}{R_2} \sum_{i \in V} m_a(i)^2 \cdot m_b(i).
\end{aligned}$$

Therefore we get,

$$\begin{aligned}
\mathrm{Var}(Z_{a,b}) &= \mathbb{E}[Z^2_{a,b}] - \mathbb{E}[Z_{a,b}]^2 \\
&\leq \sum_{i,j \in V} m_a(i) \cdot m_a(j) \cdot m_b(i) \cdot m_b(j) + \frac{1}{R_1R_2} \sum_{i \in V} m_a(i) \cdot m_b(i) \qquad (98) \\
&+ \frac{1}{R_1} \sum_{i \in V} m_a(i) \cdot m_b(i)^2 + \frac{1}{R_2} \sum_{i \in V} m_a(i)^2 \cdot m_b(i) - \left(\sum_{i \in V} m_a(i) \cdot m_b(i) \right)^2 \\
&= \frac{1}{R_1R_2} \sum_{i \in V} m_a(i) \cdot m_b(i) + \frac{1}{R_1} \sum_{i \in V} m_a(i) \cdot m_b(i)^2 + \frac{1}{R_2} \sum_{i \in V} m_a(i)^2 \cdot m_b(i) \\
&\leq \frac{1}{R_1R_2} \|m_a\|_2 \|m_b\|_2 + \frac{1}{R_1} \|m_a\|_2 \|m_b\|_4^2 + \frac{1}{R_2} \|m_a\|_4^2 \|m_b\|_2 \qquad \text{By Cauchy-Schwarz}
\end{aligned}$$

Since $G = (V, E)$ is (k, φ, ϵ) clusterable by Lemma 25 we have

$$\|m_a\|_4 \leq O\left(k^2 \cdot n^{-3/4 + (40\epsilon/\varphi^2)}\right).$$

and by Lemma 22 we have

$$\|m_a\|_2 \leq O(k \cdot n^{-1/2 + (20\epsilon/\varphi^2)}).$$

Thus we get

$$\mathrm{Var}(Z_{a,b}) \leq O\left(\frac{k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}}{R_1R_2} + \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \cdot k^5 \cdot n^{-2+(100\epsilon/\varphi^2)} \right) \qquad (99)$$

Then by Chebyshev's inequality, we get,

$$\begin{aligned}
\Pr\left[|Z_{a,b} - \mathbb{E}[Z_{a,b}]| > \sigma_{\mathrm{err}} \right] &\leq \frac{\mathrm{Var}[Z_{a,b}]}{\sigma_{\mathrm{err}}{}^2} \\
&\leq O\left(\frac{1}{\sigma_{\mathrm{err}}{}^2} \cdot \left(\frac{k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}}{R_1R_2} + \left(\frac{1}{R_1} + \frac{1}{R_2} \right) \cdot k^5 \cdot n^{-2+(100\epsilon/\varphi^2)} \right) \right) \qquad (100) \\
&\leq \frac{1}{100}.
\end{aligned}$$

The last inequality holds by our choice of R_1 and R_2 as follows where c is a large enough constant that cancels the constant hidden in $O(\cdot)$ in (100).

$$\min(R_1, R_2) \geq \frac{c \cdot k^5 \cdot n^{-2+(100\epsilon/\varphi^2)}}{\sigma_{\text{err}}^2}$$

and

$$R_1 R_2 \geq \frac{c \cdot k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}}{\sigma_{\text{err}}^2}$$

□

Lemma 27. *Let $k \geq 2$ be an integer, $\varphi \in (0,1)$ and $\epsilon \in (0,1)$. Let $G = (V, E)$ be a d-regular and (k, φ, ϵ)-clusterable graph. Let $\sigma_{err} > 0$ and let $s > 0$, $R > 0$, $t > 0$ be integers. Let $I_S = \{i_1, \ldots, i_s\}$ be a multiset of s indices chosen from $\{1, \ldots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let $c > 1$ be a large enough constant. Let $R \geq \max\left\{\frac{c \cdot k^5 \cdot n^{-2+100\epsilon/\varphi^2}}{\sigma_{err}^2}, \frac{c \cdot k \cdot n^{-1/2+20\epsilon/\varphi^2}}{\sigma_{err}}\right\}$ Let $\mathcal{G} \in \mathbb{R}^{s \times s}$ be the output of Algorithm* EstimateCollisionProbabilities(G, I_S, R, t) *(Algorithm 2). Let M be the random walk transition matrix of G. then with probability at least $1 - n^{-100}$ we have*

$$\|\mathcal{G} - (M^t S)^T (M^t S)\|_2 \leq s \cdot \sigma_{err}.$$

Proof. Note that as per line (2) and (3) of Algorithm 2 we first construct matrices $\widehat{P}_i \in \mathbb{R}^{n \times s}$ and $\widehat{Q}_i \in \mathbb{R}^{n \times s}$ using Algorithm 3. as per line (3) of Algorithm 3 matrix $\widehat{P}_i$ (or $\widehat{Q}_i$) has s columns each corresponds to a vertex $x \in S$. The column corresponding to vertex x is $\widehat{m}_x$. as per line 2 of Algorithm 3, $\widehat{m}_x$ is defined as the empirical probability distribution of running R random walks of length t starting from vertex x. Thus for any $x, y \in S$ we have the entry corresponding to the x^{th} row and y^{th} column of $\widehat{Q}_i^T \widehat{P}_i$ (or $\widehat{P}_i^T \widehat{Q}_i$) is $\langle \widehat{m}_x, \widehat{m}_y \rangle$. Since

$$R \geq \max\left\{\frac{c \cdot k^5 \cdot n^{-2+100\epsilon/\varphi^2}}{\sigma_{\text{err}}^2}, \frac{c \cdot k \cdot n^{-1/2+20\epsilon/\varphi^2}}{\sigma_{\text{err}}}\right\}$$

then by Lemma 26 with probability at least 0.99 we have

$$|\widehat{m}_x^T \widehat{m}_y - (M^t \mathbb{1}_x)^T (M^t \mathbb{1}_y)| \leq \sigma_{\text{err}}.$$

Note that as per line 4 of Algorithm 2 we define $\mathcal{G}_i := \frac{1}{2}\left(\widehat{P}_i^T \widehat{Q}_i + \widehat{Q}_i^T \widehat{P}_i\right)$. Thus for any $x, y \in I_S$ we have the entry corresponding to the x^{th} row and y^{th} column of $\mathcal{G}_i$ (i.e., $\mathcal{G}_i(x, y)$) with probability 0.99 satisfies the following:

$$|\mathcal{G}_i(x, y) - (M^t \mathbb{1}_x)^T (M^t \mathbb{1}_y)| \leq \sigma_{\text{err}}.$$

Note that as Line 5 of Algorithm 2 we define $\mathcal{G}$ as a matrix obtained by taking the entrywises median of $\mathcal{G}_i$'s over $O(\log n)$ runs. Thus with probability at least $1 - n^{-100}$ we have for all $x, y \in I_S$

$$|\mathcal{G}(x, y) - (M^t \mathbb{1}_x)^T (M^t \mathbb{1}_y)| \leq \sigma_{\text{err}}.$$

which implies

$$\|\mathcal{G} - (M^t S)^T (M^t S)^T\|_F \leq s \cdot \sigma_{\text{err}}.$$

Since the Frobenius norm of a matrix bounds its maximum eigenvalue from above we get

$$\|\mathcal{G} - (M^t S)^T (M^t S)^T\|_2 \leq s \cdot \sigma_{\text{err}}.$$

□

Recall that for a symmetric matrix A, we write $\nu_i(A)$ (resp. $\nu_{\max}(A), \nu_{\min}(A)$) to denote the i^{th} largest (resp. maximum, minimum) eigenvalue of A.

Lemma 28. *Let $k \geq 2$ be an integer, $\varphi \in (0,1)$ and $\epsilon \in (0,1)$. Let $G = (V, E)$ be a d-regular and (k, φ, ϵ)-clusterable graph. Let $t \geq \frac{20 \log n}{\varphi^2}$. Let $c > 1$ be a large enough constant and $s \geq c \cdot n^{240 \cdot \epsilon/\varphi^2} \cdot \log n \cdot k^4$. Let $I_S = \{i_1, \ldots, i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1, \ldots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let M be the random walk transition matrix of G. If $\frac{\epsilon}{\varphi^2} \leq \frac{1}{10^5}$ then with probability at least $1 - n^{-100}$ we have*

1. $\nu_k\left(\frac{n}{s}\cdot(M^tS)(M^tS)^T\right)\geq\frac{n^{-80\epsilon/\varphi^2}}{2}$
2. $\nu_{k+1}\left(\frac{n}{s}\cdot(M^tS)(M^tS)^T\right)\leq n^{-9}$.

Proof. Let $(u_1,\ldots,u_n)$ be an orthonormal basis of eigenvectors of L with corresponding eigenvalues $0\leq\lambda_1\leq\ldots\leq\lambda_n$. Observe that each u_i is also an eigenvector of M, with eigenvalue $1-\frac{\lambda_i}{2}$. Note that G is (k,φ,ϵ)-clusterable, therefore by Lemma 3 we have $\lambda_k\leq 2\epsilon$ and $\lambda_{k+1}\geq\frac{\varphi^2}{2}$. We have

$$\nu_{k+1}(M^{2t})=\left(1-\frac{\lambda_{k+1}}{2}\right)^{2t}\leq n^{-10},\text{ and}\tag{101}$$

$$\nu_k(M^{2t})=\left(1-\frac{\lambda_k}{2}\right)^{2t}\geq n^{-80\epsilon/\varphi^2}\tag{102}$$

Proof of item (1)**:** Let $A=(M^t)(M^t)^T$, and $\widetilde{A}=\frac{n}{s}\cdot(M^tS)(M^tS)^T$. By Lemma 22 we have

$$B=\|(M^t\mathbb{1}_x)(M^t\mathbb{1}_x)^T\|_2\leq\|M^t\mathbb{1}_x\|_2^2\leq O\left(k^2\cdot n^{-1+40\epsilon/\varphi^2)}\right).$$

Let $\xi=n^{-80\epsilon/\varphi^2}/2$. Therefore for large enough constant c and by choice of $s=c\cdot k^4n^{240\epsilon/\varphi^2}\log n$ we have $s\geq\frac{40n^2B^2\log n}{(\xi)^2}$. Thus Lemma 21 yields that with probability at least $1-\frac{1}{n^{100}}$ we have

$$\|A-\widetilde{A}\|_2\leq\frac{n^{-80\epsilon/\varphi^2}}{2}.\tag{103}$$

Hence, by Weyl's Inequality (see Lemma 17) we have

$$\nu_k(\widetilde{A})\geq\nu_k(A)+\nu_{\min}(\widetilde{A}-A)=\nu_k(A)-\nu_{\max}(A-\widetilde{A})=\nu_k(A)-\|A-\widetilde{A}\|_2$$

By (102) we have $\nu_k(A)=\nu_k(M^{2t})\geq n^{-10\epsilon/\varphi^2}$ and so

$$\nu_k(\widetilde{A})\geq\nu_k(A)-\|\widetilde{A}-A\|_2\geq n^{-80\epsilon/\varphi^2}-\frac{n^{-80\epsilon/\varphi^2}}{2}\geq\frac{n^{-80\epsilon/\varphi^2}}{2}.$$

Proof of item (2)**:** By Lemma 8 we have

$$\nu_{k+1}(\widetilde{A})=\frac{n}{s}\cdot\nu_{k+1}((M^tS)(M^tS)^T)=\frac{n}{s}\cdot\nu_{k+1}((M^tS)^T(M^tS))=\frac{n}{s}\cdot\nu_{k+1}(S^TM^{2t}S).$$

Recall that $1-\frac{\lambda_1}{2}\geq\cdots\geq 1-\frac{\lambda_n}{2}$ are the eigenvalues of M, and Σ is the diagonal matrix of these eigenvalues in descending order, and U is the matrix whose columns are orthonormal eigenvectors of M arranged in descending order of their eigenvalues. We have $M^{2t}=U\Sigma^{2t}U^T$. Recall that $\Sigma_{[k]}$ is $k\times k$ diagonal matrix with entries $1-\frac{\lambda_1}{2}\geq\cdots\geq 1-\frac{\lambda_k}{2}$, and $\Sigma_{-[k]}$ is a $(n-k)\times(n-k)$ diagonal matrix with entries $1-\frac{\lambda_{k+1}}{2}\geq\cdots\geq 1-\frac{\lambda_n}{2}$. We can write $U\Sigma^{2t}U=U_{[k]}\Sigma_{[k]}^{2t}U_{[k]}^T+U_{-[k]}\Sigma_{-[k]}^{2t}U_{-[k]}^T$, thus we get

$$\begin{aligned}
\nu_{k+1}(\widetilde{A})&=\frac{n}{s}\cdot\nu_{k+1}\left(S^TM^{2t}S\right)\\
&=\frac{n}{s}\cdot\nu_{k+1}\left(S^T(U\Sigma^{2t}U^T)S\right)\\
&=\frac{n}{s}\cdot\nu_{k+1}\left(S^T\left(U_{[k]}\Sigma_{[k]}^{2t}U_{[k]}^T+U_{-[k]}\Sigma_{-[k]}^{2t}U_{-[k]}^T\right)S\right)\\
&\leq\frac{n}{s}\cdot\nu_{k+1}\left(S^TU_{[k]}\Sigma_{[k]}^{2t}U_{[k]}^TS\right)+\frac{n}{s}\cdot\nu_{\max}\left(S^TU_{-[k]}\Sigma_{-[k]}^{2t}U_{-[k]}^TS\right)\quad\text{By Weyl's inequality (Lemma 17)}
\end{aligned}$$

Here $\nu_{k+1}(S^TU_{[k]}\Sigma_{[k]}^{2t}U_{[k]}^TS)=0$, because the rank of $\Sigma_{[k]}^{2t}$ is k. We then need to bound $\nu_{\max}(S^TU_{-[k]}\Sigma_{-[k]}^{2t}U_{-[k]}^TS)$. We have,

$$\begin{aligned}
&\nu_{\max}\left(S^TU_{-[k]}\Sigma_{-[k]}^{2t}U_{-[k]}^TS\right)=\nu_{\max}\left(U_{-[k]}\Sigma_{-[k]}^{2t}U_{-[k]}^TSS^\top\right) &&\text{By Lemma 8}\\
&\leq\nu_{\max}\left(U_{-[k]}\Sigma_{-[k]}^{2t}U_{-[k]}^T\right)\cdot\nu_{\max}\left(SS^\top\right) &&\text{By submultiplicativity of norm}\\
&=\nu_{\max}\left(\Sigma_{-[k]}^{2t}U_{-[k]}^TU_{-[k]}\right)\cdot\nu_{\max}\left(SS^\top\right) &&\text{By Lemma 8}\\
&=\nu_{\max}\left(\Sigma_{-[k]}^{2t}\right)\cdot\nu_{\max}\left(SS^\top\right) &&\text{Since } U_{-[k]}^TU_{-[k]}=I
\end{aligned}$$

Next, observe that $SS^\top \in \mathbb{R}^{n\times n}$ is a diagonal matrix whose $(a,a)^{\text{th}}$ entry is the multiplicity of vertex a is sampled in S. Thus, $\nu_{\max}(SS^\top)$ is the maximum multiplicity over all vertices, which is at most s. Also note that $\nu_{\max}(\Sigma_{-[k]}^{2t}) = \left(1-\frac{\lambda_{k+1}}{2}\right)^{2t}$. Thus by (101) we get,

$$\nu_{k+1}(\widetilde{A}) \le \frac{n}{s}\cdot \nu_{\max}\left(S^T U_{-[k]}\Sigma_{-[k]}^{2t} U_{-[k]}^T S\right) \le \frac{n}{s}\cdot s\cdot \left(1-\frac{\lambda_{k+1}}{2}\right)^{2t} \le n\cdot n^{-10} = n^{-9}.$$

□

Now we are ready to prove the main result of this section (Lemma 24).

Lemma 24. *Let $k\ge 2$ be an integer, $\varphi\in(0,1)$ and $\epsilon\in(0,1)$. Let $G=(V,E)$ be a d-regular and (k,φ,ϵ)-clusterable graph. Let $1/n^5<\xi<1$ and $t\ge \frac{20\log n}{\varphi^2}$. Let $c_1>1$ and $c_2>1$ be a large enough constants. Let $s\ge c_1\cdot n^{240\epsilon/\varphi^2}\cdot \log n\cdot k^4$ and $R\ge \frac{c_2\cdot k^9\cdot n^{(1/2+2100\cdot\epsilon/\varphi^2)}}{\xi^6}$. Let $I_S=\{i_1,\dots,i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1,\dots,n\}$. Let S be the $n\times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let $\mathcal{G}\in\mathbb{R}^{s\times s}$ be the output of* EstimateCollisionProbabilities(G,I_S,R,t) *(Algorithm 2). Let M be the random walk transition matrix of G. Let $\sqrt{\frac{n}{s}}\cdot M^tS=\widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ be an SVD of $\sqrt{\frac{n}{s}}\cdot M^tS$ where $\widetilde{U}\in\mathbb{R}^{n\times n}, \widetilde{\Sigma}\in\mathbb{R}^{n\times n}, \widetilde{W}\in\mathbb{R}^{s\times n}$. Let $\frac{n}{s}\cdot\mathcal{G}=\widehat{W}\widehat{\Sigma}\widehat{W}^T$ be an eigendecomposition of $\frac{n}{s}\cdot\mathcal{G}$. If $\frac{\epsilon}{\varphi^2}\le\frac{1}{10^5}$ then with probability at least $1-2\cdot n^{-100}$ matrices $\widehat{\Sigma}_{[k]}^{-2}$ and $\widetilde{\Sigma}_{[k]}^{-4}$ exist and we have*

$$\left|\left|\widehat{W}_{[k]}\widehat{\Sigma}_{[k]}^{-2}\widehat{W}_{[k]}^T - \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right|\right|_2 < \xi$$

Proof. Let $\widetilde{A}=\frac{n}{s}\cdot (M^tS)^T(M^tS)=\widetilde{W}\widetilde{\Sigma}^2\widetilde{W}^T$ and $\widehat{A}=\frac{n}{s}\cdot\mathcal{G}$. Thus we have

$$\widetilde{A}^2=\left(\frac{n}{s}\cdot (M^tS)^T\left(M^tS\right)\right)^2=\widetilde{W}\widetilde{\Sigma}^4\widetilde{W}^T$$

and

$$\widehat{A}^2=\left(\frac{n}{s}\cdot\mathcal{G}\right)^2=\widehat{W}\widehat{\Sigma}^2\widehat{W}^T.$$

Recall that for a symmetric matrix A, we write $\nu_i(A)$ to denote the i^{th} largest eigenvalue of A. We want to apply Lemma 18 to get

$$\left|\left|\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T - \widehat{W}_{[k]}\widehat{\Sigma}_{[k]}^{-2}\widehat{W}_{[k]}^T\right|\right|_2 \le \frac{32\cdot\left(\left|\left|\widetilde{A}^2-\widehat{A}^2\right|\right|_2\right)^{1/3}}{\nu_k(\widetilde{A}^2)^2}$$

Hence, we first need to verify the prerequisites of Lemma 18. Let $c_3>1$ be a large enough constant that we will define soon, and let $\sigma_{\text{err}}=\frac{\xi^3\cdot n^{(-1-10^3\cdot\epsilon/\varphi^2)}}{c_3\cdot k^2}$. Let c be a constant from Lemma 27. By the assumption of the lemma for large enough constant $c_2>1$ we have

$$R\ge\frac{c_2\cdot k^9\cdot n^{1/2+2100\cdot\epsilon/\varphi^2}}{\xi^6}\ge\max\left\{\frac{c\cdot k^5\cdot n^{-2+100\epsilon/\varphi^2}}{\sigma_{\text{err}}^2},\frac{c\cdot k\cdot n^{-1/2+20\epsilon/\varphi^2}}{\sigma_{\text{err}}}\right\}.$$

Thus we can apply Lemma 27. Hence, with probability at least $1-n^{-100}$ we have

$$\|\mathcal{G}-(M^tS)^T(M^tS)\|_2\le s\cdot\sigma_{\text{err}}. \tag{104}$$

Therefore we have

$$\begin{aligned}\|\mathcal{G}^2-\left((M^tS)^T(M^tS)\right)^2\|_2 &= \|\mathcal{G}\left(\mathcal{G}-(M^tS)^T(M^tS)\right)+\left(\mathcal{G}-(M^tS)^T(M^tS)\right)(M^tS)^T(M^tS)\|_2\\ &\le \|\mathcal{G}-(M^tS)^T(M^tS)\|_2\left(\|\mathcal{G}\|_2+\|(M^tS)^T(M^tS)\|_2\right)\\ &\le s\cdot\sigma_{\text{err}}\left((s\cdot\sigma_{\text{err}}+\|(M^tS)^T(M^tS)\|_2)+\|(M^tS)^T(M^tS)\|_2\right)\\ &= (s\cdot\sigma_{\text{err}})^2+2\cdot s\cdot\sigma_{\text{err}}\|(M^tS)^T(M^tS)\|_2 \end{aligned} \tag{105}$$

Note that

$$
\begin{aligned}
\|(M^tS)^T(M^tS)\|_2 &\le \|(M^tS)^T(M^tS)\|_F \\
&= \sqrt{\sum_{x,y\in S}((M^t\mathbb{1}_x)^T(M^t\mathbb{1}_y))^2} \\
&\le \sqrt{\sum_{x,y\in S}\|M^t\mathbb{1}_x\|_2^2\|M^t\mathbb{1}_y\|_2^2} && \text{By Cauchy Schwarz} \\
&\le O\left(\sqrt{s^2\cdot\left(k^2\cdot n^{-1+(40\epsilon/\varphi^2)}\right)^2}\right) && \text{By Lemma 22} \\
&= O\left(s\cdot k^2\cdot n^{-1+(40\epsilon/\varphi^2)}\right) && . \qquad (106)
\end{aligned}
$$

Puuting (106) and (105) and by choice of $\sigma_{\text{err}} = \frac{\xi^3\cdot n^{(-1-10^3\cdot\epsilon/\varphi^2)}}{c_3\cdot k^2}$ we get

$$
\|\widetilde{A}^2-\widehat{A}^2\|_2 = \left(\frac{n}{s}\right)^2\|\mathcal{G}^2-\left((M^tS)^T(M^tS)\right)^2\|_2 \le O\left(\frac{\xi^6\cdot n^{-2\cdot 10^3\cdot\epsilon/\varphi^2}}{(c_3)^2\cdot k^4}+\frac{\xi^3\cdot n^{-960\epsilon/\varphi^2}}{c_3}\right) = O\left(\frac{\xi^3\cdot n^{-960\epsilon/\varphi^2}}{c_3}\right) \tag{107}
$$

By Lemma 8 for any $i\in[s]$ we have

$$
\nu_i(\widetilde{A}) = \nu_i\left(\frac{n}{s}\cdot(M^tS)\left(M^tS\right)^T\right) = \nu_i\left(\frac{n}{s}\cdot(M^tS)^T\left(M^tS\right)\right)
$$

Let c_1 be the constant from Lemma 28. Since $s\ge c_1\cdot n^{240\epsilon/\varphi^2}\cdot\log n\cdot k^4$ therefore by Lemma 28 with probability at least $1-n^{-100}$ we have

$$
\nu_k\left(\widetilde{A}^2\right) = \nu_k\left(\left(\frac{n}{s}\cdot(M^tS)^T\left(M^tS\right)\right)^2\right) \ge \left(\frac{n^{-80\epsilon/\varphi^2}}{2}\right)^2 \ge \frac{n^{-160\epsilon/\varphi^2}}{4} \tag{108}
$$

and

$$
\nu_{k+1}\left(\widetilde{A}^2\right) = \nu_{k+1}\left(\left(\frac{n}{s}\cdot(M^tS)^T\left(M^tS\right)\right)^2\right) \le n^{-18} \tag{109}
$$

By the bound on the $\nu_k(\widetilde{A}^2)$ and the inequality on $\|\widetilde{A}^2-\widehat{A}^2\|_2$, we know that $\nu_k(\widehat{A}^2)$ is non-zero and so $\widehat{\Sigma}_{[k]}^{-2}$ exist. Recall that $\widetilde{A}=\widetilde{W}\widetilde{\Sigma}^2\widetilde{W}^T$. Observing that $\widetilde{A}$ is positive semi-definite, $\nu_{k+1}(\widetilde{A}^2)<\nu_k(\widetilde{A}^2)/4$, and $\|\widetilde{A}^2-\widehat{A}^2\|_2\le\frac{1}{100}\cdot\nu_k(\widetilde{A}^2)$ we can apply Lemma 18 and we get

$$
\begin{aligned}
\left\|\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T-\widehat{W}_{[k]}\widehat{\Sigma}_{[k]}^{-2}\widehat{W}_{[k]}^T\right\|_2 &\le \frac{32\cdot\left(\left\|\widetilde{A}^2-\widehat{A}^2\right\|_2\right)^{1/3}}{\nu_k(\widetilde{A}^2)^2} \\
&\le O\left(\frac{\frac{\xi\cdot n^{-320\epsilon/\varphi^2}}{c_3{}^{1/3}}}{n^{-320\epsilon/\varphi^2}}\right) && \text{By (107) and (108)} \\
&\le \xi
\end{aligned}
$$

The last inequality holds by setting c_3 to a large enough constant to cancel the constant hidden in $O\left(\frac{\xi}{c_3{}^{1/3}}\right)$. $\square$

5.5 Proof of Theorem 2

Theorem 2. *[Spectral Dot Product Oracle] Let $\epsilon,\varphi\in(0,1)$ with $\epsilon\le\frac{\varphi^2}{10^5}$. Let $G=(V,E)$ be a d-regular graph that admits a (k,φ,ϵ)-clustering $C_1,\dots,C_k$. Let $1>\xi>\frac{1}{n^5}$. Then* InitializeOracle$(G,1/2,\xi)$ *(Algorithm 4) computes in time $O(k^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\cdot(\log n)^3\cdot\frac{1}{\varphi^2}/\xi^{18})$ a sublinear space data structure $\mathcal{D}$ of size $O(k^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\cdot(\log n)^3/\xi^{12})$ such that with probability at least $1-n^{-100}$ the following property is satisfied:*

For every pair of vertices $x, y \in V$, SpectralDotProduct$(G, x, y, 1/2, \xi, \mathcal{D})$ *(Algorithm 5) computes an output value $\langle f_x, f_y \rangle_{apx}$ such that with probability at least $1 - n^{-100}$*

$$\left| \langle f_x, f_y \rangle_{apx} - \langle f_x, f_y \rangle \right| \leq \frac{\xi}{n}.$$

The running time of SpectralDotProduct$(G, x, y, 1/2, \xi, \mathcal{D})$ *is* $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$.

Furthermore, for any $0 \leq \delta \leq 1/2$, one can obtain the following trade-offs between preprocessing time and query time: Algorithm SpectralDotProduct$(G, x, y, \delta, \xi, \mathcal{D})$ *requires* $O(k^{O(1)} \cdot n^{\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$ *per query when the prepressing time of Algorithm* InitializeOracle(G, δ, ξ) *is increased to* $O(k^{O(1)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^3 \cdot \frac{1}{\varphi^2}/\xi^{18})$.

To prove Theorem 2 we need to combine Lemma 19 from Section 5.3 with the following lemma.

Lemma 29. *Let $G = (V, E)$ be a d-regular and (k, φ, ϵ)-clusterable graph. Let $0 < \delta < 1/2$, and $1/n^5 < \xi < 1$. Let $\mathcal{D}$ denote the data structure constructed by Algorithm* InitializeOracle(G, δ, ξ) *(Algorithm 4). Let $x, y \in V$. Let $\langle f_x, f_y \rangle_{apx} \in \mathbb{R}$ denote the value returned by* SpectralDotProductOracle$(G, x, y, \delta, \xi, \mathcal{D})$ *(Algorithm 5). Let $t \geq \frac{20 \log n}{\varphi^2}$. Let $c > 1$ be a large enough constant and let $s \geq c \cdot n^{240 \cdot \epsilon/\varphi^2} \cdot \log n \cdot k^4$. Let $I_S = \{i_1, \ldots, i_s\}$ be a multiset of s indices chosen independently and uniformly at random from $\{1, \ldots, n\}$. Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Let M be the random walk transition matrix of G. Let $\sqrt{\frac{n}{s}} \cdot M^t S = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ be an SVD of $\sqrt{\frac{n}{s}} \cdot M^t S$ where $\widetilde{U} \in \mathbb{R}^{n \times n}, \widetilde{\Sigma} \in \mathbb{R}^{n \times n}, \widetilde{W} \in \mathbb{R}^{s \times n}$. If $\frac{\epsilon}{\varphi^2} \leq \frac{1}{10^5}$, and Algorithm 4 succeeds, then with probability at least $1 - n^{-100}$ matrix $\widetilde{\Sigma}_{[k]}^{-4}$ exists and we have*

$$\left| \langle f_x, f_y \rangle_{apx} - (M^t \mathbb{1}_x)^T (M^t S) \left(\frac{n}{s} \cdot \widetilde{W}_{[k]} \widetilde{\Sigma}_{[k]}^{-4} \widetilde{W}_{[k]}^T \right) (M^t S)^T (M^t \mathbb{1}_y) \right| < \frac{\xi}{n}.$$

Proof. Note that as per line 7 of Algorithm 5 $\langle f_x, f_y \rangle_{apx}$ is defined as

$$\langle f_x, f_y \rangle_{apx} = \alpha_x^T \Psi \alpha_y.$$

where as per line 3 of Algorithm 4 we define matrix $\Psi \in \mathbb{R}^{s \times s}$ as

$$\Psi = \frac{n}{s} \cdot \widehat{W}_{[k]} \widehat{\Sigma}_{[k]}^{-2} \widehat{W}_{[k]}^T,$$

and $\alpha_x, \alpha_y \in \mathbb{R}^s$ are vectors obtained by taking entrywise median over all $(\widehat{Q}_i)^T(\widehat{m}_x^i)$ and $(\widehat{Q}_i)^T(\widehat{m}_y^i)$. (See line 5 and 6 of Algorithm 5). For any vertex $a \in V$ recall that m_a denote $m_a = M^t \mathbb{1}_a$. We then define

$$\mathbf{a}_x = m_x^T (M^t S), \quad A = \frac{n}{s} \cdot \widetilde{W}_{[k]} \widetilde{\Sigma}_{[k]}^{-4} \widetilde{W}_{[k]}^T, \quad \mathbf{a}_y = (M^t S)^T m_y, \text{ and}$$

$$\mathbf{e}_x = \alpha_x^T - \mathbf{a}_x, \quad E = \Psi - A, \quad \mathbf{e}_y = \alpha_y - \mathbf{a}_y$$

Thus by triangle inequality we have

$$\begin{aligned}
&\left|\left| \alpha_x^T \Psi \alpha_y - m_x^T (M^t S) \left(\frac{n}{s} \cdot \widetilde{W}_{[k]} \widetilde{\Sigma}_{[k]}^{-4} \widetilde{W}_{[k]}^T \right) (M^t S)^T m_y \right|\right|_2 \\
&= \| (\mathbf{a}_x + \mathbf{e}_x)(A + E)(\mathbf{a}_y + \mathbf{e}_y) - \mathbf{a}_x A \mathbf{a}_y \|_2 \\
&\leq \|\mathbf{e}_x\|_2 \|A\|_2 \|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2 \|E\|_2 \|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2 \|A\|_2 \|\mathbf{e}_y\|_2 \\
&+ \|\mathbf{e}_x\|_2 \|E\|_2 \|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2 \|E\|_2 \|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2 \|A\|_2 \|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2 \|E\|_2 \|\mathbf{e}_y\|_2
\end{aligned}$$

Therefore we need to bound $\|\mathbf{e}_x\|_2$, $\|\mathbf{e}_y\|_2$, $\|E\|_2$, $\|\mathbf{a}_x\|_2$, $\|\mathbf{a}_y\|_2$ and $\|A\|_2$. Let $c' > 1$ be a constant we will define soon, and let $\xi' = \frac{\xi}{c' \cdot k^4 \cdot n^{80\epsilon/\varphi^2}}$. Let c_1 be a constant in front of s and let c_2 be a constant in front of R in Lemma 24. Thus for large enough c we have $s \geq c_1 \cdot n^{240\epsilon/\varphi^2} \cdot \log n \cdot k^4$ and $R_{\text{init}} = \Theta(n^{1-\delta+3 \cdot 10^3 \cdot \epsilon/\varphi^2} \cdot k^{33}/\xi^6) \geq \frac{c_2 \cdot k^9 \cdot n^{1/2+2100 \cdot \epsilon/\varphi^2}}{\xi'^6}$ as per line 2 of Algorithm 4, hence, by Lemma 24 applied with ξ' we have with probability at least $1 - n^{-100}$, $\widehat{W}_{[k]}^T$– and $\widehat{\Sigma}_{[k]}^{-4}$ exist and we have

$$\|E\|_2 = \frac{n}{s} \cdot \left|\left| \widehat{W}_{[k]} \widehat{\Sigma}_{[k]}^{-2} \widehat{W}_{[k]}^T - \widetilde{W}_{[k]} \widetilde{\Sigma}_{[k]}^{-4} \widetilde{W}_{[k]}^T \right|\right|_2 \leq \frac{n}{s} \cdot \xi' = \frac{\xi \cdot n}{c' \cdot k^4 \cdot n^{80\epsilon/\varphi^2} \cdot s}. \tag{110}$$

Recall that for a symmetric matrix A, we write $\nu_i(A)$ (resp. $\nu_{\max}(A), \nu_{\min}(A)$) to denote the i^{th} largest (resp. maximum, minimum) eigenvalue of A. We have

$$\|A\|_2 = \frac{n}{s} \cdot \|\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\|_2 = \frac{n}{s} \cdot \nu_{\max}\left(\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right) = \frac{n}{s} \cdot \frac{1}{\nu_k\left(\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{4}\widetilde{W}_{[k]}^T\right)}$$

Note that $\frac{n}{s} \cdot (M^tS)^T(M^tS) = \widetilde{W}\widetilde{\Sigma}^2\widetilde{W}^T$. Thus by Lemma 28 item (1) we have

$$\nu_k\left(\widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{4}\widetilde{W}_{[k]}^T\right) = \nu_k\left(\widetilde{W}\widetilde{\Sigma}^4\widetilde{W}^T\right) = \nu_k\left(\left(\frac{n}{s} \cdot (M^tS)^T(M^tS)\right)^2\right) \geq \frac{n^{-160\epsilon/\varphi^2}}{4}.$$

Therefore we have

$$\|A\|_2 \leq 4 \cdot \frac{n}{s} \cdot n^{160\epsilon/\varphi^2} = \frac{4 \cdot n^{1+160\epsilon/\varphi^2}}{s}. \tag{111}$$

Since G is (k, φ, ϵ)-clusterable by Lemma 22 for any vertex $x \in V$ we have

$$\|m_x\|_2^2 \leq O\left(k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right). \tag{112}$$

Then we get

$$\begin{aligned}
\|\mathbf{a}_x\|_2 &= \|(m_x)^T(M^tS)\|_2 \\
&= \sqrt{\sum_{a\in I_S}\left((m_x)^T(m_a)\right)^2} \\
&\leq \sqrt{\sum_{a\in I_S}\|m_x\|_2^2\|m_a\|_2^2} && \text{By Cauchy Schwarz} \\
&\leq O\left(\sqrt{s \cdot \left(k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right)^2}\right) && \text{By (112)} \\
&= O\left(\sqrt{s} \cdot k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right) && (113)
\end{aligned}$$

By the same analysis we get

$$\|\mathbf{a}_y\|_2 \leq O\left(\sqrt{s} \cdot k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}\right) \tag{114}$$

Now we left to bound $\|\mathbf{e}_x\|_2$ and $\|\mathbf{e}_y\|_2$. Recall that $\mathbf{e}_x = \alpha_x - (M^t\mathbb{1}_x)^T(M^tS)$ where $\alpha_x, \alpha_y \in \mathbb{R}^s$ are vectors obtained by taking entrywises median over all $(\widehat{Q}_i)^T(\widehat{m}_x^i)$ and $(\widehat{Q}_i)^T(\widehat{m}_y^i)$. (See line 5 and 6 of Algorithm 5). Also note that as per line 3 and line 4 of Algorithm 5, $\widehat{m}_x^i$ and $\widehat{m}_y^i$ are defined as the empirical probability distribution of running R_{query} random walks of length t starting from vertex x and y. Also note that $\widehat{Q}_i$s are generated by Algorithm 3 which runs R_{init} random walks from vertices in I_S. For any $z \in I_S$ any $i \in \{1, \ldots, O(\log n)\}$ let $\mathbf{q}_z^i$ denote the column corresponding to vertex z in $\widehat{Q}_i$.

Let c_3 be a constant in front of R_1 and R_2 in Lemma 26. Let $\sigma_{\text{err}} = \frac{\xi}{c' \cdot k^2 \cdot n^{(1+200\epsilon/\varphi^2)}}$. Thus by choice of $R_{\text{init}} = \Theta(n^{1-\delta+3\cdot 10^3\cdot\epsilon/\varphi^2} \cdot k^{33}/\xi^6)$ as per line 2 of Algorithm 4 and $R_{\text{query}} = \Theta(n^{\delta+500\cdot\epsilon/\varphi^2} \cdot k^9/\xi^2)$ as per line 1 of Algorithm 5, the prerequisites of Lemma 26 are satisfied:

$$\min(R_{\text{init}}, R_{\text{query}}) \geq \frac{c_3 \cdot k^5 \cdot n^{-2+(100\epsilon/\varphi^2)}}{\sigma_{\text{err}}^2}, \text{ and, } R_{\text{init}} \cdot R_{\text{query}} \geq \frac{c_3 \cdot k^2 \cdot n^{-1+(40\epsilon/\varphi^2)}}{\sigma_{\text{err}}^2}$$

Thus we can apply Lemma 26. Hence, for any $z \in I_S$ with probability at least 0.99 we have

$$\left|(\widehat{m}_x^i)^T\mathbf{q}_z^i - (m_x)^T(m_z)\right| \leq \sigma_{\text{err}}$$

Note that as per line 5 and line 6 of Algorithm 5 we take entrywise median over all $(\widehat{Q}_i)^T(\widehat{m}_x^i)$ and $(\widehat{Q}_i)^T(\widehat{m}_y^i)$. Since we are running $O(\log n)$ copies of the same algorithm with success probability at least 0.99, thus by simple Chernoff bound with probability at least $1 - n^{-100}$ for all $z \in I_S$ we have

$$\left|\alpha_x(z) - (m_x)^T(m_z)\right| \leq \sigma_{\text{err}}$$

Therefore by choice of $\sigma_{\mathrm{err}} = \frac{\xi}{c' \cdot k^2 \cdot n^{1+200\cdot\epsilon/\varphi^2}}$ we get

$$\|\mathbf{e}_x\|_2 = \|\alpha_x - (m_x)^T(M^tS)\|_2 \leq \sqrt{s} \cdot \sigma_{\mathrm{err}} = \frac{\sqrt{s}\cdot\xi}{c' \cdot k^2 \cdot n^{(1+200\epsilon/\varphi^2)}}. \tag{115}$$

By the same analysis we get

$$\|\mathbf{e}_y\|_2 \leq \frac{\sqrt{s}\cdot\xi}{c' \cdot k^2 \cdot n^{(1+200\epsilon/\varphi^2)}}. \tag{116}$$

Putting (110), (111), (112), (113), (114), (115), and (116) and for large enough n we get:

$$\begin{aligned}
&\left|\left|\langle f_x, f_y\rangle_{apx} - \cdot m_x^T(M^tS)\left(\frac{n}{s}\cdot \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(M^tS)^T m_y\right|\right| \leq \\
&\|\mathbf{e}_x\|_2\|A\|_2\|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2\|E\|_2\|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2\|A\|_2\|\mathbf{e}_y\|_2 + \\
&\|\mathbf{e}_x\|_2\|E\|_2\|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2\|E\|_2\|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2\|A\|_2\|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2\|E\|_2\|\mathbf{e}_y\|_2 \\
&\leq 2\cdot\left(\frac{\sqrt{s}\cdot\xi}{c'\cdot k^2\cdot n^{(1+200\epsilon/\varphi^2)}}\right)\left(\frac{4\cdot n^{1+160\epsilon/\varphi^2}}{s}\right)\cdot O\left(\sqrt{s}\cdot k^2\cdot n^{-1+(40\epsilon/\varphi^2)}\right) \\
&+ 2\cdot\left(\frac{\sqrt{s}\cdot n^{(80\epsilon/\varphi^2)}k^2}{n}\right)\left(\frac{\xi\cdot n}{c'\cdot k^4\cdot n^{80\epsilon/\varphi^2}\cdot s}\right)\left(\frac{\xi}{c'\cdot\sqrt{s}\cdot n^{(1+20\epsilon/\varphi^2)}}\right) \\
&+ \left(\frac{\sqrt{s}\cdot\xi}{c'\cdot k^2\cdot n^{(1+200\epsilon/\varphi^2)}}\right)^2\left(\frac{4\cdot n^{1+160\epsilon/\varphi^2}}{s}\right) \\
&+ O\left(\sqrt{s}\cdot k^2\cdot n^{-1+(40\epsilon/\varphi^2)}\right)^2\left(\frac{\xi\cdot n}{c'\cdot k^4\cdot n^{80\epsilon/\varphi^2}\cdot s}\right) \\
&+ \left(\frac{\sqrt{s}\cdot\xi}{c'\cdot k^2\cdot n^{(1+200\epsilon/\varphi^2)}}\right)^2\left(\frac{\xi\cdot n}{c'\cdot k^4\cdot n^{80\epsilon/\varphi^2}\cdot s}\right) \\
&\leq O\left(\frac{\xi}{c'\cdot n}\right) \\
&\leq \frac{\xi}{n}.
\end{aligned}$$

The last inequality holds by setting c' to a large enough constant to cancel the hidden constant of $O\left(\frac{\xi}{c'\cdot n}\right)$. □

Now we are able to complete the proof of Theorem 2.

Theorem 2. *[Spectral Dot Product Oracle] Let $\epsilon, \varphi \in (0,1)$ with $\epsilon \leq \frac{\varphi^2}{10^5}$. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $1 > \xi > \frac{1}{n^5}$. Then* InitializeOracle$(G, 1/2, \xi)$ *(Algorithm 4) computes in time $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^3 \cdot \frac{1}{\varphi^2}/\xi^{18})$ a sublinear space data structure $\mathcal{D}$ of size $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^3/\xi^{12})$ such that with probability at least $1 - n^{-100}$ the following property is satisfied:*

For every pair of vertices $x, y \in V$, SpectralDotProduct$(G, x, y, 1/2, \xi, \mathcal{D})$ *(Algorithm 5) computes an output value $\langle f_x, f_y\rangle_{apx}$ such that with probability at least $1 - n^{-100}$*

$$\left|\ \langle f_x, f_y\rangle_{apx} - \langle f_x, f_y\rangle\right| \leq \frac{\xi}{n}.$$

The running time of SpectralDotProduct$(G, x, y, 1/2, \xi, \mathcal{D})$ *is $O(k^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$.*

Furthermore, for any $0 \leq \delta \leq 1/2$, one can obtain the following trade-offs between preprocessing time and query time: Algorithm SpectralDotProduct$(G, x, y, \delta, \xi, \mathcal{D})$ *requires $O(k^{O(1)} \cdot n^{\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2}/\xi^{12})$ per query when the prepressing time of Algorithm* InitializeOracle(G, δ, ξ) *is increased to $O(k^{O(1)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)} \cdot (\log n)^3 \cdot \frac{1}{\varphi^2}/\xi^{18})$.*

Proof of Theorem 2. **Correctness:** Note that as per line 3 of Algorithm 4 we set $s = \Theta(n^{1500\cdot\epsilon/\varphi^2} \cdot \log n \cdot k^{16}/\xi^6)$. Recall that $I_S = \{i_1, \ldots, i_s\}$ is the multiset of s vertices each sampled uniformly at random (see line 4 of Algorithm 4). Let S be the $n \times s$ matrix whose j-th column equals $\mathbb{1}_{i_j}$. Recall that M is the random walk transition matrix of G. Let $\sqrt{\frac{n}{s}} \cdot M^t S = \widetilde{U}\widetilde{\Sigma}\widetilde{W}^T$ be the eigendecomposition of $\sqrt{\frac{n}{s}} \cdot M^t S$. We define

$$e_1 = \left|(M^t\mathbb{1}_x)^T(M^tS)\left(\frac{n}{s}\cdot \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(M^tS)^T(M^t\mathbb{1}_y) - \mathbb{1}_x^T U_{[k]}U_{[k]}^T\mathbb{1}_y\right|$$

and

$$e_2 = \left|\langle f_x, f_y\rangle_{apx} - (M^t\mathbb{1}_x)^T(M^tS)\left(\frac{n}{s}\cdot \widetilde{W}_{[k]}\widetilde{\Sigma}_{[k]}^{-4}\widetilde{W}_{[k]}^T\right)(M^tS)^T(M^t\mathbb{1}_y)\right|$$

By triangle inequality we have

$$\left|\langle f_x, f_y\rangle_{apx} - \langle f_x, f_y\rangle\right| = \left|\langle f_x, f_y\rangle_{apx} - \mathbb{1}_x^T U_{[k]}U_{[k]}^T\mathbb{1}_y\right| \le e_1 + e_2.$$

Let $\xi' = \xi/2$. Let c be a constant in front of s in Lemma 19 and c' be a constant in front of s in Lemma 29. Note that as per line 3 of Algorithm 4 we set $s = \Theta(n^{1500\cdot\epsilon/\varphi^2}\cdot \log n \cdot k^{16}/\xi^6)$. Since $\frac{\epsilon}{\varphi^2} \le \frac{1}{10^5}$ and $s \ge c\cdot n^{1500\epsilon/\varphi^2}\cdot \log n\cdot k^{16}/{\xi'}^6$ by Lemma 19 with probability at least $1-n^{-100}$ we have $e_1 \le \frac{\xi'}{n} = \frac{\xi}{2\cdot n}$. Since $s \ge c'\cdot n^{240\epsilon/\varphi^2}\cdot\log n\cdot k^4$, by Lemma 29 with probability at least $1-2\cdot n^{-100}$ we have $e_2 \le \frac{\xi}{2\cdot n}$. Thus with probability at least $1-3\cdot n^{-100}$ we have

$$\left|\langle f_x, f_y\rangle_{apx} - \langle f_x, f_y\rangle\right| \le e_1 + e_2 \le \frac{\xi}{2\cdot n} + \frac{\xi}{2\cdot n} \le \frac{\xi}{n}.$$

Space and runtime of InitializeOracle: Algorithm InitializeOracle(G,δ,ξ) (Algorithm 4) samples a set I_S. Then as per line 6 of Algorithm 4 it estimates the empirical probability distribution of random walks starting from any vertex $x \in I_S$ for $O(\log n)$ times. To that end as per line 2 of Algorithm 3 it runs R_{init} random walks of length t from each vertex $x \in I_S$. So it takes $O(\log n\cdot s\cdot R_{\text{init}}\cdot t)$ time and requires $O(\log n\cdot s\cdot R_{\text{init}})$ space to store endpoints of random walks. Then as per line 7 of Algorithm 4 it estimates matrix $\mathcal{G}$ such that the entry corresponding to the x^{th} row and y^{th} column of $\mathcal{G}$ is an estimation of pairwise collision probability of random walks starting from $x, y \in I_S$. To compute $\mathcal{G}$ we call Algorithm EstimateCollisionProbabilities$(G, I_S, R_{\text{init}}, t)$ (Algorithm 2) for $O(\log n)$ times. Algorithm 2 runs R_{init} random walks of length t from each vertex $x \in I_S$, hence, It takes $O(s\cdot R_{\text{init}}\cdot t\cdot\log n)$ time and it requires $O(s^2\cdot\log n)$ space to store matrix $\mathcal{G}$. Then as per line 8 of Algorithm 4 we compute the SVD of matrix $\mathcal{G}$ in time $O(s^3)$. Thus overall Algorithm 4 runs in time $O\left(\log n\cdot s\cdot R_{\text{init}}\cdot t + s^3\right)$. Thus, by choice of $t = \Theta\left(\frac{\log n}{\varphi^2}\right)$, $R_{\text{init}} = \Theta(n^{1-\delta+3\cdot10^3\cdot\epsilon/\varphi^2}\cdot k^{33}/\xi^6)$ and $s = \Theta(n^{1500\cdot\epsilon/\varphi^2}\cdot\log n\cdot k^{16}/\xi^6)$ as in Algorithm 4 we get that Algorithm 4 runs in time $O\left(\log n\cdot s\cdot R_{\text{init}}\cdot t + s^3\right) = O\left(k^{O(1)}\cdot(\frac{1}{\xi})^{18}\cdot n^{1-\delta+O(\epsilon/\varphi^2)}\cdot\log^3 n\cdot\frac{1}{\varphi^2}\right)$ and returns a data structure of size $O\left(s^2 + \log n\cdot s\cdot R_{\text{init}}\right) = O\left(k^{O(1)}\cdot(\frac{1}{\xi})^{12}\cdot n^{1-\delta+O(\epsilon/\varphi^2)}\cdot\log^2 n\right)$.

Space and runtime of SpectralDotProductOracle: Algorithm SpectralDotProductOracle$(G,x,y,\delta,\xi,\mathcal{D})$(Algorithm 5) repeats $O(\log n)$ copies of the following procedure: it runs R_{query} random walks of lenght t from vertex x and vertex y, then it computes $\widehat{m}_x\cdot\widehat{Q}_i$ and $\widehat{m}_y\cdot\widehat{Q}_i$. Since $\widehat{Q}_i \in \mathbb{R}^{n\times s}$ has s columns and since $\widehat{m}_x$ has at most R_{query} non-zero entries, thus one can compute $\widehat{m}_x\cdot\widehat{Q}_i$ in time $R_{\text{query}}\cdot s$. Finally Algorithm 5 take entrywises median of computed vectors (see line 5 and line 6 of Algorithm 5), and returns value $\alpha_x\Psi\alpha_y$ (see line 7 of Algorithm 5). Since $\alpha_x, \alpha_y \in \mathbb{R}^s$ and $\Psi \in \mathbb{R}^{s\times s}$ one can compute $\alpha_x\Psi\alpha_y$ in time $O(s^2)$. Thus overall Algorithm 5 takes $O\left(t\cdot R_{\text{query}}\cdot\log n + s\cdot R_{\text{query}}\cdot\log n + s^2\right)$ time and $O\left(R_{\text{query}}\cdot\log n + s\cdot R_{\text{query}}\cdot\log n + s^2\right)$ space. Thus, by choice of $t = \Theta\left(\frac{\log n}{\varphi^2}\right)$, $R_{\text{query}} = \Theta(n^{\delta+500\cdot\epsilon/\varphi^2}\cdot k^9/\xi^2)$ and $s = \Theta(n^{1500\cdot\epsilon/\varphi^2}\cdot\log n\cdot k^{16}/\xi^6)$ as in Algorithm 4 and Algorithm 5 we get that the Algorithm 5 runs in time $O\left(k^{O(1)}\cdot(\frac{1}{\xi})^{12}\cdot n^{\delta+O(\epsilon/\varphi^2)}\cdot\log^2 n\cdot\frac{1}{\varphi^2}\right)$ and returns a data structure of size $O\left(k^{O(1)}\cdot(\frac{1}{\xi})^{12}\cdot n^{\delta+O(\epsilon/\varphi^2)}\cdot\log^2 n\right)$.

□

5.6 Computing approximate norms and spectral dot products (Proof of Theorem 6)

To design the clustering algorithm in Section 6, since we cannot evaluate the dot-product of the spectral embedding exactly in sublinear time, we prove that it is enough to have access to approximate dot-product of the spectral embedding. In Algorithm 7, Algorithm 9 and throughout the analysis of in Section 6 we will use $\langle \cdot, \cdot \rangle_{apx}$ to denote approximate spectral dot products and $\|\cdot\|_{apx}$ to denote the approximate norm of a vector. Let $r \in [k]$ and $B, B_1, \ldots, B_r \subseteq V$. Let $\widehat{\mu}, \widehat{\mu}_1, \ldots, \widehat{\mu}_r \in \mathbb{R}^k$ where $\widehat{\mu} = \frac{\sum_{z \in B} f_z}{|B|}$ and $\widehat{\mu}_i = \frac{\sum_{z \in B_i} f_z}{|B_i|}$. All dot products we will try to approximate in Section 6 will be of the form $\left\langle f_x, \widehat{\Pi}(\widehat{\mu}) \right\rangle$ and all the norms that we approximate are of the form $\left\| \widehat{\Pi}(\widehat{\mu}) \right\|_{apx}$, where $x \in V$ and $\widehat{\Pi}$ is defined as a orthogonal projection onto $span(\{\widehat{\mu}_1, \ldots, \widehat{\mu}_r\})^{\perp}$. To compute such dot products we call Algorithm 6 in the following way (see Corollary 1):

$$\left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx} := \frac{1}{|B|} \cdot \sum_{y \in B} \left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx}, \tag{117}$$

$$\left\| \widehat{\Pi}\widehat{\mu} \right\|_{apx}^2 := \frac{1}{|B|} \cdot \sum_{x \in B} \left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx}. \tag{118}$$

Algorithm 6 DotProductOracleOnSubspace$(G, x, y, \delta, \xi, \mathcal{D}, B_1, \ldots, B_r)$ $\triangleright$ Need: $\epsilon/\varphi^2 \leq \frac{1}{10^5}$
$\triangleright$ $\mathcal{D} := \{\Psi, \widehat{Q}_1, \ldots, \widehat{Q}_{O(\log n)}\}$

1: Let $X \in \mathbb{R}^{r \times r}, h_x \in \mathbb{R}^r, h_y \in \mathbb{R}^r$.
2: Let $\xi' := \Theta(\xi \cdot n^{(-80\epsilon/\varphi^2)} \cdot k^{-6})$
3: **for** i, j in $[r]$ **do**
4: $\quad X(i, j) := \frac{1}{|B_i||B_j|} \cdot \sum_{z_i \in B_i} \sum_{z_j \in B_j}$ SpectralDotProduct$(G, z_i, z_j, \delta, \xi', \mathcal{D})$
5: $\quad$ $\triangleright$ $X(i, j) = \langle \widehat{\mu}_i, \widehat{\mu}_j \rangle_{apx}$
6: **for** i in $[r]$ **do**
7: $\quad h_x(i) := \frac{1}{|B_i|} \cdot \sum_{z_i \in B_i}$ SpectralDotProduct$(G, z_i, x, \delta, \xi', \mathcal{D})$ $\triangleright$ $h_x(i) = \langle \widehat{\mu}_i, f_x \rangle_{apx}$
8: $\quad h_y(i) := \frac{1}{|B_i|} \cdot \sum_{z_i \in B_i}$ SpectralDotProduct$(G, z_i, y, \delta, \xi', \mathcal{D})$ $\triangleright$ $h_y(i) = \langle \widehat{\mu}_i, f_y \rangle_{apx}$
9: **return** $\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} :=$ SpectralDotProduct$(G, x, y, \delta, \xi', \mathcal{D}) - h_x^T X^{-1} h_y$

The following Lemma is a generalization of Lemma 14 to the approximation of the cluster means (i.e, $\widehat{\mu}_1, \ldots, \widehat{\mu}_k$), where $\widehat{\mu}_i \in \mathbb{R}^k$ is a vector that approximates the center of cluster C_i (i.e., μ_i) such that $||\widehat{\mu}_i - \mu_i||_2$ is small.

Lemma 30. *Let $k \geq 2$ be an integer, $\varphi \in (0, 1)$, and $\epsilon \in (0, 1)$. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $\mu_1, \ldots, \mu_k$ denote the cluster means of $C_1, \ldots, C_k$. Let $0 < \zeta < \frac{\sqrt{\epsilon}}{20 \cdot k \cdot \varphi}$. Let $\widehat{\mu}_1, \ldots, \widehat{\mu}_k \in \mathbb{R}^k$ denote an approximation of the cluster means such that for each $i \in [k]$, $||\mu_i - \widehat{\mu}_i||_2 \leq \zeta ||\mu_i||_2$. Let $S \subseteq \{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$. Let $|S| = r$ and $\widehat{H} \in \mathbb{R}^{k \times r}$ denote a matrix whose columns are the vectors in S. Let $\sigma : [r] \to [k]$ denote a mapping from the the columns of H to the corresponding cluster. Let $\widehat{W} \in \mathbb{R}^{r \times r}$ be a diagonal matrix such that $\widehat{W}(i, i) = \sqrt{|C_{\sigma(i)}|}$. Let $\widehat{Z} = \widehat{H}\widehat{W}$. Then for any vector $x \in \mathbb{R}^r$ with $||x||_2 = 1$ we have*

1. $|x^T(\widehat{Z}^T\widehat{Z} - I)x| \leq \frac{5\sqrt{\epsilon}}{\varphi}$
2. $|x^T((\widehat{Z}^T\widehat{Z})^{-1} - I)x| \leq \frac{5\sqrt{\epsilon}}{\varphi}$.

Proof. **Proof of item** (1) Let $Y \in \mathbb{R}^{k \times k}$ be a matrix, whose i-th column is equal to $\sqrt{C_i} \cdot \mu_i$. By Lemma 9 item (2) for any vector $\alpha \in \mathbb{R}^k$ with $||\alpha||_2 = 1$ we have

$$|\alpha^T(Y^TY - I)\alpha| \leq \frac{4\sqrt{\epsilon}}{\varphi} \tag{119}$$

Let $\widehat{Y} \in \mathbb{R}^{k \times k}$ be a matrix, whose i-th column is equal to $\sqrt{C_i} \cdot \widehat{\mu}_i$. Note that for any $i, j \in [k]$ we have

$(Y^TY)(i,j) = \sqrt{|C_i||C_j|}\,\langle \mu_i, \mu_j\rangle$ and $(\widehat{Y}^T\widehat{Y})(i,j) = \sqrt{|C_i||C_j|}\,\langle \widehat{\mu}_i, \widehat{\mu}_j\rangle$. Therefore for any $i \in [k]$ we have

$$\begin{aligned}
\left|(Y^TY)(i,i) - (\widehat{Y}^T\widehat{Y})(i,i)\right| &= |C_i|\,\left|||\mu_i||_2^2 - ||\widehat{\mu}_i||_2^2\right| \\
&\le |C_i| \cdot |(||\mu_i||_2 - ||\widehat{\mu}_i||_2)(||\mu_i||_2 + ||\widehat{\mu}_i||_2)| \\
&\le |C_i| \cdot |(\zeta||\mu_i||_2)(||\mu_i||_2 + (1+\zeta)||\mu_i||_2)| && \text{Since } ||\widehat{\mu}_i||_2 \le (1+\zeta)||\mu_i||_2 \\
&\le 3\cdot\zeta|C_i|\cdot||\mu_i||_2^2 && \text{Since } \zeta < 1 \\
&\le 6\cdot\zeta && \text{By Lemma 7 } ||\mu_i||_2^2 \le \frac{2}{|C_i|}
\end{aligned}$$

Also for any $i \ne j \in [k]$ we have

$$\begin{aligned}
&\left|(Y^TY)(i,j) - (\widehat{Y}^T\widehat{Y})(i,j)\right| && (120)\\
&= \sqrt{|C_i||C_j|}\cdot|\langle\widehat{\mu}_i,\widehat{\mu}_j\rangle - \langle\mu_i,\mu_j\rangle| \\
&= \sqrt{|C_i||C_j|}\cdot|\langle\mu_i + (\widehat{\mu}_i - \mu_i), \mu_j + (\widehat{\mu}_j - \mu_j)\rangle - \langle\mu_i,\mu_j\rangle| \\
&\le \sqrt{|C_i||C_j|}\cdot(|\,\langle\widehat{\mu}_i - \mu_i, \widehat{\mu}_j - \mu_j\rangle\,| + |\,\langle\widehat{\mu}_i - \mu_i, \mu_j\rangle\,| + |\,\langle\widehat{\mu}_j - \mu_j, \mu_i\rangle\,|) && \text{By triangle inequality} \\
&\le \sqrt{|C_i||C_j|}\cdot(||\widehat{\mu}_i - \mu_i||_2||\widehat{\mu}_j - \mu_j||_2 + ||\widehat{\mu}_i - \mu_i||_2||\mu_j||_2 + ||\widehat{\mu}_j - \mu_j||_2||\mu_i||_2) && \text{By Cauchy-Schwarz} \\
&\le \sqrt{|C_i||C_j|}\cdot(\zeta^2 + 2\zeta)\,(||\mu_i||_2||\mu_j||_2) && \text{Since } ||\widehat{\mu}_i - \mu_i||_2 \le \zeta||\mu_i||_2 \text{ for all } i \\
&\le \sqrt{|C_i||C_j|}\cdot 6\cdot\zeta\cdot\frac{1}{\sqrt{|C_i||C_j|}} && \text{By Lemma 7 } ||\mu_i||_2^2 \le \frac{2}{|C_i|} \text{ for all } i \\
&\le 6\cdot\zeta && (121)
\end{aligned}$$

Therefore we have

$$\begin{aligned}
||(Y^TY) - (\widehat{Y}^T\widehat{Y})||_2 &\le ||(Y^TY) - (\widehat{Y}^T\widehat{Y})||_F \\
&\le \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}\Big((Y^TY)(i,j) - (\widehat{Y}^T\widehat{Y})(i,j)\Big)^2} \\
&\le 6\cdot k\cdot\zeta \\
&\le \frac{\sqrt{\epsilon}}{2\varphi} && \text{Since } \zeta \le \frac{\sqrt{\epsilon}}{20\cdot k\cdot\varphi}
\end{aligned}$$

Thus for any $\alpha \in \mathbb{R}^k$ with $||\alpha||_2 = 1$ we have

$$\left|\alpha^T\left((Y^TY) - (\widehat{Y}^T\widehat{Y})\right)\alpha\right| \le \frac{\sqrt{\epsilon}}{2\varphi} \tag{122}$$

Putting (122) and (119) together we get

$$\left|\alpha^T\left(\widehat{Y}^T\widehat{Y} - I\right)\alpha\right| \le 4.5\frac{\sqrt{\epsilon}}{\varphi}$$

Let $x \in \mathbb{R}^r$ be a vector with $||x||_2 = 1$, and let $\alpha \in \mathbb{R}^k$ be a vector that is $x_j = \alpha_j$ if $\widehat{\mu}_j \in S$ and otherwise $x_j = 0$. Thus we have $||\alpha||_2 = ||x||_2 = 1$ and $\widehat{Y}z = \widehat{Z}x$. Hence, we get

$$|x^T(\widehat{Z}^T\widehat{Z} - I)x| = |\alpha^T(\widehat{Y}^T\widehat{Y} - I)\alpha| \le \frac{4.5\sqrt{\epsilon}}{\varphi}$$

Proof of item (2) For any vector $x \in \mathbb{R}^r$ with $||x||_2 = 1$ we have

$$1 - \frac{4.5\sqrt{\epsilon}}{\varphi} \le x^T(\widehat{Z}^T\widehat{Z})x \le 1 + \frac{4.5\sqrt{\epsilon}}{\varphi} \tag{123}$$

Note that $\widehat{Z}^T\widehat{Z}$ is symmetric and positive semidefinit. Also note that $\widehat{Z}^T\widehat{Z}$ is spectrally close to I, hence, $\widehat{Z}^T\widehat{Z}$ is invertible. Thus by (123) and Lemma 13 for any vector $x \in \mathbb{R}^r$ we have

$$1 - \frac{5\sqrt{\epsilon}}{\varphi} \leq x^T(\widehat{Z}^T\widehat{Z})^{-1}x \leq 1 + \frac{5\sqrt{\epsilon}}{\varphi}$$

Therefore we get

$$|x^T((\widehat{Z}^T\widehat{Z})^{-1} - I)x| \leq \frac{5\sqrt{\epsilon}}{\varphi}.$$

□

Theorem 6. *Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \dots, C_k$. Let $k \geq 2$ be an integer, $\varphi \in (0, 1)$, $\frac{1}{n^5} < \xi < 1$, and $\frac{\epsilon}{\varphi^2}$ be smaller than a positive absolute constant. Then there exists an event $\mathcal{E}$ such that $\mathcal{E}$ happens with probability $1 - n^{-48}$ and conditioned on $\mathcal{E}$ the following holds.*

Let $r \in [k]$. Let $\delta \in (0, 1)$. Let $B_1, \dots, B_r$ denote multisets of points. Let $b = \max_{i \in r} |B_i|$. Let $\sigma : [r] \to [k]$ denote a mapping from the set B to the cluster $C = \sigma(B)$. Suppose that for all $i \in [r]$, $B_i \subseteq \sigma(B_i)$ and for all $i \neq j \in [r]$, $\sigma(B_i) \neq \sigma(B_j)$. Let $\widehat{\mu}_i = \frac{1}{|B_i|} \cdot \sum_{z \in B_i} f_z$. Suppose that for each $i \in [r]$, $||\widehat{\mu}_i - \mu_{\sigma(i)}||_2 \leq \frac{\sqrt{\epsilon}}{20 \cdot k \cdot \varphi} ||\mu_i||_2$. Let $\widehat{\Pi}$ is defined as a orthogonal projection onto then $span(\{\widehat{\mu}_1, \dots, \widehat{\mu}_r\})^\perp$. Then for all $x, y \in V$ we have

$$\left| \left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} f_y \right\rangle \right| \leq \frac{\xi}{n},$$

where $\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} :=$ DotProductOracleOnSubspace$(G, x, y, \delta, \xi, \mathcal{D}, B_1, \dots, B_r)$. *Algorithm 6 runs in time $O(b^2 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$.*

Proof. **Runtime:** Note that Algorithm 6, first computes matrix $X \in \mathbb{R}^{r \times r}$, and vectors $h_x, h_y \in \mathbb{R}^k$. To compute $X(i,j)$ for any $i, j \in [r]$, as per line 4 of Algorithm 6, we run SpectralDotProduct$(G, z_i, z_j, \delta, \xi', \mathcal{D})$ for all $z_i \in B_i$ and $z_j \in B_j$, where $|B_i| \leq b$ and $|B_j| \leq b$.

Note that by Theorem 2, Algorithm SpectralDotProduct$(G, z_i, z_j, \delta, \xi', \mathcal{D})$ runs in time $O(k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi'^{12})$. Thus one can compute the matrix X^{-1} in time $O(k^3 + k^2 \cdot b^2 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi'^{12})$. Also, to compute $h_x(i)$ (respectively, $h_y(i)$) for any $i \in [r]$, as per line 7 and line 8 of Algorithm 6, we run SpectralDotProduct$(G, x, z, \delta, \xi', \mathcal{D})$ for all $z \in B_i$ (respectively, $z \in B_j$). Thus one can compute h_x and h_y in time $k \cdot b \cdot (k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi} / \xi'^{12})$. As per line (2) of Algorithm 6 we set $\xi' := \Theta(\xi \cdot n^{(-80\epsilon/\varphi^2)} \cdot k^{-6})$. Therefore the runtime of the algoritm is $O(k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$.

Correctness: Let $x, y \in V$. Let $H \in \mathbb{R}^{k \times r}$ be a matrix whose columns are $\widehat{\mu}_1, \dots, \widehat{\mu}_r$. Then we have $H\left(H^T H\right)^{-1} H^T$ is the orthogonal projection matrix onto $span(\{\widehat{\mu}_1, \dots, \widehat{\mu}_r\})$. Let $W \in \mathbb{R}^{r \times r}$ denote a matrix such that for any $i \in [r]$, $W(i,i) = \sqrt{|C_{\sigma(i)}|}$. Note that

$$(HW)\left((HW)^T(HW)\right)^{-1}(HW)^T = HW\left(W^{-1}\left(H^T H\right)^{-1} W^{-1}\right) WH^T = H\left(H^T H\right)^{-1} H^T$$

Thus we have $(HW)\left((HW)^T(HW)\right)^{-1}(HW)^T$ is the orthogonal projection matrix onto $span(\{\widehat{\mu}_1, \dots, \widehat{\mu}_r\})$ and we get

$$\widehat{\Pi} = I - HW\left(WH^T HW\right)^{-1} WH^T$$

Therefore, we have

$$\left\langle f_x, \widehat{\Pi} f_y \right\rangle = \langle f_x, f_y \rangle - f_x^T HW\left(WH^T HW\right)^{-1} WH^T f_y \tag{124}$$

Let $\langle f_x, f_y \rangle_{apx} :=$ SpectralDotProduct$(G, x, y, \delta, \xi', \mathcal{D})$. Then as per line 9 of Algorithm 6 we have

$$\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} := \langle f_x, f_y \rangle_{apx} - h_x^T X^{-1} h_y, \tag{125}$$

where as per line (4) of Algorithm 6 for any $i, j \in [r]$ we have $X(i,j) = \langle \widehat{\mu}_i, \widehat{\mu}_j \rangle_{apx}$, and as per line (7) and line (8) of Algorithm 6 for any $i \in [r]$ we have $h_x(i) = \langle \widehat{\mu}_i, f_x \rangle_{apx}$ and $h_y(i) = \langle \widehat{\mu}_i, f_y \rangle_{apx}$. Note that

$$h_x^T X^{-1} h_y = h_x^T W W^{-1} X^{-1} W^{-1} W h_y = h_x^T W (WXW)^{-1} W h_y$$

Therefore by (125), (124) and triangle inequality we have

$$\left|\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} f_y \right\rangle\right| \leq |\left\langle f_x, f_y \right\rangle_{apx} - \left\langle f_x, f_y \right\rangle| + \left|h_x^T W (WXW)^{-1} W h_y - f_x^T HW \left(WH^T HW\right)^{-1} WH^T f_y\right|$$

Note that by Theorem 2 and by union bound over all pair of vertices with probability at least $1 - n^{-100} \cdot n^2$ for all $a, b \in V$ we have

$$|\left\langle f_a, f_b \right\rangle_{apx} - \left\langle f_a, f_b \right\rangle| \leq \frac{\xi'}{n} \tag{126}$$

We define

$$\mathbf{a}_x = f_x^T HW, \quad A = (WH^T HW)^{-1}, \quad \mathbf{a}_y = WH^T f_y, \text{ and}$$
$$\mathbf{e}_x = h_x^T W - \mathbf{a}_x, \quad E = (WXW)^{-1} - A, \quad \mathbf{e}_y = Wh_y - \mathbf{a}_y$$

Thus by triangle inequality we have

$$\begin{aligned}
&\left|h_x^T W (WXW)^{-1} W h_y - f_x^T HW (WH^T HW)^{-1} WH^T f_y\right| = \\
&\| \left(\mathbf{a}_x + \mathbf{e}_x\right) \left(A + E\right) \left(\mathbf{a}_y + \mathbf{e}_y\right) - \mathbf{a}_x A \mathbf{a}_y \|_2 \leq \\
&\|\mathbf{e}_x\|_2 \|A\|_2 \|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2 \|E\|_2 \|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2 \|A\|_2 \|\mathbf{e}_y\|_2 + \\
&\|\mathbf{e}_x\|_2 \|E\|_2 \|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2 \|E\|_2 \|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2 \|A\|_2 \|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2 \|E\|_2 \|\mathbf{e}_y\|_2
\end{aligned} \tag{127}$$

Thus we need to bound $||\mathbf{a}_x||_2, ||\mathbf{a}_y||_2, ||\mathbf{e}_x||_2, ||\mathbf{e}_y||_2, ||A||_2, ||E||_2$. Note that $||\mathbf{a}_x||_2 = ||f_x^T HW||_2$, Thus we have $||\mathbf{a}_x||_2 \leq ||f_x^T H||_2 ||W||_2$. Note that

$$||W||_2 \leq \max_i W(i,i) = \max_i \sqrt{|C_i|} \leq \sqrt{n} \tag{128}$$

Then we bound $||f_x^T H||_2$. Note that $||f_x^T H||_2 = \sqrt{\sum_{i=1}^r \left\langle f_x, \widehat{\mu}_i \right\rangle^2}$. We first bound $\langle f_x, \widehat{\mu}_i \rangle$.

$$\begin{aligned}
\langle f_x, \widehat{\mu}_i \rangle &= \frac{1}{|B_i|} \cdot \sum_{z \in B_i} \langle f_x, f_z \rangle \\
&\leq \frac{1}{|B_i|} \sum_{z \in B_i} ||f_x||_2 ||f_z||_2 \\
&\leq \frac{1}{|B_i|} \cdot \sum_{z \in B_i} \sqrt{k^2 \cdot ||f_x||_\infty^2 ||f_z||_\infty^2} \\
&\leq \frac{1}{|B_i|} \cdot |B_i| \cdot k \cdot O\left(\frac{k \cdot n^{40\epsilon/\varphi^2}}{n}\right) \qquad \text{By Lemma 5 and since } \min_{i \in k} |C_i| \geq \Omega\left(\frac{n}{k}\right) \\
&\leq O(k^2 \cdot n^{-1+40\epsilon/\varphi^2})
\end{aligned}$$

Since, $r < k$, we get

$$||f_x^T H||_2 = \sqrt{\sum_{i=1}^r \left\langle f_x, \widehat{\mu}_i \right\rangle^2} \leq \sqrt{k} \cdot O(k^2 \cdot n^{-1+40\epsilon/\varphi^2}) \leq O(k^{2.5} \cdot n^{-1+40\epsilon/\varphi^2}) \tag{129}$$

Thus we get

$$||\mathbf{a}_x||_2 = ||f_x^T HW||_2 \leq ||f_x^T H||_2 ||W||_2 \leq O\left(k^{2.5} \cdot n^{-1/2+40\epsilon/\varphi^2}\right) \tag{130}$$

By the same computation we also have

$$||\mathbf{a}_y||_2 \leq O\left(k^{2.5} \cdot n^{-1/2+40\epsilon/\varphi^2}\right) \tag{131}$$

Next we bound $||\mathbf{e}_x||_2$. We have $\mathbf{e}_x = h_x^T W - f_x^T HW$. Thus we get $||\mathbf{e}_x||_2 \leq ||h_x^T - f_x^T H||_2 ||W||_2$. By (128) we have a bound on $||W||_2$. Note that for any $i \in r$, we have $h_x(i) = \frac{1}{|B_i|} \sum_{z \in B_i} \langle f_x, f_z \rangle_{apx}$ and

$(f_x^T H)(i) = \frac{1}{|B_i|}\sum_{z\in B_i} \langle f_x, f_z\rangle$. Therefore with probability at least $1-n^{-98}$ we have

$$\begin{aligned}
|h_x(i) - (f_x^T H)(i)| &= \left|\frac{1}{b}\sum_{z\in B_i}(\langle f_x, f_z\rangle_{apx} - \langle f_x, f_z\rangle)\right| \\
&\le \frac{1}{|B_i|}\sum_{z\in B_i} |\langle f_x, f_z\rangle_{apx} - \langle f_x, f_z\rangle| && \text{By triangle inequality} \\
&\le \frac{1}{|B_i|}\cdot |B_i| \cdot \frac{\xi'}{n} && \text{By (126)}
\end{aligned}$$

Since $r\le k$, we have

$$||h_x^T - f_x^T H||_2 = \sqrt{\sum_{i=1}^r (h_x(i) - \mathbf{a}_x(i))^2} \le \sqrt{k}\cdot\frac{\xi'}{n}$$

Therefore by (128) we have

$$||\mathbf{e}_x||_2 \le ||h_x^T - f_x^T H||_2 ||W||_2 \le \frac{\xi'\sqrt{k}}{\sqrt{n}} \tag{132}$$

By the same computation we also have

$$||\mathbf{e}_y||_2 \le \frac{\xi'\sqrt{k}}{\sqrt{n}} \tag{133}$$

Next we bound $||A||_2$. Note that $A = ((HW)^T(HW))^{-1}$. By Lemma 30 item (2) for any vector $x\in\mathbb{R}^r$ with $||x||_2 = 1$ we have

$$\left|x^T\left(\left((HW)^T(HW)\right)^{-1} - I\right)x\right| \le \frac{5\sqrt{\epsilon}}{\varphi}$$

Therefore

$$||A||_2 = ||((HW)^T(HW))^{-1}||_2 \le 1 + \frac{5\sqrt{\epsilon}}{\varphi} \le 2 \tag{134}$$

Now we bound $||E||_2 = ||(WXW)^{-1} - (WH^THW)^{-1}||_2$. For any $i,j\in[r]$ we have

$$(WXW)(i,j) = \sqrt{|C_{\sigma(B_i)}||C_{\sigma(B_j)}|}\cdot\frac{1}{|B_i|\cdot|B_j|}\cdot\sum_{z_i\in B_i, z_j\in B_j}\left\langle f_{z_i}, f_{z_j}\right\rangle_{apx}$$

and

$$(WH^THW)(i,j) = \sqrt{|C_{\sigma(B_i)}||C_{\sigma(B_j)}|}\cdot\frac{1}{|B_i|\cdot|B_j|}\cdot\sum_{z_i\in B_i, z_j\in B_j}\left\langle f_{z_i}, f_{z_j}\right\rangle$$

Therefore with probability at least $1-n^{-98}$ we have

$$\begin{aligned}
&|(WXW)(i,j) - (WH^THW)(i,j)| \\
&= \left|\sqrt{|C_{\sigma(B_i)}||C_{\sigma(B_j)}|}\cdot\frac{1}{|B_i|\cdot|B_j|}\sum_{z_i\in B_i, z_j\in B_j}(\hat{f}_{z_iz_j} - \left\langle f_{z_i}, f_{z_j}\right\rangle)\right| \\
&\le \sqrt{|C_{\sigma(B_i)}||C_{\sigma(B_j)}|}\cdot\frac{1}{|B_i|\cdot|B_j|}\sum_{z_i\in B_i, z_j\in B_j}|\hat{f}_{z_iz_j} - \left\langle f_{z_i}, f_{z_j}\right\rangle| && \text{By triangle inequality} \\
&\le n\cdot\frac{1}{|B_i|\cdot|B_j|}\cdot|B_i|\cdot|B_j|\cdot\frac{\xi'}{n} && \text{By (126) and since } |C|\le n \qquad (135)
\end{aligned}$$

Since $r\le k$ and by (135) we get

$$\begin{aligned}
\left|||WXW - WH^THW||_2\right| &\le ||WXW - WH^THW||_F \\
&\le \sqrt{\sum_{i=1}^r\sum_{j=1}^r ((WXW)(i,j) - (WH^THW)(i,j))^2} \\
&\le k\cdot\xi'
\end{aligned}$$

Thus for any vector $x \in \mathbb{R}^r$ with $||x||_2 = 1$ we have

$$x^T(WH^THW)x - k \cdot \xi' \leq x^T(WXW)x \leq x^T(WH^THW)x + k \cdot \xi' \tag{136}$$

By Lemma 30 item (1) for any vector $x \in \mathbb{R}^r$ with $||x||_2 = 1$ we have

$$|x^T\left((HW)^T(HW) - I\right)x| \leq \frac{5\sqrt{\epsilon}}{\varphi}$$

Hence we have

$$x^T(HW)^T(HW)x \geq 1 - \frac{5\sqrt{\epsilon}}{\varphi} \geq \frac{1}{2} \tag{137}$$

Therfore by (136) and (137) we get for any vector $x \in \mathbb{R}^r$ with $||x||_2 = 1$ we have

$$(1 - 2 \cdot k \cdot \xi') \cdot x^T(WH^THW)x \leq x^T(WXW)x \leq (1 + 2 \cdot k \cdot \xi') \cdot x^T(WH^THW)x \tag{138}$$

Note that WH^THW is a symmetric matrix. Also note that by definition of X in line 4 of Algorithm 6, X is a symmetric matrix, hence, WXW is symmetric and positive semidefinit. Also note that WXW is spectrally close to WH^THW and I, hence, WXW is invertible. Thus by (138) and Lemma 13 we have

$$(1 - 4 \cdot k \cdot \xi') \cdot x^T(WH^THW)^{-1}x \leq x^T(WXW)^{-1}x \leq (1 + 4 \cdot k \cdot \xi') \cdot x^T(WH^THW)^{-1}x$$

Therefore by (134) we have

$$||E||_2 = ||(WH^THW)^{-1} - (WXW)^{-1}||_2 \leq 4 \cdot k \cdot \xi' \cdot ||(WH^THW)^{-1}||_2 = 8 \cdot k \cdot \xi' \tag{139}$$

Putting (139), (134), (132), (133), (130), (131) and(127) together, with probability at least $1 - n^{-50}$ we have

$$\begin{aligned}
&\left|h_x^T W(WXW)^{-1}Wh_y - f_x^T HW(WH^THW)^{-1}WH^Tf_y\right| = \\
&\|\left(\mathbf{a}_x + \mathbf{e}_x\right)(A+E)\left(\mathbf{a}_y + \mathbf{e}_y\right) - \mathbf{a}_x A \mathbf{a}_y\|_2 \leq \\
&\|\mathbf{e}_x\|_2\|A\|_2\|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2\|E\|_2\|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2\|A\|_2\|\mathbf{e}_y\|_2 + \\
&\|\mathbf{e}_x\|_2\|E\|_2\|\mathbf{a}_y\|_2 + \|\mathbf{a}_x\|_2\|E\|_2\|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2\|A\|_2\|\mathbf{e}_y\|_2 + \|\mathbf{e}_x\|_2\|E\|_2\|\mathbf{e}_y\|_2 \\
&\leq O\left(\xi' \cdot \frac{\sqrt{k}}{\sqrt{n}} \cdot k^{2.5} \cdot n^{-1/2+40\epsilon/\varphi^2}\right) + O\left(k \cdot \xi' \cdot k^5 \cdot n^{-1+80\epsilon/\varphi^2}\right) \\
&+ O\left(\xi' \cdot \frac{\sqrt{k}}{\sqrt{n}} \cdot k \cdot \xi' \cdot k^{2.5} \cdot n^{-1/2+40\epsilon/\varphi^2}\right) + O\left(\xi'^2 \cdot \frac{k}{n}\right) + O\left(\xi'^2 \cdot \frac{k}{n} \cdot k \cdot \xi'\right) \\
&\leq O\left(\frac{\xi' \cdot k^6 \cdot n^{80\epsilon/\varphi^2}}{n}\right) \\
&\leq \frac{1}{2} \cdot \frac{\xi}{n}
\end{aligned} \tag{140}$$

The last inequality holds by setting $\xi' = \frac{\xi \cdot n^{(-80\epsilon/\varphi^2)} \cdot k^{-6}}{c}$ as per line of Algorithm 6 where c is a large enough constant to cancel the constant hidden in $O\left(\frac{\xi' \cdot k^6 \cdot n^{80\epsilon/\varphi^2}}{n}\right)$.

Therefore with probability at least $1 - n^{-98} \geq 1 - n^{-50}$ we have

$$\begin{aligned}
&\left|\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} f_y \right\rangle\right| \\
&\leq |\widehat{f}_{xy} - \langle f_x, f_y \rangle| + \left|h_x^T W(WXW)^{-1}Wh_y - f_x^T HW\left(WH^THW\right)^{-1}WH^Tf_y\right| \\
&\leq \frac{\xi'}{n} + \frac{1}{2} \cdot \frac{\xi}{n} \\
&\leq \frac{\xi}{n} \qquad \text{By (126), (140), and since } \xi' < \xi/2
\end{aligned} \tag{141}$$

Now let $\mathcal{E}$ be the event that for all $x, y \in V$ we have $|\left\langle f_x, \widehat{\Pi} f_y \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} f_y \right\rangle| \leq \frac{\xi}{n}$. Then by (141) and the union bound we get that $\mathcal{E}$ happens with probability at least $1 - n^{-48}$ and it is the claimed high probability event from the statement.

□

Corollary 1. *Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let $k \geq 2$ be an integer, $\varphi \in (0, 1)$, $\delta \in (0, 1)$, $\frac{1}{n^5} < \xi < 1$, $\frac{\epsilon}{\varphi^2}$ be smaller than a positive absolute constant. Let $\mathcal{E}$ be the event that happens with probability $1 - n^{-48}$ that is guaranteed by Theorem 6. Then conditioned on $\mathcal{E}$ the following conditions hold.*

Let $r \in [k]$. Let $B_1, \ldots, B_r, B'$ denote multisets of points. Let $b = \max\{|B_1|, \ldots, |B_r|, |B'|\}$. Let $\sigma : [r] \to [k]$ denote a mapping from the set B to the cluster $C = \sigma(B)$. Suppose that for all $i \in [r]$, $B_i \subseteq \sigma(B_i)$ and for all $i \neq j \in [r]$, $\sigma(B_i) \neq \sigma(B_j)$. Let $\widehat{\mu}_i = \frac{1}{|B_i|} \cdot \sum_{z \in B} f_z$ for all $i \in [r]$, and let $\widehat{\mu} = \frac{1}{|B'|} \cdot \sum_{z \in B_i} f_z$. Suppose that for each $i \in [r]$, $||\widehat{\mu}_i - \mu_{\sigma(i)}||_2 \leq \frac{\sqrt{\epsilon}}{20 \cdot k \cdot \varphi} ||\mu_i||_2$. Let $\widehat{\Pi}$ is defined as a orthogonal projection onto then span$(\{\widehat{\mu}_1, \ldots, \widehat{\mu}_r\})^\perp$. Then the following hold:

1. *There exits an algorithm that runs in time $O(b^3 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$ and for any $x \in V$ returns a value $\left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx}$ such that*
$$\left| \left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle \right| \leq \frac{\xi}{n}.$$

2. *There exits an algorithm that runs in time $O(b^4 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$ and returns a value $\left\| \widehat{\Pi} \widehat{\mu} \right\|_{apx}^2$ such that $\left| \left\| \widehat{\Pi} \widehat{\mu} \right\|_{apx}^2 - ||\widehat{\Pi} \widehat{\mu}||_2^2 \right| \leq \frac{\xi}{n}$.*

Proof. **Proof of 1:** To compute $\left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx}$ we call Algorithm 6, b times in the following way:

$$\left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx} := \frac{1}{|B|} \cdot \sum_{y \in B} \textsc{DotProductOracleOnSubspace}(G, x, y, \delta, \mathcal{D}, \xi, B_1, \ldots, B_r) \tag{142}$$

The runtime of Algorithm 6 is $O(b^2 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$, thus the runtime of computation of $\left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx}$ is $O(b^3 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$. Moreover by Theorem 6 and the assumption that $\mathcal{E}$ holds we have

$$\begin{aligned}
\left| \left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle \right| &= \left| \frac{1}{|B'|} \sum_{y \in B'} \left\langle f_x, \widehat{\Pi} y \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle \right| \\
&\leq \frac{1}{|B'|} \sum_{y \in B'} \left| \left\langle f_x, \widehat{\Pi} y \right\rangle_{apx} - \left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle \right| && \text{By triangle inequality} \\
&\leq \frac{1}{|B'|} \cdot |B'| \cdot \frac{\xi}{n} && \text{By Theorem 6} \\
&\leq \frac{\xi}{n}
\end{aligned}$$

Proof of 2: To compute $\left\| \widehat{\Pi} \widehat{\mu} \right\|_{apx}^2$ we call the procedure from item (1) b times in the following way:

$$\left\| \widehat{\Pi} \widehat{\mu} \right\|_{apx}^2 := \frac{1}{|B|} \cdot \sum_{x \in B} \left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx}. \tag{143}$$

The runtime of the procedure from item (1) is $O(b^3 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$, thus the runtime of computation of $\left\langle f_x, \widehat{\Pi} \widehat{\mu} \right\rangle_{apx}$ is $O(b^4 \cdot k^{O(1)} \cdot n^{\delta + O(\epsilon/\varphi^2)} \cdot (\log n)^2 \cdot \frac{1}{\varphi^2} / \xi^{12})$. Moreover by item (1) we have

$$
\begin{aligned}
\left| \left\| \widehat{\Pi}\widehat{\mu} \right\|_{apx}^{2} - ||\widehat{\Pi}\widehat{\mu}||_2^2 \right| &= \left| \left\langle \widehat{\mu}, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx} - \left\langle \widehat{\mu}, \widehat{\Pi}\widehat{\mu} \right\rangle \right| \\
&= \left| \frac{1}{|B'|} \cdot \sum_{x \in B'} \left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx} - \frac{1}{|B'|} \cdot \sum_{x \in B'} \left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle \right| \\
&\leq \frac{1}{|B'|} \cdot \sum_{x \in B'} \left| \left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx} - \sum_{x \in B'} \left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle \right| && \text{By triangle inequality} \\
&\leq \frac{1}{|B'|} \cdot |B'| \cdot \frac{\xi}{n} && \text{By item (1)} \\
&\leq \frac{\xi}{n}.
\end{aligned}
$$

□

6 The main algorithm and its analysis

In this section we show that, by having access to approximate spectral dot-products for a (k, φ, ϵ)-clusterable graph G, we can assign each vertex in G to a cluster in sublinear time so that the resulting collection of clusters is, with high probability, a good approximation of a (k, φ, ϵ)-clustering of G. In particular, we can show that the fraction of wrong assignments per cluster is at most $C \cdot \frac{\epsilon}{\varphi^3} \cdot \log(k)$, for some constant $C > 0$. In the next subsection we describe our algorithm then in the remaining part of the section we present its analysis.

6.1 The Algorithm (Partitioning Scheme, Algorithm 7)

We first present an idealized version of the sublinear clustering scheme defined by Algorithm 7 and Algorithm 10. In this section to simplify presentation we assume φ to be constant.

The algorithm can be thought of as consisting of 3 parts. The first part, described in paragraph **Idealized Clustering Algorithm**, is a procedure that explicitly, in iterative fashion, produces a k-clustering of G. More precisely it recovers clusters in $O(\log(k))$ stages, where for every i after the i-th stage at most $k/2^i$ clusters are left unrecovered. The algorithm can be thought of as a version of carving of halfspaces in $\mathbb{R}^k$ and it relies on the knowledge of cluster means $\mu_1, \ldots, \mu_k$ (recall that $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} f_x$). That is why in paragraph **Finding approximate centers** we show how to compute approximations of μ_i's. To find good approximation to μ_i's we need to test many candidate sets $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$, which also means considering many candidate clusterings. This is a problem as we want our procedure to run in sublinear time but the idealized partitioning algorithm constructs clusterings explicitly! To solve this we explain in paragraph **Verifying a clustering** how to emulate the partitioning algorithm to test that, for a set of $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$, it indeed induces a good clustering.

Idealized Clustering Algorithm. Assume that the we have access to cluster means $\{\mu_1, \ldots, \mu_k\}$ and dot product evaluations. The algorithm proceeds in $O(\log(k))$ stages, in the first stage it considers k candidate sets $\widehat{C}_i$, where $x \in \widehat{C}_i$ iff f_x has big correlation with μ_i but small correlation with all other μ_j's. More precisely $x \in \widehat{C}_i$ iff:

$$\langle f_x, \mu_i \rangle \geq 0.93||\mu_i||^2 \text{ and for all } j \neq i \ \langle f_x, \mu_j \rangle < 0.93||\mu_j||^2.$$

Note that by definition all these clusters are disjoint. Moreover we are able to show (see Lemma 37) that at least $k/2$ out of $\widehat{C}_i$'s are good approximate clusters, that is for each one of them there exists j such that $|\widehat{C}_i \triangle C_j| \leq O(\epsilon) \cdot |C_j|$. At this point we return these good clusters, remove the corresponding vertices from the graph, remove the corresponding μ's from the set $\{\mu_1, \ldots, \mu_k\}$ of still alive centers and proceed to the next stage.

In the next stage we restrict our attention to a lower dimensional subspace Π of $\mathbb{R}^k$. Intuitively we want to project out all the directions corresponding to the removed cluster centers. Recall that μ_i's are close to being orthogonal (see Lemma 12 and 7) so projecting the returned directions out is almost equivalent to considering the subspace $\Pi := \text{span}(\{\mu_1, \ldots, \mu_b\})$, where $\{\mu_1, \ldots, \mu_b\}$ is the set of still alive μ's. Now the algorithm considers b candidate clusters where the condition for x being in a cluster i changes to:

$$\langle f_x, \Pi\mu_i \rangle \geq 0.93||\Pi\mu_i||^2 \text{ and for all } j \in [b], j \neq i \ \langle f_x, \Pi\mu_j \rangle < 0.93||\Pi\mu_j||^2.$$

We are still able to show (also Lemma 37) that at least $b/2$ out of them are good approximate clusters. That is for each i there exists j such that $|\widehat{C}_i \triangle C_j| \leq O(\epsilon) \cdot |C_j|$ but this time the constant hidden in the O notation is bigger than in the first stage. In general at any stage t the bound degrades to $O(\epsilon \cdot t)$. At the end of the stage we proceed in a similar fashion by returning the clusters, removing the corresponding vertices and μ's and considering a lower dimensional subspace of Π in the next stage.

The algorithm continues in such a fashion for $O(\log(k))$ steps, as we guarantee that in each stage at least half of the remaining cluster means is removed. Thus the final guarantee is: there exists a permutation π on k elements such that for every i:

$$|\widehat{C}_{\pi(i)} \triangle C_i| \leq O\left(\epsilon \log(k)\right) \cdot |C_i|.$$

The decreasing (in the inclusion sense) sequence of subspaces $(\Pi_1, \ldots, \Pi_{\log(k)})$ corresponds to the subspaces constructed in Algorithm 7, while this offline algorithm as a whole corresponds to the sublinear Algorithm 10 that implicitly tries to construct a sequence of subspaces that (with respect to Algorithm 7) defines a good clustering.

Finding approximate centers. Note that cluster means are defined by the clustering, so it may seem that finding approximate means is a difficult operation. However, there is a relatively simple solution to this. In Algorithm 10 we find approximate cluster means by sampling $O(\frac{\varphi^2}{\epsilon}k^4\log(k))$ points, guessing cluster memberships and considering the means of the samples as cluster centers. We use that the mean of a random sample of a cluster is typically close to the true mean of its cluster and so our sample means will provide a good estimation of the true means. We also remark that sampling a single vertex from each cluster does not seem to provide a sufficiently good estimate, i.e. we require to take the mean of a sample *set*.

Verifying a clustering. We also need a procedure that given an implicit sequence of subspaces $(\Pi_1, \dots, \Pi_{\log(k)})$ checks whether they indeed define (via Algorithm 7) a good clustering. In fact, for every guess of cluster centers and the corresponding (as implicitly created by Algorithm 10) sequence of Π's we need to be able to check efficiently if the resulting clustering is a good approximation of a (k, φ, ϵ)-clustering. Since we would like to do this in sublinear time as well, we need to do this verification by random sampling. Then we design a procedure that consists of two steps. In a first step, we check if the cluster sizes are not too small. This is only a technical step, which is needed to make sure that the later steps work. The main step is to test whether every cluster has small outer conductance (Algorithm 11). In order to do so, we sample vertices uniformly at random and check whether they are contained in the cluster that is currently checked. If this is the case, we sample a random edge incident to the sample vertex. This way, we obtain a random edge incident to a random vertex from the current cluster (this follows since the conditional distribution is uniform over the cluster). We use standard concentration bounds to prove that we get a good approximation.

In the partitioning scheme and in the analysis a useful definition are subsets of vertices called threshold sets. A threshold set of a point y is the set of vertices with dot products (or approximate dot product) with y being above a specific threshold, more formally:

Definition 8 (**Threshold sets**)**.** Let $G = (V, E)$ be a (k, φ, ϵ)-clusterable graph (*as in Definition 2*). Recall that $f_x = F\mathbb{1}_x$. For $y \in \mathbb{R}^k, \theta \in \mathbb{R}^+$ we define:

$$C_{y,\theta} := \{x \in V : \langle f_x, y\rangle \geq \theta ||y||^2\}$$

Definition 9 (**Approximate threshold sets**)**.** Let $G = (V, E)$ be a (k, φ, ϵ)-clusterable graph (*as in Definition 2*). Recall that $f_x = F\mathbb{1}_x$. For $\theta \in \mathbb{R}^+$ and $y \in \mathbb{R}^k$ such that $y = \widehat{\Pi}(\widehat{\mu})$, where $\widehat{\Pi}$ is the orthogonal projection onto $span(\{\widehat{\mu}_1, \dots, \widehat{\mu}_b\})^\perp$ and each $\widehat{\mu}, \widehat{\mu}_1, \dots, \widehat{\mu}_b$ is an average of a set of embedded vertices:

$$C^{apx}_{y,\theta} := \{x \in V : \langle f_x, y\rangle_{apx} \geq \theta\, \|y\|^2_{apx}\}. \tag{144}$$

Recall that a discussion of how $\langle\cdot,\cdot\rangle_{apx}$ and $\|\cdot\|_{apx}$ are computed is presented in Section 5.6.

Algorithm 7 HyperplanePartitioning$(x, (T_1, T_2, \dots, T_b))$
▷ T_i's are sets of $\widehat{\mu}_j$ where $\widehat{\mu}_j$'s are given as sets of points
▷ see Section 5.6 for the reason of such representation

```
for i = 1 to b do
    Let Π be the projection onto the span(⋃_{j<i} T_j)^⊥.
    Let S_i = ⋃_{j≥i} T_j
    for μ̂ ∈ T_i do
        if x ∈ C^{apx}_{Πμ̂,0.93} \ ⋃_{μ̂'∈S_i\{μ̂}} C^{apx}_{Πμ̂',0.93} then    ▷ see (144) for definition of C^{apx}_{y,θ}
            return μ̂
```

HyperplanePartitioning is the algorithm that, after preprocessing, is used to assign vertices to clusters. In the preprocessing step (see ComputeOrderedPartition in Section 6.3) an ordered partition $(T_1, \dots, T_b)$ of approximate cluster means $\{\widehat{\mu}_1, \dots, \widehat{\mu}_k\}$ is computed. HyperplanePartitioning invoked with this ordered partition as a parameter induces a collection of clusters as follows:

Definition 10 (**Implicit clustering**)**.** For an ordered partition $(T_1, \dots, T_b)$ of approximate cluster means $\{\widehat{\mu}_1, \dots, \widehat{\mu}_k\}$ we say that $(T_1, \dots, T_b)$ **induces a collection of clusters** $\{\widehat{C}_{\widehat{\mu}_1}, \dots, \widehat{C}_{\widehat{\mu}_k}\}$ if for all $i \in [k]$:

$$\widehat{C}_{\widehat{\mu}_i} = \{x \in V : \textsc{HyperplanePartitioning}(x, (T_1, \dots, T_b)) = \widehat{\mu}_i\}.$$

Remark 6. *Ordered partition $(T_1, \ldots, T_b)$, precomputed in the preprocessing step (assuming access to $\{\mu_1, \ldots, \mu_k\}$), will correspond to the **Idealized Clustering Algorithm** in the following sense. Number of sets in the partition (i.e. b) corresponds to the number of stages of **Idealized Clustering Algorithm** and for every $i \in [b]$ T_i contains exactly the μ's returned in stage i.*

In the rest of this section we explain how to compute an ordered partition $(T_1, \ldots, T_b)$ of a set of approximate centers $(\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_k)$ such that the induced clustering $\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ satisfies that there exists a permutation π on k elements such that for all $i \in [k]$:

$$\left|\widehat{C}_{\widehat{\mu}_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|.$$

We start, in Subsection 6.2, by studying geometric properties of our clustering instance. Recall, that we denote with μ_i the center of cluster C_i in the spectral embedding. We show that, for specific choices of θ, the threshold sets of μ_i have large intersection with the cluster C_i and small intersections with all other cluster C_j. This fact intuitively suggests that our partitioning algorithm works. Unfortunately, as discussed in the technical overview, this is not enough to prove a per cluster guarantee. For this reason in Subsection 6.3 we analyze the overlap structure of $\{C_{\mu_1,\theta}, \ldots, C_{\mu_k,\theta}\}$ more carefully and we give an algorithm (see ComputeOrderedPartition) that given real centers $\{\mu_1, \ldots, \mu_k\}$ and access to exact dot product evaluations computes an ordered partition of $\{\mu_1, \ldots, \mu_k\}$ that induces a valid clustering. In Subsection 6.4 we present an algorithm that guesses the cluster memberships for a set of randomly selected nodes and, using those guesses, approximates cluster centers. Interestingly, we can show, in Subsection 6.4.1, that for the set of correct guesses the algorithm returns a good approximation of the cluster centers. Finally in Subsection 6.5 we show that we can find an ordered partition that induces a good clustering even if we have access only to approximate quantities. That is we show that even if we have access only to approximate means $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$ and the dot product evaluations are only approximately correct then we can find an ordered partition $(T_1, \ldots, T_b)$ that induces a good collection of clusters. The last ingredient is to show that we are able to check if the clustering induced by a specific ordered partition is good. To solve this problem, we design an efficient and simple sampling algorithm which is also analyzed in Subsection 6.5.

6.2 Bounding intersections of $C_{\mu_i,\theta}$ with true clusters C_i

In this subsection we show that, for specific choices of θ, the threshold sets of μ_i (recall that μ_i's are cluster means in the spectral embedding) have large intersection with C_i and small intersections with other clusters. The main idea behind the proof is to use the bounds on dot product of cluster centers presented in Lemma 7. In particular, we use Lemma 6 to relate $\frac{\epsilon}{\varphi^2}$ with the directional variance of the spectral embedding in the direction of μ_i (i.e. $\sum_{x \in C_i} \langle f_x - \mu_i, \alpha \rangle^2$). Then we use the definition of threshold set to upper and lower bound $\langle f_x, \frac{\mu_i}{\|\mu_i\|} \rangle$ and Lemma 7 to upper and lower bound the dot product between cluster centers. By combining the bounds we obtain the following result:

Lemma 31. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $\{C_1, \ldots, C_k\}$. If μ_i's are cluster means then the following conditions hold. Let $S \subset \{\mu_1, \ldots, \mu_k\}$. Let Π denote the orthogonal projection matrix on to the span$(S)^\perp$. Let $\mu \in \{\mu_1, \ldots, \mu_k\} \setminus S$. Let C denote the cluster corresponding to the center μ. Let*

$$\widehat{C} := \{x \in V : \langle \Pi f_x, \Pi \mu \rangle \geq 0.96 \|\Pi \mu\|_2^2\}$$

then we have:

$$\left|C \setminus \widehat{C}\right| \leq \frac{10^4 \epsilon}{\varphi^2} |C|.$$

Proof. Let $x \in C \setminus \widehat{C}$. Then:

$$\begin{aligned}
\left|\left\langle \mu - f_x, \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle\right| &= \left|\left\langle \Pi(\mu - f_x), \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle\right| \\
&\geq 0.04 \cdot \|\Pi\mu\|_2 && \text{Since } \langle \Pi f_x, \Pi\mu\rangle < 0.96\|\Pi\mu\|_2^2 \\
&\geq 0.04 \cdot \left(1 - 24\frac{\sqrt{\epsilon}}{\varphi}\right)||\mu||_2 && \text{By Lemma 12} \\
&\geq 0.04 \cdot \left(1 - 40\frac{\sqrt{\epsilon}}{\varphi}\right)\sqrt{\frac{1}{|C|}} && \text{By Lemma 7} \\
&\geq 0.02 \cdot \sqrt{\frac{1}{|C|}} && \text{Since } \frac{\epsilon}{\varphi^2} \text{ is sufficiently small}
\end{aligned}$$

Then by Lemma 6 applied to direction $\alpha = \frac{\Pi\mu}{\|\Pi\mu\|_2}$ we have $\sum_{i=1}^{k}\sum_{x\in C_i}\langle f_x - \mu_i, \alpha\rangle^2 \leq \frac{4\epsilon}{\varphi^2}$. On the other hand

$$\frac{4\epsilon}{\varphi^2} \geq \sum_{i=1}^{k}\sum_{x\in C_i}\langle f_x - \mu_i, \alpha\rangle^2 \geq \sum_{x\in C\setminus\widehat{C}}\left\langle f_x - \mu, \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle^2 \geq 0.0004 \cdot \frac{|C\setminus\widehat{C}|}{|C|}.$$

Using the above we conclude with $|C \setminus \widehat{C}| \leq 10^4\frac{\epsilon}{\varphi^2}|C|$. □

Remark 7. *Notice that the constants in Lemma 32 are different, they are equal* 0.96 *and* 0.9*. The reason is that the real tests for membership in Algorithm 7 are performed with constant* 0.93 *and the slacks are needed as we have access only to approximate dot products. See* (217) *for the formal reason.*

Lemma 32. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V,E)$ be a d-regular graph that admits a (k,φ,ϵ)-clustering $\{C_1,\ldots,C_k\}$. If μ_i's are cluster means then the following conditions hold. Let $S \subset \{\mu_1,\ldots,\mu_k\}$. Let Π denote the projection matrix on to span$(S)^\perp$. Let $\mu \in \{\mu_1,\ldots,\mu_k\}\setminus S$. Let C denote the cluster corresponding to the center μ. Let*

$$\widehat{C} := \{x \in V : \langle \Pi f_x, \Pi\mu\rangle \geq 0.9\|\Pi\mu\|_2^2\}$$

then we have:

$$\left|\widehat{C} \cap (V\setminus C)\right| \leq 100\frac{\epsilon}{\varphi^2}|C|.$$

Proof. Let $x \in \widehat{C} \cap (V \setminus C)$. Then there exists cluster $C' \neq C$ such that $x \in C'$. Let μ' be the cluster mean of C'. Then:

$$\begin{aligned}
\left|\left\langle f_x - \mu', \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle\right| &\geq \left|\left\langle \Pi f_x, \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle\right| - \left|\left\langle \Pi\mu', \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle\right| && \text{By triangle inequality} \\
&\geq 0.9\|\Pi\mu\|_2 - \left|\left\langle \Pi\mu', \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle\right| && \text{As } x \in \widehat{C}
\end{aligned}$$

Note that either $\mu' \in S$ and then $\Pi\mu' = 0$ and in turn $|\langle \Pi\mu', \Pi\mu\rangle| = 0$ or $\mu' \notin S$ and then $|\langle \Pi\mu', \Pi\mu\rangle| \leq \frac{60\sqrt{\epsilon}}{\varphi^2}\frac{1}{\sqrt{|C|\cdot|C'|}}$ by Lemma 12. Thus we have

$$\begin{aligned}
\left|\left\langle f_x - \mu', \frac{\Pi\mu}{\|\Pi\mu\|_2}\right\rangle\right| &\geq 0.9\|\Pi\mu\|_2 - \frac{60\sqrt{\epsilon}}{\varphi^2}\frac{1}{\sqrt{|C|\cdot|C'|}}\frac{1}{\|\Pi\mu\|_2} \\
&\geq 0.8\frac{1}{\sqrt{|C|}} - \frac{120\sqrt{\epsilon}}{\varphi^2}\frac{1}{\sqrt{|C|\cdot|C'|}}\cdot\sqrt{|C|} && \text{by Lemma 12 and Lemma 7, } \|\Pi\mu\|_2 \geq \frac{1}{2\cdot\sqrt{|C|}} \\
&\geq 0.2\sqrt{\frac{1}{|C|}} && \text{Since } \frac{\epsilon}{\varphi^2} \text{ sufficiently small and } \frac{|C|}{|C'|} \text{ constant}
\end{aligned} \tag{145}$$

Then by Lemma 6 applied to direction $\alpha = \frac{\Pi\mu}{\|\Pi\mu\|_2}$ we have $\sum_{i=1}^{k}\sum_{x\in C_i}\langle f_x - \mu_i, \alpha\rangle^2 \leq \frac{4\epsilon}{\varphi^2}$. On the other hand using (145) we get

$$\frac{4\epsilon}{\varphi^2} \geq \sum_{i=1}^{k}\sum_{x\in C_i}\langle f_x - \mu_i, \alpha\rangle^2 \geq \sum_{x\in\widehat{C}\cap(V\setminus C)}\left\langle f_x - \mu_x, \frac{\Pi\mu_x}{\|\Pi\mu_x\|_2}\right\rangle^2 \geq 0.04 \cdot \frac{|\widehat{C}\cap(V\setminus C)|}{|C|}.$$

Therefore we have $\left|\widehat{C} \cap (V \setminus C)\right| \leq 100\frac{\epsilon}{\varphi^2}|C|$. $\square$

6.3 Partitioning scheme works with *exact* cluster means & dot products

The goal of this section is to present the main ideas behind the algorithms and the analysis. In this section we make a couple of simplifying assumptions. We assume that:

- We have access to real centers $\{\mu_1, \dots, \mu_k\}$,
- Dot products computed by the algorithm are exact,
- A test, that relies on computing outer-conductance of candidate sets, for assessing the quality of clusters is perfect.

Whenever we use one (or more) of these assumptions we state them explicitly in the Lemmas. Later in Section 6.5 we show that we can get rid of all of these assumptions.

In the previous section we showed geometric properties of the threshold sets. Recall that threshold sets are defined as follows:

$$C_{y,\theta} := \{x \in V : \langle f_x, y\rangle \geq \theta ||y||^2\}.$$

In this section, using these properties of threshold sets, we show an algorithm that given exact centers, access to real dot products and a perfect primitive for computing outer-conductance computes an ordered partition $(T_1, \dots, T_b)$ of $\{\mu_1, \dots, \mu_k\}$ such that $(T_1, \dots, T_b)$ induces a good collection of clusters.

Algorithm 8 COMPUTEORDEREDPARTITION$(G, \widehat{\mu}_1, \widehat{\mu}_2, \dots, \widehat{\mu}_k, s_1, s_2)$ ▷ $\widehat{\mu}_i$'s given as sets of points
▷ s_1 is # sampled points for size estimation
▷ s_2 is # of sampled points for conductance estimation

$S := \{\hat{\mu}_1, \dots, \hat{\mu}_k\}$
for $i = 1$ to $\lceil \log(k) \rceil$ **do**
    $T_i := \emptyset$
    for $\widehat{\mu} \in S$ **do**
        $\psi :=$ OUTERCONDUCTANCE$(G, \widehat{\mu}, (T_1, T_2, \dots, T_{i-1}), S, s_1, s_2)$ ▷ Algorithm 11
        if $\psi \leq O(\frac{\epsilon}{\varphi^2} \cdot \log(k))$ **then**
            $T_i := T_i \cup \{\widehat{\mu}\}$
    $S := S \setminus T_i$
    if $S = \emptyset$ **then**
        return (TRUE, $(T_1, \dots, T_i)$)
return (FALSE, $\perp$)

To explain and analyze COMPUTEORDEREDPARTITION we first need to introduce another algorithm and some definitions.

Definition 11. For a set $\{a_1, \dots, a_i\}$ we say a sequence $(S_1, \dots, S_p)$ is an ordered partial partition of $\{a_1, \dots, a_i\}$ if:

- $\bigcup_{j \in [p]} S_j \subseteq \{a_1, \dots, a_i\}$,
- S_i's are pairwise disjoint.

Intuitively Algorithm ISINSIDE emulates CLASSIFYBYHYPERPLANEPARTITIONING on ordered partial partition $(T_1, \dots, T_b)$. This intuition is made formal, after introducing Definition 12, in Remark 8. For this we need additional notation for clusters that are implicitly created by ISINSIDE. We define:

Definition 12 (Candidate cluster). For an ordered partial partition $P = (T_1, \dots, T_p)$ of approximate cluster means $\{\widehat{\mu}_1, \dots, \widehat{\mu}_k\}$ and $\widehat{\mu} \in \{\widehat{\mu}_1, \dots, \widehat{\mu}_k\} \setminus \bigcup_{i \in [p]} T_i$ we say that $\widehat{C}^P_{\widehat{\mu}}$ is a **candidate cluster corresponding to $\widehat{\mu}$ with respect to P** if:

$$\widehat{C}^P_{\widehat{\mu}} = \left\{ x \in V : \text{ISINSIDE}\left(x, \widehat{\mu}, P, \{\widehat{\mu}_1, \dots, \widehat{\mu}_k\} \setminus \bigcup_{i \in [p]} T_i \right) = \text{TRUE} \right\}.$$

Furthermore we define: $V^P := V \setminus \bigcup_{j<p} \bigcup_{\widehat{\mu} \in T_j} \widehat{C}^{(T_1, \dots, T_{j-1})}_{\widehat{\mu}}$.

Algorithm 9 IsInside$(x, \widehat{\mu}, (T_1, T_2, \ldots, T_b), S)$

▷ T_i's are sets of $\widehat{\mu}_j$ where $\widehat{\mu}_j$'s are given as sets of points
▷ see Section 5.6 for the reason of such representation
▷ S = set of not yet processed centers, $\widehat{\mu} \in S$

1: **for** $i = 1$ to b **do**
2: Let Π be the projection onto the span$(\bigcup_{j<i} T_j)^{\perp}$.
3: Let $S_i = \left(\bigcup_{j \geq i} T_j\right) \cup S$
4: **for** $\hat{\mu}_i \in T_i$ **do**
5: **if** $x \in C^{apx}_{\Pi\widehat{\mu}_i, 0.93} \setminus \bigcup_{\widehat{\mu}' \in S_i \setminus \{\hat{\mu}_i\}} C^{apx}_{\Pi\widehat{\mu}', 0.93}$ **then** ▷ see (144) for definition of $C^{apx}_{y,\theta}$
6: **return** False
7: Let Π be the projection onto the span$(\bigcup_{j \leq b} T_j)^{\perp}$.
8: **if** $x \in C^{apx}_{\Pi\widehat{\mu}, 0.93} \setminus \bigcup_{\widehat{\mu}' \in S \setminus \{\widehat{\mu}\}} C^{apx}_{\Pi\widehat{\mu}', 0.93}$ **then** ▷ see (144) for definition of $C^{apx}_{y,\theta}$
9: **return** True
10: **return** False

Algorithm IsInside receives a vertex x, the centre of a cluster $\widehat{\mu}$, and an ordered partial partition, then it tests if vertex x is not recovered by any of the previous stages (see line (5) of Algorithm 9) and can be recovered at the current stage using $\widehat{\mu}$. More formally, it can be recovered at the current stage if it only belongs to the candidate cluster corresponding to the center $\widehat{\mu}$ (see line (8) of Algorithm 9).

Remark 8. *Note that Definitions 10 and 12 are compatible in the following sense. For an ordered partition $(T_1, \ldots, T_b)$ of approximate cluster means $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$ that induces a collection of clusters $\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ it is true that:*

$$\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\} = \bigcup_{i \in [b]} \bigcup_{\widehat{\mu} \in T_i} \{\widehat{C}^{(T_1, \ldots, T_{i-1})}_{\widehat{\mu}}\},$$

Equipped with Definition 12 we are ready to explain Algorithm ComputeOrderedPartition. The Algorithm proceeds in $O(\log(k))$ stages. It maintains a set S of approximate cluster means, that initially is equal to $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$, from which $\widehat{\mu}$'s are removed after every stage. At every stage i a collection of sets

$$\mathcal{C}_i := \bigcup_{\widehat{\mu} \in S} \{\widehat{C}^{(T_1, \ldots, T_{i-1})}_{\widehat{\mu}}\},$$

is implicitly considered. In fact sets in this collection are, by definition, pairwise disjoint (see Definition 12 and line: 8 of IsInside). $\widehat{C}^{(T_1, \ldots, T_{i-1})}_{\widehat{\mu}}$'s are defined as threshold sets (see Definition 8) that are made disjoint by removing intersections. The main idea behind the Algorithm is to use properties from Section 6.2 so that we can show that $\widehat{C}^{(T_1, \ldots, T_{i-1})}_{\widehat{\mu}}$'s match some C_j's well. Unfortunately after removing the intersections the above property might not hold for every cluster in $\mathcal{C}_i$. In the rest of this section we show however that it is true for a constant fraction of sets from $\mathcal{C}_i$. The Algorithm ComputeOrderedPartition proceeds by discarding, from set S, the $\widehat{\mu}$'s for which $\widehat{C}^{(T_1, \ldots, T_{i-1})}_{\widehat{\mu}}$ matches some C_j's well and implicitly removes the vertices of $\widehat{C}^{(T_1, \ldots, T_{i-1})}_{\widehat{\mu}}$ from consideration. Moreover it projects out the directions corresponding to the removed $\widehat{\mu}$'s and restricts its attention to a lower dimensional subspace Π of $\mathbb{R}^k$ (see **Idealized Clustering Algorithm** from Section 6.1 for comparison). The Algorithm doesn't know which sets from $\mathcal{C}_i$ are good as it runs in sublinear time. That is why we develop a simple sampling procedure that computes outer-conductance of candidate clusters (see Algorithm 11). Then the Algorithm removes the $\widehat{\mu}$'s for which the corresponding $\widehat{C}^{(T_1, \ldots, T_{i-1})}_{\widehat{\mu}}$ have small outer-conductance. We conclude using the robustness property of (k, φ, ϵ)-clusterable graphs (Lemma 16) that these tests are enough.

The rest of this subsection is devoted to showing that if ComputeOrderedPartition is called with $(\hat{\mu}_1, \ldots, \hat{\mu}_k)$ equal to $(\mu_1, \ldots, \mu_k)$ and the algorithm has access to real dot products then ComputeOrderedPartition returns True and an ordered partition $(T_1, \ldots, T_b)$ (of $\{\mu_1, \ldots, \mu_k\}$) that induces a collection of pairwise disjoint clusters $\{\widehat{C}_{\mu_1}, \ldots, \widehat{C}_{\mu_k}\}$ such that for every i:

$$\phi\left(\widehat{C}_{\mu_i}\right) \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right). \tag{146}$$

Then using Lemma 16 we get that there exists a permutation π such that for all $i \in [k]$:

$$\left|\widehat{C}_{\mu_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|. \tag{147}$$

The core of the argument is an averaging argument that, for every linear subspace of $\mathbb{R}^k$, bounds the average distance of embedded points to their centers in this subspace. What is important is that the bound depends linearly on the dimensionality of the subspace.

Lemma 33. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $\{C_1, \ldots, C_k\}$. Then for all $L \subseteq \mathbb{R}^k$ - a linear subspace of $\mathbb{R}^k$, Π the orthogonal projection onto L we have:*

$$\sum_{x \in V} \|\Pi f_x - \Pi \mu_x\|_2^2 \leq O\left(dim(L) \cdot \frac{\epsilon}{\varphi^2}\right)$$

Proof. Let $b := dim(L)$ and $\{w_1, \ldots, w_b\}$ be any orthonormal basis of L and recall that for $x \in V$ μ_x is the cluster mean of the cluster which x belongs to. Then

$$\begin{aligned}
\sum_{x \in V} \|\Pi f_x - \Pi \mu_x\|_2^2 &= \sum_{x \in V} \sum_{i=1}^{b} \langle f_x - \mu_x, w_i \rangle^2 \\
&= \sum_{i=1}^{b} \sum_{x \in V} \langle f_x - \mu_x, w_i \rangle^2 \\
&\leq b \cdot \frac{4\epsilon}{\varphi^2} && \text{By Lemma 6}
\end{aligned}$$

□

In order to show (146) we need to show that a constant fraction of candidate sets $\widehat{C}_{\mu}^{(T_1, \ldots, T_{i-1})}$'s match some C_j's well. To do that we argue that that sets of the form $C_{\Pi\widehat{\mu}, 0.9}$ (where Π is the orthogonal projection onto the span$(\bigcup_{j<i} T_j)^{\perp}$) don't overlap too much. We do this in two steps. First in Lemma 34 and Lemma 35 we show that points from the intersections are far from their centers. Then in Lemma 36 below we show that having too many such vertices would contradict Lemma 33.

Lemma 34. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $\{C_1, \ldots, C_k\}$. Let $\{v_1, \ldots, v_k\} \in \mathbb{R}^k$ be a set of vectors satisfying:*

- $|\langle v_i, v_j \rangle| \leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) \frac{1}{\sqrt{|C_i||C_j|}}$
- $\left| ||v_i||^2 - \frac{1}{|C_i|} \right| \leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) \frac{1}{|C_i|}$

Then for every pair $i \neq j \in [k]$ for every $\theta \in (0,1)$ if $\alpha := \frac{v_i \frac{\|v_j\|}{\|v_i\|} + v_j \frac{\|v_i\|}{\|v_j\|}}{\sqrt{\|v_i\|^2 + \|v_j\|^2}}$ and $I := C_{v_i, \theta} \cap C_{v_j, \theta} = \{x \in V : \langle f_x, v_i \rangle \geq \theta \|v_i\|^2 \wedge \langle f_x, v_j \rangle \geq \theta \|v_j\|^2\}$ then the following conditions hold:

1. *Correlation of vector v_p with the direction α is as follows:*
   - *for all $p \in [k] \setminus \{i, j\}$, $\left\langle \frac{\alpha}{\|\alpha\|}, v_p \right\rangle \leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) \cdot \frac{\|v_i\| \cdot \|v_j\|}{\sqrt{\|v_i\|^2 + \|v_j\|^2}}$, for all $i \neq j \in [k]$*
   - *for all $p \in \{i, j\}$, $\left\langle \frac{\alpha}{\|\alpha\|}, v_p \right\rangle \leq \left(1 + O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right) \cdot \frac{\|v_i\| \cdot \|v_j\|}{\sqrt{\|v_i\|^2 + \|v_j\|^2}}$ for all $i \in [k]$*
2. *Spectral embeddings of vertices from set I have big correlation with direction α.*

$$\min_{x \in I} \left\langle \frac{\alpha}{\|\alpha\|}, f_x \right\rangle \geq \left(2\theta - O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right) \cdot \frac{\|v_i\| \cdot \|v_j\|}{\sqrt{\|v_i\|^2 + \|v_j\|^2}}$$

Proof. For all $p \in [k]$ let $\widetilde{v}_p := v_p/||v_p||$. Let $\gamma := \frac{||v_j||}{\sqrt{||v_i||^2+||v_j||^2}}$, $\alpha := \gamma\widetilde{v}_i + \sqrt{1-\gamma^2}\widetilde{v}_j$, and $\widetilde{\alpha} := \alpha/||\alpha||$. Fix $i \neq j \in [1,\ldots,k]$. First we show that since v_i's are close to orthogonal we have $||\alpha||^2 \approx 1$. More precisely we will upper bound $|||\alpha||^2 - 1|$

$$
\begin{aligned}
\left|||\alpha||^2 - 1\right| &= \left|\gamma^2||\widetilde{v}_i||^2 + (1-\gamma^2)||\widetilde{v}_j||^2 + 2\gamma\sqrt{1-\gamma^2}\left\langle \widetilde{v}_i, \widetilde{v}_j\right\rangle - 1\right| \\
&= \frac{2\left\langle v_i, v_j\right\rangle}{||v_i||^2 + ||v_j||^2} && \text{as } ||\widetilde{v}_i|| = ||\widetilde{v}_j|| = 1 \\
&\leq \frac{2 \cdot O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\frac{1}{\sqrt{|C_i||C_j|}}}{\left(1 - O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right)\left(\frac{1}{|C_i|} + \frac{1}{|C_j|}\right)} && \text{By assumptions} \\
&\leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\frac{\sqrt{|C_i||C_j|}}{|C_i| + |C_j|} \\
&\leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) && \text{as } \frac{\sqrt{|C_i||C_j|}}{\max(|C_i|,|C_j|)} \leq 1 \qquad (148)
\end{aligned}
$$

Observe the following fact:

$$\sqrt{1-\gamma^2}\cdot||v_j|| = \gamma \cdot ||v_i|| \tag{149}$$

Next notice the following:

$$\langle \alpha, v_i\rangle = \gamma||v_i|| + \langle \widetilde{v}_i, \widetilde{v}_j\rangle \cdot \sqrt{1-\gamma^2}||v_i|| \tag{150}$$

$$\langle \alpha, v_j\rangle = \langle \widetilde{v}_i, \widetilde{v}_j\rangle \cdot \gamma||v_j|| + \sqrt{1-\gamma^2}||v_j|| \tag{151}$$

For all $p \in \{1,2,\ldots,k\} \setminus \{i,j\}$

$$\langle \alpha, v_p\rangle = \langle \widetilde{v}_i, \widetilde{v}_p\rangle \cdot \gamma||v_p|| + \langle \widetilde{v}_j, \widetilde{v}_p\rangle\sqrt{1-\gamma^2}||v_p|| \tag{152}$$

Moreover for all $p \neq q \in [1,\ldots,k]$ we have

$$
\begin{aligned}
\left|\frac{1}{||v_p||^2}\cdot\langle v_q, v_p\rangle\right| &\leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\frac{1}{\sqrt{|C_q||C_p|}}|C_p|\frac{1}{\left(1 - O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right)} && \text{By assumptions} \\
&\leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\sqrt{\frac{|C_p|}{|C_q|}} && \text{for small enough } \frac{\epsilon}{\varphi^2} \\
&\leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) && \text{as } \frac{|C_p|}{|C_q|} = O(1) \qquad (153)
\end{aligned}
$$

Using the above we can prove:

$$
\begin{aligned}
\left|\langle \widetilde{v}_i, \widetilde{v}_j\rangle \cdot \sqrt{1-\gamma^2}||v_i||\right| &= \left|\sqrt{1-\gamma^2}\cdot||v_j||\cdot\frac{1}{||v_j||^2}\cdot\langle v_i, v_j\rangle\right| \\
&\leq \sqrt{1-\gamma^2}\cdot||v_j||\cdot O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) && \text{By (153)} \\
&= O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\cdot\gamma\cdot||v_i|| && \text{By (149)} \qquad (154)
\end{aligned}
$$

And similarly we show:

$$
\begin{aligned}
|\langle \widetilde{v}_i, \widetilde{v}_j\rangle \cdot \gamma||v_j||| &= \left|\gamma\cdot||v_i||\cdot\frac{1}{||v_i||^2}\cdot\langle v_i, v_j\rangle\right| \\
&\leq O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\cdot\gamma\cdot||v_i|| && \text{By (153)} \qquad (155)
\end{aligned}
$$

For all $p \in \{1, 2, \ldots, k\} \setminus \{i, j\}$ we get

$$\begin{aligned}
|\langle \alpha, v_p \rangle| &\le \left| \langle \widetilde{v}_i, \widetilde{v}_p \rangle \cdot \gamma ||v_p|| \right| + \left| \langle \widetilde{v}_j, \widetilde{v}_p \rangle \sqrt{1-\gamma^2} ||v_p|| \right| && \text{By (152)} \\
&= \left| \langle v_i, v_p \rangle \cdot \frac{1}{||v_i||^2} ||v_i|| \gamma \right| + \left| \langle v_j, v_p \rangle \frac{1}{||v_j||^2} ||v_j|| \sqrt{1-\gamma^2} \right| \\
&\le O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) \cdot \gamma \cdot ||v_i|| && \text{By (153) and (149)} \qquad (156)
\end{aligned}$$

Combining (150), (151), (154), (155) and (156) we get that for all $p \in \{i, j\}$ we have

$$\langle \alpha, v_p \rangle \le \left(1 + O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right) \cdot \gamma \cdot ||v_i|| \tag{157}$$

and for all $p \in \{1, \ldots, k\} \setminus \{i, j\}$

$$\langle \alpha, v_p \rangle \le O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) \cdot \gamma \cdot ||v_i|| \tag{158}$$

Now using (148) we get that for all $p \in \{i, j\}$

$$\langle \widetilde{\alpha}, v_p \rangle \le \frac{1}{\sqrt{1 - O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)}} \left(1 + O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right) \cdot \gamma \cdot ||v_i|| \le \left(1 + O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right) \cdot \frac{||v_i|| ||v_j||}{\sqrt{||v_i||^2 + ||v_j||^2}}$$

and for all $p \in \{1, \ldots, k\} \setminus \{i, j\}$

$$\langle \widetilde{\alpha}, v_p \rangle \le \frac{1}{\sqrt{1 - O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)}} O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) \cdot \gamma \cdot ||v_i|| \le O\left(\frac{\sqrt{\epsilon}}{\varphi}\right) \cdot \frac{||v_i|| ||v_j||}{\sqrt{||v_i||^2 + ||v_j||^2}}$$

These two inequalities establish the first statement of the Claim.

Recall that

$$I = \{x \in V : \langle f_x, v_i \rangle \ge \theta ||v_i||^2 \wedge \langle f_x, v_j \rangle \ge \theta ||v_j||^2\}$$

Now let $x \in I$. Then observe

$$\begin{aligned}
\langle \alpha, f_x \rangle &= \langle \gamma \cdot \widetilde{v}_i, f_x \rangle + \left\langle \sqrt{1-\gamma^2} \cdot \widetilde{v}_j, f_x \right\rangle \\
&\ge \gamma \cdot \theta \cdot ||v_i|| + \sqrt{1-\gamma^2} \cdot \theta \cdot ||v_j|| && \text{because } x \in I \\
&= 2\theta \cdot \gamma \cdot ||v_i|| && \text{by (149)}
\end{aligned}$$

Hence

$$\begin{aligned}
\langle \widetilde{\alpha}, f_x \rangle &\ge \frac{1}{\sqrt{1 + O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)}} 2\theta \cdot \gamma \cdot ||v_i|| && \text{By (148)} \\
&\ge \left(2\theta - O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right) \cdot \gamma \cdot ||v_i||
\end{aligned}$$

□

Now we use technical Lemma 34 to show that vertices from the intersections of $C_{\Pi\mu,0.9}$'s are far from their centers.

Lemma 35. *Let $k \ge 2$, $\varphi \in (0, 1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $\{C_1, \ldots, C_k\}$. If μ_i's are cluster means then the following conditions hold. For all $S \subset \{\mu_1, \ldots, \mu_k\}$ if $L := span(S)^\perp$ and Π is the projection on L then if $x \in V$ is such that*

$$\langle \Pi f_x, \Pi \mu_i \rangle \ge 0.9 \|\Pi \mu_i\|_2^2 \wedge \langle \Pi f_x, \Pi \mu_j \rangle \ge 0.9 \|\Pi \mu_j\|_2^2$$

for some $\mu_i, \mu_j \in \{\mu_1, \ldots, \mu_k\} \setminus S, \mu_i \ne \mu_j$. Then:

$$\|\Pi f_x - \Pi \mu_x\| \ge 0.3 \sqrt{\frac{1}{\max_{p \in [k]} |C_p|}}$$

Proof. Let $x \in V$ be such that $\langle \Pi f_x, \Pi\mu_i\rangle \geq 0.9\|\Pi\mu_i\|_2^2$ and $\langle \Pi f_x, \Pi\mu_j\rangle \geq 0.9\|\Pi\mu_j\|_2^2$. Note that by Lemma 12 set $\{\Pi\mu_1, \ldots, \Pi\mu_k\}$ satisfies assumptions of Lemma 34. So applying Lemma 34 for $\theta = 0.9$ we get that there exists $\alpha \in \text{span}\{\Pi\mu_i, \Pi\mu_j\}, ||\alpha|| = 1$ such that:

- $\langle \alpha, f_x\rangle = \langle \alpha, \Pi f_x\rangle \geq (1.8 - O(\frac{\sqrt{\epsilon}}{\varphi})) \cdot \frac{\|\Pi\mu_i\| \cdot \|\Pi\mu_j\|}{\sqrt{\|\Pi\mu_i\|^2 + \|\Pi\mu_j\|^2}}$
- $\langle \alpha, \Pi\mu_p\rangle \leq (1 + O(\frac{\sqrt{\epsilon}}{\varphi})) \cdot \frac{\|\Pi\mu_i\| \cdot \|\Pi\mu_j\|}{\sqrt{\|\Pi\mu_i\|^2 + \|\Pi\mu_j\|^2}}$, for all $p \in [k]$

Thus we get

$$\begin{aligned}
||\Pi f_x - \Pi\mu_x|| &\geq |\, \langle \alpha, \Pi f_x\rangle - \langle \alpha, \Pi\mu_x\rangle \,| \\
&\geq \left(0.8 - O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right) \cdot \frac{\|\Pi\mu_i\| \cdot \|\Pi\mu_j\|}{\sqrt{\|\Pi\mu_i\|^2 + \|\Pi\mu_j\|^2}} \\
&\geq 0.75 \cdot \frac{\|\Pi\mu_i\| \cdot \|\Pi\mu_j\|}{\sqrt{\|\Pi\mu_i\|^2 + \|\Pi\mu_j\|^2}} \qquad \text{By assumption that } \frac{\epsilon}{\varphi^2} \text{ small} \qquad (159)
\end{aligned}$$

without loss of generality we can assume $||\Pi\mu_i|| \geq ||\Pi\mu_j||$. Then we get:

$$\begin{aligned}
\frac{\|\Pi\mu_i\| \cdot \|\Pi\mu_j\|}{\sqrt{\|\Pi\mu_i\|^2 + \|\Pi\mu_j\|^2}} &= \frac{\|\Pi\mu_j\|}{\sqrt{1 + \|\Pi\mu_j\|^2/||\Pi\mu_i||^2}} \\
&\geq \frac{1}{\sqrt{2}}||\Pi\mu_j|| \\
&\geq \frac{1}{2\sqrt{\max_{p\in[k]} |C_p|}} \qquad \text{Lemma 12, assumption that } \frac{\epsilon}{\varphi^2} \text{ small} \qquad (160)
\end{aligned}$$

Combining (159) and (160) we get:

$$||\Pi f_x - \Pi\mu_x|| \geq 0.3 \cdot \frac{1}{\sqrt{\max_{p\in[k]} |C_p|}}$$

□

Combining Lemma 33 and Lemma 35 we show that sets $C_{\Pi\mu,0.9}$'s don't overlap much.

Lemma 36. *Let $k \geq 2$, $\varphi \in (0, 1)$ and $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $\{C_1, \ldots, C_k\}$. If μ_i's are cluster means then the following conditions hold. For all $S \subset \{\mu_1, \ldots, \mu_k\}$ if $L := span(S)^\perp$, $dim(L) = b$ and Π is projection on L then:*

$$\left| \bigcup_{\substack{\mu,\mu' \in \{\mu_1,\ldots,\mu_k\}\setminus S \\ \mu \neq \mu'}} C_{\Pi\mu,0.9} \cap C_{\Pi\mu',0.9} \right| \leq O\left(b \cdot \frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k}.$$

Proof. Let $x \in V$ be such that $\langle \Pi f_x, \Pi\mu\rangle \geq 0.9\|\Pi\mu\|_2^2$ and $\langle \Pi f_x, \Pi\mu'\rangle \geq 0.9\|\Pi\mu'\|_2^2$ for some $\mu, \mu' \in \{\mu_1, \ldots, \mu_k\} \setminus S$. Then by Lemma 35 we get that

$$\|\Pi f_x - \Pi\mu_x\| \geq 0.3\sqrt{\frac{1}{\max_{p\in[k]} |C_p|}}. \tag{161}$$

On the other hand Lemma 33 guarantees:

$$\sum_{x\in V} \|\Pi f_x - \Pi\mu_x\|_2^2 \leq O\left(dim(L) \cdot \frac{\epsilon}{\varphi^2}\right) \tag{162}$$

Combining (161), (162) and the fact that $\frac{\max_{p\in[k]} |C_p|}{\min_{p\in[k]} |C_p|} = O(1)$ we get

$$\left| \bigcup_{\mu,\mu' \in \{\mu_1,\ldots,\mu_k\}\setminus S} C_{\Pi\mu,0.93} \cap C_{\Pi\mu',0.93} \right| \leq O\left(b \cdot \frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k}$$

□

Our bounds above enable the following analysis. At every stage of the for loop from line 4 of Algorithm 8 at least half of the candidate clusters:

$$\mathcal{C}_i := \bigcup_{\widehat{\mu}\in S}\{\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_{i-1})}\},$$

passes the test from line 6 of Algorithm 8, which means that they have small outer-conductance and satisfy condition (146).

Lemma 37. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}\cdot\log(k)$ be smaller than a sufficiently small constant. Let $G=(V,E)$ be a d-regular graph that admits a (k,φ,ϵ)-clustering $\{C_1,\ldots,C_k\}$.*

If ComputeOrderedPartition*$(G,\widehat{\mu}_1,\widehat{\mu}_2,\ldots,\widehat{\mu}_k,s_1,s_2)$ is invoked with $(\hat{\mu}_1,\ldots,\hat{\mu}_k)=(\mu_1,\ldots,\mu_k)$ and we assume that all tests Algorithm 8 performs $\left(i.e. \left\langle f_x,\widehat{\Pi}\widehat{\mu}\right\rangle_{apx} \overset{?}{\geq} 0.93\left\|\widehat{\Pi}\widehat{\mu}\right\|_{apx}^2\right)$ are exact and* OuterConductance *computes outer-conductance precisely then there exists an absolute constant Υ such that the following conditions hold.*

For any $i\in[0..\log(k)]$ assume that at the beginning of the i-th iteration of the for loop from line 4 of Algorithm 8 $|S|=b$ and, up to renaming of μ's, $S=\{\mu_1,\ldots,\mu_b\}$, the corresponding clusters are $\mathcal{C}=\{C_1,\ldots,C_b\}$ respectively and the ordered partial partition of μ's is equal to $(T_1,\ldots,T_{i-1})$. Then if for every $C\in\mathcal{C}$ we have that $|V^{(T_1,\ldots,T_{i-1})}\cap C| \geq \left(1-\Upsilon\cdot i\cdot\frac{\epsilon}{\varphi^2}\right)|C|$ then at the beginning of $(i+1)$-th iteration:

1. *$|S|\leq b/2$ (that is at least half of the remaining cluster means were removed in i-th iteration),*
2. *for every $\mu\in S$ the corresponding cluster C satisfies $|V^{(T_1,\ldots,T_i)}\cap C|\geq\left(1-\Upsilon\cdot(i+1)\cdot\frac{\epsilon}{\varphi^2}\right)|C|$, where $(T_1,\ldots,T_i)$ is the ordered partial partition of μ's created in the first i iterations.*

Proof. Let $i\in[0..\log(k)]$, without loss of generality we can assume that $S=\{\mu_1,\ldots,\mu_b\}$ (if not we can rename the μ's) at the beginning of the i-th iteration and the corresponding clusters be $\mathcal{C}=\{C_1,\ldots,C_b\}$ respectively. Assume that for every $C\in\mathcal{C}$ we have that $|V^{(T_1,\ldots,T_{i-1})}\cap C|\geq\left(1-\Upsilon\cdot i\cdot\frac{\epsilon}{\varphi^2}\right)|C|$. We start by showing the first part of the Lemma.

At least half of the cluster means is removed from S:

Let $\mu\in S$, Π_i be the orthogonal projection onto the span$(\bigcup_{j<i}T_j)^\perp$, where $(T_1,\ldots,T_{i-1})$ is the ordered partial partition of $\{\mu_1,\ldots,\mu_k\}$ created before iteration i by ComputeOrderedPartition. For brevity we will refer to $(T_1,\ldots,T_{i-1})$ as P in this proof. Let

$$I := \bigcup_{\mu',\mu''\in\{\mu_1,\ldots,\mu_b\}} C_{\Pi_i\mu',0.93}\cap C_{\Pi_i\mu'',0.93}.$$

By Lemma 36 we have that

$$|I|\leq O\left(b\cdot\frac{\epsilon}{\varphi^2}\right)\cdot\frac{n}{k}$$

So by Markov inequality we get that there exists a subset of clusters $\mathcal{R}\subseteq\mathcal{C}$ such that $|\mathcal{R}|\geq b/2$ and for every $C\in\mathcal{R}$ we have that

$$|C\cap I|\leq 2\cdot O\left(\frac{\epsilon}{\varphi^2}\right)\cdot\frac{n}{k} \tag{163}$$

We will argue that for any order of the for loop from line 4 of Algorithm 8 it is true that for every $C\in\mathcal{R}$ with corresponding mean μ the candidate cluster $\widehat{C}_\mu^P$ satisfies the if statement from line 6.

First note that behavior of the algorithm is independent of the order of the for loop from line 4 of Algorithm 8 as by definition $\widehat{C}_\mu^P$'s for $\mu\in S$ are pairwise disjoint. Now let $C\in\mathcal{R}$, μ be the corresponding mean to C and $\widehat{C}_\mu^P$ be the candidate cluster corresponding to μ with respect to $P=(T_1,\ldots,T_{i-1})$. By inductive assumption $|V^P\cap C|\geq\left(1-\Upsilon\cdot i\cdot\frac{\epsilon}{\varphi^2}\right)|C|$ so by (163), Lemma 31 and the fact that $\frac{\max_{p\in[k]}|C_p|}{\min_{p\in[k]}|C_p|}=O(1)$ we get that:

$$\begin{aligned}|\widehat{C}_\mu^P\cap C| &\geq \left(1-\Upsilon\cdot i\cdot\frac{\epsilon}{\varphi^2}\right)|C| - O\left(\frac{\epsilon}{\varphi^2}\right)\frac{n}{k} - O\left(\frac{\epsilon}{\varphi^2}\right)|C| \\ &\geq \left(1-O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right)\right)|C| \end{aligned} \tag{164}$$

To prove that $\widehat{C}_\mu^P$ passes the outer-conductance test we also need to show that $\widehat{C}_\mu^P$ doesn't contain a lot of points from $V^P \setminus C$. By Lemma 32 we get that:

$$|\widehat{C}_\mu^P \cap (V^P \setminus C)| \leq |\widehat{C}_\mu^P \cap (V \setminus C)| \leq O\left(\frac{\epsilon}{\varphi^2}\right)|C|. \tag{165}$$

Combining (165) and (164) we get that:

$$|\widehat{C}_\mu^P \triangle C| \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)|C| \tag{166}$$

Now we want to argue that $\widehat{C}_\mu^P$ passes the outerconductance test from line 6 of Algorithm 8. From the definition of outer conductance:

$$\begin{aligned}
\phi(\widehat{C}_\mu^P) &\leq \frac{E(C, V \setminus C) + d|\widehat{C}_\mu^P \triangle C|}{d(|C| - |\widehat{C}_\mu^P \triangle C|)} \\
&\leq \frac{E(C, V \setminus C) + d \cdot O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)|C|}{d(|C| - O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)|C|)} && \text{from (166)} \\
&\leq \frac{O\left(\frac{\epsilon}{\varphi^2}\right) + O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)}{1 - O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)} && \text{because } \frac{E(C, V \setminus C)}{d|C|} \leq O\left(\frac{\epsilon}{\varphi^2}\right) \\
&\leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right) && \text{for sufficiently small } \frac{\epsilon}{\varphi^2} \cdot \log(k)
\end{aligned}$$

and it follows that

$$\phi(\widehat{C}_\mu^P) \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right),$$

which means that $\widehat{C}_\mu^P$ passes the test as we assumed that OuterConductance computes outer-conductance precisely.

Clusters corresponding to unremoved μ's satisfy condition 2:
Now we prove that for every μ that was not removed from set S only small fraction of its corresponding cluster is removed.

Let $\mu \in S$ be such that it is not removed in the i-th step. Let Π_i be the orthogonal projection onto the $\text{span}(\bigcup_{j<i} T_j)^\perp$. Let $C \in \mathcal{C}$ be the cluster corresponding to μ. By assumption $|V^P \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$. Now let $x \in V^{(T_1, \ldots, T_{i-1})} \setminus V^{(T_1, \ldots, T_i)}$, where $(T_1, \ldots, T_i)$ is the partial partition of μ's created in the first i-th steps of the for loop. We get that there exists $\mu' \in \{\mu_1, \ldots, \mu_b\}$ such that $x \in \widehat{C}_{\mu'}^P$ (recall that $\widehat{C}_{\mu'}^P$ is the candidate cluster corresponding to μ' with respect to $P = (T_1, \ldots, T_{i-1})$). Recall (Definition 12) that $\widehat{C}_{\mu'}^P$ is defined as:

$$\widehat{C}_{\mu'}^P = \left\{ x \in V : \textsc{IsInside}\left(x, \mu', P, \{\mu_1, \ldots, \mu_k\} \setminus \bigcup_{j \in [i-1]} T_j\right) = \textsc{True} \right\}.$$

This in particular means (see line 8: of Algorithm IsInside) that:

$$\widehat{C}_{\mu'}^P \subseteq C_{\Pi_i \mu', 0.93} \setminus \bigcup_{\mu'' \in S \setminus \{\mu'\}} C_{\Pi_i \mu'', 0.93},$$

which, as $\mu \in S \setminus \{\mu'\}$, gives us that:

$$\widehat{C}_{\mu'}^P \cap C_{\Pi_i \mu, 0.93} = \emptyset,$$

and finally, using Definition 8, we have:

$$\langle f_x, \Pi_i \mu \rangle < 0.93 ||\Pi_i \mu||^2. \tag{167}$$

But by Lemma 31:

$$|\{x \in C : \langle \Pi_i f_x, \Pi_i \mu \rangle < 0.93 \|\Pi_i \mu\|_2^2\}| \leq O\left(\frac{\epsilon}{\varphi^2}\right) \cdot |C| \tag{168}$$

Combining (167) and (168) we get that:

$$|C \cap (V^{(T_1,\dots,T_{i-1})} \setminus V^{(T_1,\dots,T_i)})| \leq O\left(\frac{\epsilon}{\varphi^2}\right)|C|. \tag{169}$$

By assumption that $|V^{(T_1,\dots,T_{i-1})} \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$ and (169) we get that:

$$|V^{(T_1,\dots,T_i)} \cap C| \geq \left(1 - \Upsilon \cdot (i+1) \cdot \frac{\epsilon}{\varphi^2}\right)|C|,$$

provided that Υ is bigger than the constant from O notation in (169), which is the same constant as the one in the statement of Lemma 31.

□

Remark 9. *Note that in this section we assume that the Algorithm has access to real centers* $\{\mu_1,\dots,\mu_k\}$*. If it was the case in the final algorithm we could in fact prove a stronger guarantee, i.e. "Algorithm 8 returns* TRUE *and an ordered partition* $(T_1,\dots,T_b)$ *(of* $\{\mu_1,\dots,\mu_k\}$*) that induces a collection of pairwise disjoint clusters* $\{\widehat{C}_{\mu_1},\dots,\widehat{C}_{\mu_k}\}$ *such that there exists a permutation* π *such that for all* $i \in [k]$*:*

$$\left|\widehat{C}_{\mu_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right)|C_{\pi(i)}|\text{''}.$$

Compare the above statement with with (147) *and the main theorem of this section, Theorem 7. The reason we present it this way is the following.*

The final algorithm doesn't have access to μ*'s but instead tests many candidate sets* $\{\widehat{\mu}_1,\dots,\widehat{\mu}_k\}$*. Moreover Algorithm 8 returns an ordered partition* $(T_1,\dots,T_b)$ *that induces a collection of clusters* $\{\widehat{C}_1,\dots,\widehat{C}_k\}$ *whenever every set from this collection passes the test from line 6 of* COMPUTEORDEREDPARTITION*, that is when for every* $\widehat{C} \in \{\widehat{C}_1,\dots,\widehat{C}_k\}$*:*

$$\phi\left(\widehat{C}\right) \leq O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right).$$

This in particular means that Algorithm 8 may return TRUE *even for a set* $\{\widehat{\mu}_1,\dots,\widehat{\mu}_k\}$ *that is* ***not*** *a good approximation to* $\{\mu_1,\dots,\mu_k\}$*.*

Because of that, once we know that COMPUTEORDEREDPARTITION *invoked with* $\{\mu_1,\dots,\mu_k\}$ *returns an ordered partition* $(T_1,\dots,T_b)$ *that induces a collection of clusters* $\{\widehat{C}_1,\dots,\widehat{C}_k\}$*, when proving the final result of this section (Theorem 7) the only thing we assume about* $\widehat{C}$*'s is that they passed the outer-conductance test. And that is why we use Lemma 16 and we "loose" a factor* $\frac{1}{\varphi}$ *in the final guarantee.*

Moreover structuring the argument in this way helps the presentation as later, in Section 6.5, the proof will follow a similar structure.

The following Theorem concludes this subsection by showing (147). It does so by induction using Lemma 37 as an inductive step. At the end it uses Lemma 16 to go from the guarantees for outer-conductance to guarantees for recovery.

Theorem 7. *Let* $k \geq 2$*,* $\varphi \in (0,1)$ *and* $\frac{\epsilon}{\varphi^2}\log(k)$ *be smaller than a sufficiently small constant. Let* $G = (V,E)$ *be a d-regular graph that admits a* (k,φ,ϵ)*-clustering* $\{C_1,\dots,C_k\}$*.*

If COMPUTEORDEREDPARTITION$(G,\widehat{\mu}_1,\widehat{\mu}_2,\dots,\widehat{\mu}_k,s_1,s_2)$ *is invoked with* $(\hat{\mu}_1,\dots,\hat{\mu}_k) = (\mu_1,\dots,\mu_k)$ *and we assume that all tests Algorithm 8 performs* $\left(\text{i.e. } \left\langle f_x, \widehat{\Pi}\widehat{\mu}\right\rangle_{apx} \overset{?}{\geq} 0.93\left\|\widehat{\Pi}\widehat{\mu}\right\|^2_{apx}\right)$ *are exact and* OUTERCONDUCTANCE *computes outer-conductance precisely then the following conditions hold.*

COMPUTEORDEREDPARTITION *returns* $(\text{TRUE},(T_1,\dots,T_b))$ *such that* $(T_1,\dots,T_b)$ *induces a collection of clusters* $\{\widehat{C}_{\mu_1},\dots,\widehat{C}_{\mu_k}\}$ *such that there exists a permutation* π *on* k *elements such that for all* $i \in [k]$*:*

$$\left|\widehat{C}_{\mu_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3}\cdot\log(k)\right)|C_{\pi(i)}|$$

and

$$\phi(\widehat{C}_{\mu_i}) \leq O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right).$$

Proof. Note that for $i = 0$ in the for loop in line 2 of ComputeOrderedPartition S and clusters $\{C_1, \dots, C_k\}$ trivially satisfy assumptions of Lemma 37. So using Lemma 37 and induction we get that for every $i \in [0..\lceil \log(k) \rceil]$ at the beginning of the i-th iteration:

- $|S| \le k/2^i$,
- for every $\mu \in S$ and the corresponding cluster C we have $|V^{(T_1,\dots,T_{i-1})} \cap C| \ge \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$ (where Υ is the constant from the statement of Lemma 37).

In particular this means that after at most $\lceil \log(k) \rceil$ iterations set S becomes empty. This also means that ComputeOrderedPartition returns in line 10, so it returns True and the ordered partial partition $(T_1, \dots, T_b)$ is in fact an ordered partition of $\{\mu_1, \dots, \mu_k\}$.

Note that by definition (see Definition 10) all the approximate clusters $\{\widehat{C}_{\mu_1}, \dots, \widehat{C}_{\mu_k}\}$ are pairwise disjoint and moreover for every constructed cluster $\widehat{C} \in \{\widehat{C}_{\mu_1}, \dots, \widehat{C}_{\mu_k}\}$ we have:

$$\phi(\widehat{C}) \le O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right),$$

as it passed the test in line 6 of ComputeOrderedPartition. So by Lemma 16 it means that there exists a permutation π on k elements such that for all $i \in [k]$:

$$\left|\widehat{C}_{\mu_i} \triangle C_{\pi(i)}\right| \le O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right)|C_{\pi(i)}|.$$

□

6.4 Finding the cluster means

In the previous subsection we showed that ComputeOrderedPartition succeeds if we have access to real cluster centers (i.e. μ_i's). In this section we present a search procedure for finding the centers.

The main idea behind our algorithm is to guess the clustering assignment of few random nodes and use this assignment to compute the approximate cluster means. More precisely, the first step of our algorithm is to learn the spectral embedding as described in Section 5. Then we sample $s = \Omega(\frac{\varphi^2}{\epsilon} \cdot k^4 \log(k))$ random nodes and we consider all the possible clustering assignments for them. For each assignment, we implicitly define the cluster center for a specific cluster as $\widehat{\mu}_i := \frac{1}{|P_i|}\sum_{x \in P_i} f_x$.

Remark 10. *We note that in* FindCenters *we don't necessarily find $\mu_1, \dots, \mu_k$ exactly but we are able to show (see Section 6.4.1) that it finds a good approximation to μ_i's. Then in Section 6.5 we show that such approximation is sufficient for the partitioning scheme to work.*

Algorithm 10 FindCenters(G, η, δ)

```
InitializeOracle(G, δ)
for t ∈ [1 ... log(2/η)] do
    S := Random sample of vertices of V of size s = Θ((φ²/ε) k⁴ log(k))
    for (P_1, P_2, ..., P_k) ∈ Partitions(S) do
        for i = 1 to k do
            μ̂_i := (1/|P_i|) Σ_{x∈P_i} f_x          ▷ Note that we compute the centers only implicitly.
        (r, C) :=
        ComputeOrderedPartition(G, (μ̂_1, μ̂_2, ..., μ̂_k), Θ((φ²/ε) k⁵ log²(k) log(1/η)), Θ((φ⁴/ε²) k⁵ log²(k) log(1/η)))
        if r = True then
            return C
```

6.4.1 Quality of cluster means approximation

In the previous Section 6.3 we showed that the partitioning scheme works if we can find $\mu_1, \dots, \mu_k$ exactly. In this section we show that it is possible to estimate the cluster means with a small error factor (i.e $\mu_i \approx \widehat{\mu}_i$). Later in Section 6.5 we show that such an approximation to μ_i's is enough for the partitioning scheme to work.

In the rest of this section we show that if PARTITIONS(S) (see Algorithm 10) computes a correct guess of cluster assignments then the cluster means computed in line (6) are close to the real cluster means with constant probability. Then we repeat the procedure $O(\log(1/\eta))$ times to achieve success probability of at least $1-\eta$.

In particular, in Lemma 39 we show using Matrix Bernstein that if we have enough samples in a cluster i then $\|\mu_i - \widehat{\mu}_i\|_2 \leq \zeta \cdot \|\mu_i\|_2$. Then we prove that if we sample enough random nodes we have enough samples in every cluster.

Before proving Lemma 39 we show a tail bound for the spectral projection of a node that will be useful to apply Matrix Bernstein.

Lemma 38. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2}\log(k)$ be smaller than a sufficiently small constant. Let $G=(V,E)$ be a d-regular and a (k,φ,ϵ)-clusterable graph. Let $\beta > 1$ .Let*

$$T = \left\{ x \in V : ||f_x||_\infty \geq \beta \cdot \sqrt{\frac{10}{\min_{i\in[k]}|C_i|}} \right\}.$$

Then we have $|T| \leq k \cdot \left(\frac{\beta}{2}\right)^{-\varphi^2/20\cdot\epsilon} \cdot (\min_{i\in[k]}|C_i|)$.

Proof. Recall that $f_x = U_{[k]}^T \mathbb{1}_x$, and u_i denote the i^{th} column of $U_{[k]}$. Thus we have $\|f_x\|_\infty = \max_{i\in[k]}\{u_i(x)\}$. Let $s_{\min} = \min_{i\in k}|C_i|$. We define

$$T_i = \left\{ x \in V : |u_i(x)| \geq \beta \cdot \sqrt{\frac{10}{s_{\min}}} \right\}$$

Therefore, by Lemma 4 we have $|T_i| \leq \left(\frac{\beta}{2}\right)^{-\varphi^2/20\cdot\epsilon} \cdot s_{\min}$. Note that $T = \bigcup_{i=1}^k T_i$. Therefore we have

$$|T| \leq k \cdot \left(\frac{\beta}{2}\right)^{-\varphi^2/20\cdot\epsilon} \cdot s_{\min}$$

□

Now we are ready to derive a bound on the difference between μ_i and $\widehat{\mu}_i$.

Lemma 39. *Let $\zeta,\delta \in (0,1)$, $k \geq 2$, $\varphi \in (0,1)$, $\frac{\epsilon \log k}{\varphi^2}$ be smaller than a positive sufficiently small constant. Let $G=(V,E)$ be a d-regular graph that admits a (k,φ,ϵ)-clustering $C_1,\ldots,C_k$. Let $s \geq c \cdot \left(k \cdot \log\left(\frac{k}{\delta}\right) \cdot \left(\frac{1}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)} \cdot \left(\frac{1}{\zeta}\right)^2\right)^{1/(1-(80\cdot\epsilon/\varphi^2))}$ for large enough constant c. Let $S=\{x_1,x_2,\ldots,x_s\}$ be the multiset with s vertices sampled uniformly at random from cluster C. Let $\mu = \frac{1}{|C|}\sum_{x\in C} f_x$ denote the cluster mean, and let $\widehat{\mu} = \frac{1}{s}\sum_{i=1}^s f_x$ denote the empirical cluster mean. Then with probability at least $1-\delta$ we have*

$$\|\mu - \widehat{\mu}\|_2 \leq \zeta \cdot \|\mu\|_2$$

Proof. Let $s_{\min} := \min_{i\in[k]}|C_i|$. We define

$$C' = \left\{ x \in C : ||f_x||_\infty \leq 2 \cdot \left(\frac{s\cdot k}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)} \cdot \sqrt{\frac{10}{s_{\min}}} \right\}$$

Note that by Lemma 38 and by choice of $\beta = 2\cdot\left(\frac{s\cdot k}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)}$ we have

$$|C \setminus C'| \leq k \cdot \left(\frac{\beta}{2}\right)^{-\varphi^2/(20\cdot\epsilon)} \cdot s_{\min} \leq k \cdot \left(\frac{s\cdot k}{\delta}\right)^{-2} \cdot |C| = (k^{-1}\cdot s^{-2}\cdot\delta^2)\cdot|C|$$

Thus we have

$$|C'| \geq \left(1 - (k^{-1}\cdot s^{-2}\cdot\delta^2)\right)|C| \tag{170}$$

Let $\mu' = \frac{1}{|C'|}\sum_{x\in C'} f_x$. By triangle inequality we have

$$||\widehat{\mu} - \mu||_2 \leq ||\widehat{\mu} - \mu'||_2 + ||\mu' - \mu||_2 \tag{171}$$

In the rest of the proof we will upper bound both of these terms by $\frac{\zeta}{2} \cdot ||\mu||_2$.

Step 1: We first prove $||\widehat{\mu} - \mu'||_2 \leq \frac{\zeta}{2} \cdot ||\mu||_2$. By the assumption of the lemma for sufficiently small $\frac{\epsilon \log k}{\varphi^2}$ we have $k^{(40\cdot\epsilon/\varphi^2)} \leq 2$. Thus for any $x \in C'$ we have $||f_x||_\infty \leq \left(\frac{s}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)} \cdot \sqrt{\frac{160}{s_{\min}}}$. Therefore by triangle inequality we have

$$||\mu'||_2 = \left|\left| \frac{1}{|C'|} \cdot \sum_{x\in C'} f_x \right|\right| \leq \frac{1}{|C'|} \cdot \sum_{x\in C'} ||f_x||_2 \leq \frac{\sqrt{k}}{|C'|} \cdot \sum_{x\in C'} ||f_x||_\infty \leq \left(\frac{s}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)} \cdot \sqrt{\frac{160\cdot k}{s_{\min}}}. \tag{172}$$

By (170) and by union bound over all samples in S with probability at least $1 - s\cdot(k^{-1}\cdot s^{-2}\cdot\delta^2) = 1 - s^{-1}\cdot k^{-1}\cdot\delta^2 \geq 1 - \frac{\delta}{2}$ for all $x_i \in S$ we have $x_i \in C'$, hence, $||f_x||_\infty \leq \left(\frac{s}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)} \cdot \sqrt{\frac{160}{s_{\min}}}$. Thus with probability at least $1-\frac{\delta}{2}$, S is chosen uniformly at random from C' so for all $x_i \in S$ we have

$$|f_x||_\infty \leq \left(\frac{s}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)} \cdot \sqrt{\frac{160}{s_{\min}}} \tag{173}$$

In the rest of the proof of step 1 we assume $S \subseteq C'$ which holds with probability at least $1-\frac{\delta}{2}$. Therefore conditioned on $S \subseteq C'$ we have $\mathbb{E}[f_{x_i}] = \mu'$.

$$\|\widehat{\mu} - \mu'\|_2 = \left\| \sum_{i=1}^{s} \left(\frac{f_{x_i}}{s} - \mu' \right) \right\|_2 .$$

We define $\mathbf{z}_i = \frac{f_{x_i}}{s} - \frac{\mu'}{s}$, so $\|\widehat{\mu} - \mu'\|_2 = \|\sum_{i=1}^{s}\mathbf{z}_i\|_2$. Observe that $\mathbb{E}\left[\mathbf{z}_i\right] = \mathbb{E}\left[\frac{f_{x_i}}{s}\right] - \frac{\mu'}{s} = 0$, thus we can apply Lemma 20. Therefore we get

$$\mathbb{P}\left[||\widehat{\mu} - \mu'||_2 > q\right] = \mathbb{P}\left[\|\sum_{i=1}^{s}\mathbf{z}_i\|_2 > q\right] \leq (k+1)\cdot\exp\left(\frac{\frac{-q^2}{2}}{\sigma^2 + \frac{bq}{3}}\right), \tag{174}$$

where $\sigma^2 = \max\{\|\sum_{i=1}^{s}\mathbb{E}[\mathbf{z}_i\mathbf{z}_i^T]\|_2, \|\sum_{i=1}^{s}\mathbb{E}[\mathbf{z}_i^T\mathbf{z}_i]\|_2\}$ and b is an upper bound on $\|\mathbf{z}_i\|_2$ for all random variables $\mathbf{z}_i$. Therefore we need to upperbound $\|\mathbf{z}_i\|_2$ and σ^2. Note that

$$\|\mathbf{z}_i\|_2 = \left\|\frac{f_{x_i}}{s} - \frac{\mu'}{s}\right\|_2 \leq \left\|\frac{f_{x_i}}{s}\right\|_2 + \left\|\frac{\mu'}{s}\right\|_2 \leq \frac{\sqrt{k}}{s}\cdot\|f_{x_i}\|_\infty + \frac{1}{s}\cdot||\mu'||_2 \tag{175}$$

Therefore by (172), (173) and (175) we have

$$\|\mathbf{z}_i\|_2 \leq \frac{\sqrt{k}}{s}\cdot\|f_{x_i}\|_\infty + \frac{1}{s}\cdot||\mu'||_2 \leq \frac{2}{s}\cdot\left(\frac{s}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)}\cdot\sqrt{\frac{160\cdot k}{s_{\min}}}, \tag{176}$$

Thus $b \leq \frac{2}{s}\cdot\left(\frac{s}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)}\cdot\sqrt{\frac{160\cdot k}{s_{\min}}}$. We also need to upper bound σ^2. By (176) we get

$$\sigma^2 = \max\{\|\sum_{i=1}^{s}\mathbb{E}[\mathbf{z}_i\mathbf{z}_i^T]\|_2, \|\sum_{i=1}^{s}\mathbb{E}[\mathbf{z}_i^T\mathbf{z}_i]\|_2\} = s\cdot\mathbb{E}\left[\|\mathbf{z}_i\|_2^2\right] \leq s\cdot\frac{4}{s^2}\cdot\left(\frac{s}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)}\cdot\frac{160\cdot k}{s_{\min}}. \tag{177}$$

We set $q = \frac{\zeta}{2}\cdot||\mu||_2$. Having upper bound for σ^2 by (177) and on b by (176) we can apply Lemma 20 and we get

$$\begin{aligned}
\mathbb{P}\left[||\widehat{\mu} - \mu'||_2 > \frac{\zeta}{2}\cdot||\mu||_2\right] &\leq (k+1)\cdot\exp\left(\frac{\frac{-q^2}{2}}{\sigma^2 + \frac{bq}{3}}\right) \\
&\leq (k+1)\cdot\exp\left(\frac{\frac{-\zeta^2\cdot||\mu||_2^2}{8}}{\frac{640\cdot k\cdot\left(\frac{s}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)}}{s\cdot s_{\min}} + \frac{\zeta}{2}\cdot||\mu||_2\cdot\frac{2\cdot\left(\frac{s}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)}}{3\cdot s}\sqrt{\frac{160\cdot k}{s_{\min}}}}\right)
\end{aligned} \tag{178}$$

By Lemma 7 for small enough $\frac{\epsilon}{\varphi^2}$ we have $\|\mu\|_2^2 \geq \frac{1}{2\cdot|C|}$ and since $\min_{i,j} \frac{|C_i|}{|C_j|} \geq \Omega(1)$. Thus for a small enough constant c' we have

$$s_{\min} \cdot ||\mu||_2^2 \geq \frac{s_{\min}}{2 \cdot |C|} \geq c', \tag{179}$$

Thus by (179) and by choice of $s^{(1-80\cdot\epsilon/\varphi^2)} \geq \frac{10^6}{c'} \cdot k \cdot \log\left(\frac{k}{\delta}\right) \cdot \left(\frac{1}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)} \cdot \left(\frac{1}{\zeta}\right)^2 \geq \frac{10^6 \cdot k \cdot \log\left(\frac{k}{\delta}\right) \cdot \left(\frac{1}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)} \cdot \left(\frac{1}{\zeta}\right)^2}{s_{\min} \cdot ||\mu||_2^2}$ we get

$$\frac{\zeta^2 \cdot ||\mu||_2^2}{8} \geq 400 \cdot \log\left(\frac{k}{\delta}\right) \cdot \left(\frac{640 \cdot k \cdot \left(\frac{s}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)}}{s \cdot s_{\min}}\right) \tag{180}$$

and

$$\frac{\zeta^2 \cdot ||\mu||_2^2}{8} \geq 400 \cdot \log\left(\frac{k}{\delta}\right) \left(\frac{\zeta}{2} \cdot ||\mu||_2 \cdot \frac{2 \cdot \left(\frac{s}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)}}{3 \cdot s} \sqrt{\frac{160 \cdot k}{s_{\min}}}\right) \tag{181}$$

Therefore since $s \geq c \cdot \left(k \cdot \log\left(\frac{k}{\delta}\right) \cdot \left(\frac{1}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)} \cdot \left(\frac{1}{\zeta}\right)^2\right)^{1/(1-(80\cdot\epsilon/\varphi^2))}$ for large enough constant c, and putting (178), (180) and (181) together we get

$$\mathbb{P}\left[||\widehat{\mu} - \mu'||_2 > \frac{\zeta}{2} \cdot ||\mu||_2\right] \leq (k+1) \cdot e^{-200\cdot\log\left(\frac{k}{\delta}\right)} \leq \left(\frac{\delta}{k}\right)^{100}$$

Thus with probability at least $1 - \frac{\delta}{2} - \left(\frac{\delta}{k}\right)^{100} \geq 1 - \delta$ we have

$$\|\widehat{\mu} - \mu'\|_2 \leq \frac{\zeta}{2} \cdot \|\mu\|_2. \tag{182}$$

Step 2: Next we want to bound $\|\mu - \mu'\|_2$. We have

$$\begin{aligned}
\|\mu' - \mu\|_2 &= \left\|\frac{1}{|C'|}\sum_{x\in C'} f_x - \frac{1}{|C|}\sum_{x\in C} f_x\right\|_2 \\
&\leq \left\|\frac{1}{|C'|}\sum_{x\in C} f_x - \frac{1}{|C|}\sum_{x\in C} f_x\right\|_2 + \left\|\frac{1}{|C'|}\sum_{x\in C\setminus C'} f_x\right\|_2 && \text{By triangle inequality} \\
&\leq \left(\frac{1}{1-(k^{-1}\cdot s^{-2}\cdot\delta^2)} - 1\right)||\mu||_2 + \left\|\frac{1}{|C'|}\sum_{x\in C\setminus C'} f_x\right\|_2 && \text{Since } |C'| \geq \left(1 - (k^{-1}\cdot s^{-2}\cdot\delta^2)\right)|C| \text{ by (170)} \\
&\leq 2\cdot(k^{-1}\cdot s^{-2}\cdot\delta^2)\cdot\|\mu\|_2 + \left\|\frac{1}{|C'|}\sum_{x\in C\setminus C'} f_x\right\|_2 && (183)
\end{aligned}$$

It thus remains to upper bound the second term. We now note that

$$\left\|\frac{1}{|C'|}\sum_{x\in C\setminus C'} f_x\right\|_2 \leq \frac{1}{|C'|}\sum_{x\in C\setminus C'} ||f_x||_2 \leq \frac{\sqrt{k}}{|C'|}\sum_{x\in C\setminus C'} \|f_x\|_\infty \tag{184}$$

For any $y \geq 1$ we define

$$T(y) = \left\{x \in V : ||f_x||_\infty \geq 2 \cdot y \cdot \left(\frac{s\cdot k}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)} \cdot \sqrt{\frac{10}{s_{\min}}}\right\}$$

Therefore, by Lemma 38 we have

$$|T(y)| \leq k \cdot \left(\frac{2 \cdot y \cdot \left(\frac{s\cdot k}{\delta}\right)^{(40\cdot\epsilon/\varphi^2)}}{2}\right)^{-\varphi^2/(20\cdot\epsilon)} \cdot s_{\min} = \left(\frac{s\cdot k}{\delta}\right)^{-2} \cdot y^{-\varphi^2/(20\cdot\epsilon)} \cdot s_{\min}. \tag{185}$$

Using the bound on $|T(y)|$ above, we now get

$$\sum_{x \in C \setminus C'} \|f_x\|_\infty \tag{186}$$

$$\leq \int_1^\infty \left(y \cdot \left(\frac{s \cdot k}{\delta}\right)^{(40 \cdot \epsilon/\varphi^2)} \cdot \sqrt{\frac{40}{s_{\min}}} \right) \cdot |T(y)| \cdot dy \qquad \text{By definition of } T(y) \text{ and } C'$$

$$\leq \sqrt{\frac{160}{s_{\min}}} \cdot \left(\frac{s}{\delta}\right)^{(40 \cdot \epsilon/\varphi^2)} \cdot \int_1^\infty y \cdot |T(y)| \cdot dy \qquad \text{Since } k^{(40 \cdot \epsilon/\varphi^2)} \leq 2 \text{ for small enough } \frac{\epsilon \cdot \log k}{\varphi^2}$$

$$\leq \sqrt{\frac{160}{s_{\min}}} \cdot \left(\frac{s}{\delta}\right)^{(40 \cdot \epsilon/\varphi^2)} \cdot \int_1^\infty \left(\frac{s \cdot k}{\delta}\right)^{-2} \cdot y^{\left(1-\varphi^2/(20 \cdot \epsilon)\right)} \cdot s_{\min} \cdot dy \qquad \text{By (185)}$$

$$\leq \sqrt{\frac{160}{s_{\min}}} \cdot s_{\min} \cdot \left(\frac{s}{\delta}\right)^{(40 \cdot \epsilon/\varphi^2)} \cdot \left(\frac{s \cdot k}{\delta}\right)^{-2} \frac{1}{\varphi^2/(20 \cdot \epsilon) - 2} \qquad \text{Since for any } c < 0, \int_1^\infty y^c dy = \frac{-1}{c+1}$$

$$\leq k^{-2} \cdot s^{-1} \cdot \sqrt{s_{\min}} \qquad \text{For small enough } \frac{\epsilon}{\varphi^2} \tag{187}$$

Therefore we get

$$\begin{aligned}
\left\| \frac{1}{|C'|} \sum_{x \in C \setminus C'} f_x \right\|_2 &\leq \frac{\sqrt{k}}{|C'|} \sum_{x \in C \setminus C'} \|f_x\|_\infty && \text{By (184)} \\
&\leq \frac{\sqrt{k} \cdot k^{-2} \cdot s^{-1} \cdot \sqrt{s_{\min}}}{|C'|} && \text{By (187)} \\
&\leq \frac{2 \cdot k^{-1} \cdot s^{-1}}{\sqrt{|C|}} \cdot \frac{\sqrt{s_{\min}}}{\sqrt{|C|}} && \text{By (170)} \\
&\leq \frac{k^{-1} \cdot s^{-1}}{\sqrt{|C|}} && \text{Since } |C| \geq s_{\min} \\
&\leq 2 \cdot k^{-1} \cdot s^{-1} \cdot ||\mu||_2 && \text{By Lemma 7 } ||\mu||_2 \geq \frac{1}{2 \cdot \sqrt{|C|}}
\end{aligned}$$

Therefore by (183) we have

$$\|\mu' - \mu\|_2 \leq 2 \cdot (k^{-1} \cdot s^{-2} \cdot \delta^2) \|\mu\|_2 + \left\| \frac{1}{|C'|} \sum_{x \in C \setminus C'} f_x \right\|_2 \leq 2 \left(k^{-1} \cdot s^{-2} \cdot \delta^2 + \cdot k^{-1} \cdot s^{-1}\right) ||\mu||_2 \leq \frac{\zeta}{2} \cdot ||\mu||_2 \tag{188}$$

The last inequality holds since $s \geq 8 \cdot \left(\frac{1}{\zeta}\right)^2$, hence, $2\left(k^{-1} \cdot s^{-2} \cdot \delta^2 + \cdot k^{-1} \cdot s^{-1}\right) \leq \frac{\zeta}{2}$. Putting (171), (182) and (188) together with probability at least $1 - \delta$ we get

$$||\widehat{\mu} - \mu||_2 \leq ||\widehat{\mu} - \mu'||_2 + ||\mu' - \mu||_2 \leq \frac{\zeta}{2} \cdot ||\mu||_2 + \frac{\zeta}{2} \cdot ||\mu||_2 \leq \zeta \cdot ||\mu||_2$$

□

To conclude our argument we show that if we sample enough nodes, we have a large number of samples in each cluster.

Lemma 40. *Let $k \geq 2$, $\varphi \in (0,1)$, $\frac{\epsilon \log k}{\varphi^2}$ be smaller than a positive sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Let S be the multiset of $s \in \Omega(k \log k)$ vertices each sampled independently at random from V. Then with probability at least $\frac{9}{10}$, for every $i \in [k]$,*

$$|S \cap C_i| \geq \frac{0.9 \cdot s}{k} \cdot \min_{p,q \in [k]} \frac{|C_p|}{|C_q|}.$$

Proof. For $i \in [k]$, and $1 \leq r \leq s$, let X_i^r be a random variable which is 1 if the r-th sampled vertex is in C_i, and 0 otherwise. Thus $\mathbb{E}[X_i^r] = \frac{|C_i|}{n}$. Observe that $|S \cap C_i|$ is a random variable defined as $\sum_{r=1}^s X_i^r$, where its expectation is given by

$$\mathbb{E}[|S \cap C_i|] = \sum_{r=1}^s \mathbb{E}[X_i^r] = s \cdot \frac{|C_i|}{n} \geq \frac{s \cdot s_{\min}}{k \cdot s_{\max}}.$$

Notice that random variables X_i^r are independent, Therefore, by Chernoff bound,

$$\Pr\left[|S\cap C_i| < \frac{9s}{10}\cdot\frac{|C_i|}{n}\right] \le \exp\left(-\frac{1}{200}\cdot\frac{s\cdot s_{\min}}{k\cdot s_{\max}}\right).$$

By union bound and since $s = 500\cdot k\cdot\log k\cdot\frac{s_{\max}}{s_{\min}}$ we have

$$\Pr\left[\exists i\colon\ |S\cap C_i| < \frac{9s}{10}\cdot\frac{|C_i|}{n}\right] \le k\cdot\exp\left(-\frac{1}{200}\cdot\frac{s\cdot s_{\min}}{k\cdot s_{\max}}\right) \le \frac{1}{10}.$$

Therefore with probability at least $\frac{9}{10}$ for all $i\in[k]$ we have

$$|S\cap C_i| \ge \frac{9\cdot s}{10}\cdot\frac{|C_i|}{n} \ge \frac{0.9\cdot s}{k}\cdot\frac{s_{\min}}{s_{\max}}$$

□

6.4.2 Approximate Centers are strongly orthogonal

The main result of this section is Lemma 41 that generalizes Lemma 12 to the approximate of cluster means.

Lemma 41. *Let $k\ge 2$ be an integer, $\varphi\in(0,1)$, and $\epsilon\in(0,1)$. Let $G=(V,E)$ be a d-regular graph that admits a (k,φ,ϵ)-clustering $C_1,\dots,C_k$. Let $0<\zeta<\frac{\sqrt{\epsilon}}{20\cdot k\cdot\varphi}$. Let $\mu_1,\dots,\mu_k$ denote the cluster means of $C_1,\dots,C_k$. Let $\widehat{\mu}_1,\dots,\widehat{\mu}_k\in\mathbb{R}^k$ denote an approximation of the cluster means such that for each $i\in[k]$, $||\mu_i-\widehat{\mu}_i||_2\le\zeta||\mu_i||_2$. Let $S\subset\{\widehat{\mu}_1,\dots,\widehat{\mu}_k\}$ denote a subset of cluster means. Let $\widehat{\Pi}\in\mathbb{R}^{k\times k}$ denote the orthogonal projection matrix into the span$(S)^\perp$. Then the following holds:*

1. *For all $\widehat{\mu}_i\in\{\widehat{\mu}_1,\dots,\widehat{\mu}_k\}\setminus S$ we have $\left|\|\widehat{\Pi}\widehat{\mu}_i\|_2^2-||\widehat{\mu}_i||_2^2\right|\le\frac{20\sqrt{\epsilon}}{\varphi}\cdot||\widehat{\mu}_i||_2^2$.*
2. *For all $\widehat{\mu}_i\ne\widehat{\mu}_j\in\{\widehat{\mu}_1,\dots,\widehat{\mu}_k\}\setminus S$ we have $|\langle\widehat{\Pi}\widehat{\mu}_i,\widehat{\Pi}\widehat{\mu}_j\rangle|\le\frac{50\sqrt{\epsilon}}{\varphi}\cdot\frac{1}{\sqrt{|C_i|\cdot|C_j|}}$.*

To prove Lemma 41 we use Lemma 30 from Section 4 and we prove Lemma 42.

Lemma 42. *Let $k\ge 2$ be an integer, $\varphi\in(0,1)$, and $\epsilon\in(0,1)$. Let $G=(V,E)$ be a d-regular graph that admits a (k,φ,ϵ)-clustering $C_1,\dots,C_k$. Let $0<\zeta<\frac{\sqrt{\epsilon}}{20\cdot k\cdot\varphi}$. Let $\widehat{\mu}_1,\dots,\widehat{\mu}_k\in\mathbb{R}^k$ denote an approximation of the cluster means such that for each $i\in[k]$, $||\mu_i-\widehat{\mu}_i||_2\le\zeta||\mu_i||_2$. Let $S=\{\widehat{\mu}_1,\dots,\widehat{\mu}_k\}\setminus\{\widehat{\mu}_i\}$. Let $\widehat{H}=[\widehat{\mu}_1,\widehat{\mu}_2,\dots,\widehat{\mu}_{i-1},\widehat{\mu}_{i+1},\dots,\widehat{\mu}_k]$ denote a matrix such that its columns are the vectors in S. Let $\widehat{W}\in\mathbb{R}^{(k-1)\times(k-1)}$ denote a diagonal matrix such that for all $j<i$ we have $\widehat{W}(j,j)=\sqrt{|C_j|}$ and for all $j\ge i$ we have $\widehat{W}(j,j)=\sqrt{|C_{j+1}|}$. Let $\widehat{Z}=\widehat{H}\widehat{W}$. Then we have*

$$\widehat{\mu}_i^T\widehat{Z}\widehat{Z}^T\widehat{\mu}_i\le\frac{10\sqrt{\epsilon}}{\varphi}\cdot||\widehat{\mu}_i||_2^2.$$

Proof. Note that $\widehat{Z}\widehat{Z}^T=(\sum_{j=1}^k|C_j|\widehat{\mu}_j\widehat{\mu}_j^T)-|C_i|\widehat{\mu}_i\widehat{\mu}_i^T$. Thus we have

$$\widehat{\mu}_i^T\widehat{Z}\widehat{Z}^T\widehat{\mu}_i=\widehat{\mu}_i^T\left(\sum_{j=1}^k|C_j|\widehat{\mu}_j\widehat{\mu}_j^T\right)\widehat{\mu}_i-|C_i|\cdot||\widehat{\mu}_i||_2^4. \tag{189}$$

By Lemma 9 for any vector x with $||x||_2=1$ we have

$$x^T\left(\sum_{j=1}^k|C_j|\mu_j\mu_j^T-I\right)x\le\frac{4\sqrt{\epsilon}}{\varphi} \tag{190}$$

Note that

$$
\begin{aligned}
& || \sum_{j=1}^{k} |C_j| \widehat{\mu}_j \widehat{\mu}_j^T - \sum_{j=1}^{k} |C_j| \mu_j \mu_j^T ||_2 \\
& \leq \sum_{j=1}^{k} |C_j| \cdot || \widehat{\mu}_j \widehat{\mu}_j^T - \mu_j \mu_j^T ||_2 && \text{By triangle inequality} \\
& = \sum_{j=1}^{k} |C_j| \left(|| \left(\mu_j + (\widehat{\mu}_j - \mu_j) \right) \left(\mu_j + (\widehat{\mu}_j - \mu_j) \right)^T - \mu_j \mu_j^T ||_2 \right) \\
& \leq \sum_{j=1}^{k} |C_j| \left(|| \left(\widehat{\mu}_j - \mu_j \right) \left(\widehat{\mu}_j - \mu_j \right)^T ||_2 + || \mu_j \left(\widehat{\mu}_j - \mu_j \right)^T ||_2 + || \left(\widehat{\mu}_j - \mu_j \right) \mu_j^T ||_2 \right) && \text{By triangle inequality} \\
& \leq \sum_{j=1}^{k} |C_j| \cdot (\zeta^2 + 2\zeta) \cdot ||\mu_j||_2^2 && \text{Since } ||\widehat{\mu}_j - \mu_j||_2 \leq \zeta ||\mu_j||_2 \\
& \leq \sum_{j=1}^{k} |C_j| \cdot 6 \cdot \zeta \cdot \frac{1}{|C_j|} && \text{By Lemma 7 } ||\mu_j||_2^2 \leq \frac{2}{|C_i|} \\
& \leq 6 \cdot \zeta \cdot k \\
& \leq \frac{\sqrt{\epsilon}}{2\varphi} && \text{Since } \zeta \leq \frac{\sqrt{\epsilon}}{20 \cdot k \cdot \varphi}
\end{aligned}
$$

Thus for any vector x with $||x||_2 = 1$ we have

$$
x^T \left(\sum_{j=1}^{k} |C_j| \widehat{\mu}_j \widehat{\mu}_j^T - \sum_{j=1}^{k} |C_j| \mu_j \mu_j^T \right) x \leq \frac{\sqrt{\epsilon}}{2\varphi} \tag{191}
$$

Putting (191) and (190) for any vector any vector x with $||x||_2 = 1$ we have that

$$
x^T \left(\sum_{j=1}^{k} |C_j| \widehat{\mu}_j \widehat{\mu}_j^T - I \right) x \leq \frac{5\sqrt{\epsilon}}{\varphi}
$$

Hence we can write

$$
\widehat{\mu}_i^T \left(\sum_{j=1}^{k} |C_j| \widehat{\mu}_j \widehat{\mu}_j^T \right) \widehat{\mu}_i = \widehat{\mu}_i^T \left(\sum_{j=1}^{k} |C_j| \widehat{\mu}_j \widehat{\mu}_j^T - I \right) \widehat{\mu}_i + \widehat{\mu}_i^T \widehat{\mu}_i \leq \left(1 + \frac{5\sqrt{\epsilon}}{\varphi} \right) ||\widehat{\mu}_i||_2^2
$$

Therefore by (189) we get

$$
\widehat{\mu}_i^T \widehat{Z} \widehat{Z}^T \widehat{\mu}_i = \widehat{\mu}_i^T \left(\sum_{j=1}^{k} |C_j| \widehat{\mu}_j \widehat{\mu}_j^T \right) \widehat{\mu}_i - |C_i| \cdot ||\widehat{\mu}_i||_2^4 \leq \left(1 + \frac{5\sqrt{\epsilon}}{\varphi} - |C_i| \cdot ||\widehat{\mu}_i||_2^2 \right) ||\widehat{\mu}_i||_2^2
$$

By Lemma 7, and since $||\widehat{\mu}_i|| \geq (1-\zeta)||\mu_i||_2$ and $\zeta \leq \frac{\sqrt{\epsilon}}{20 \cdot k \cdot \varphi}$ we have that

$$
|C_i| \cdot ||\widehat{\mu}_i||_2^2 \geq \left(1 - \frac{4\sqrt{\epsilon}}{\varphi} \right) (1-\zeta)^2 \geq 1 - \frac{5\sqrt{\epsilon}}{\varphi}
$$

Thus we get

$$
\widehat{\mu}_i^T \widehat{Z} \widehat{Z}^T \widehat{\mu}_i \leq \left(1 + \frac{5\sqrt{\epsilon}}{\varphi} - |C_i| \cdot ||\widehat{\mu}_i||_2^2 \right) ||\widehat{\mu}_i||_2^2 \leq \left(1 + \frac{5\sqrt{\epsilon}}{\varphi} - 1 + \frac{5\sqrt{\epsilon}}{\varphi} \right) ||\widehat{\mu}_i||_2^2 \leq \frac{10\sqrt{\epsilon}}{\varphi} \cdot ||\widehat{\mu}_i||_2^2
$$

□

We now prove the main result of this section (Lemma 41).

Lemma 41. *Let $k \geq 2$ be an integer, $\varphi \in (0,1)$, and $\epsilon \in (0,1)$. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \dots, C_k$. Let $0 < \zeta < \frac{\sqrt{\epsilon}}{20 \cdot k \cdot \varphi}$. Let $\mu_1, \dots, \mu_k$ denote the cluster means of $C_1, \dots, C_k$. Let $\widehat{\mu}_1, \dots, \widehat{\mu}_k \in \mathbb{R}^k$ denote an approximation of the cluster means such that for each $i \in [k]$, $||\mu_i - \widehat{\mu}_i||_2 \leq \zeta ||\mu_i||_2$. Let $S \subset \{\widehat{\mu}_1, \dots, \widehat{\mu}_k\}$ denote a subset of cluster means. Let $\widehat{\Pi} \in \mathbb{R}^{k \times k}$ denote the orthogonal projection matrix into the span$(S)^\perp$. Then the following holds:*

1. *For all $\widehat{\mu}_i \in \{\widehat{\mu}_1, \dots, \widehat{\mu}_k\} \setminus S$ we have $\left| ||\widehat{\Pi}\widehat{\mu}_i||_2^2 - ||\widehat{\mu}_i||_2^2 \right| \leq \frac{20\sqrt{\epsilon}}{\varphi} \cdot ||\widehat{\mu}_i||_2^2$.*

2. *For all $\widehat{\mu}_i \neq \widehat{\mu}_j \in \{\widehat{\mu}_1, \dots, \widehat{\mu}_k\} \setminus S$ we have $|\langle \widehat{\Pi}\widehat{\mu}_i, \widehat{\Pi}\widehat{\mu}_j \rangle| \leq \frac{50\sqrt{\epsilon}}{\varphi} \cdot \frac{1}{\sqrt{|C_i| \cdot |C_j|}}$.*

Proof. **Proof of item (1):** Since $\widehat{\Pi}$ is a orthogonal projection matrix we have $||\widehat{\Pi}||_2 = 1$. Hence, we have

$$||\widehat{\Pi}\widehat{\mu}_i||_2^2 \leq ||\widehat{\mu}_i||_2^2 \leq \left(1 + \frac{20\sqrt{\epsilon}}{\varphi}\right) ||\widehat{\mu}_i||_2^2.$$

Thus it's left to prove $||\widehat{\Pi}\widehat{\mu}_i||_2^2 \geq \left(1 - \frac{20\sqrt{\epsilon}}{\varphi}\right) ||\widehat{\mu}_i||_2^2$. Note that by Pythagoras $||\widehat{\Pi}\widehat{\mu}_i||_2^2 = ||\widehat{\mu}_i||_2^2 - ||(I - \widehat{\Pi})\widehat{\mu}_i||_2^2$. We will prove $||(I - \widehat{\Pi})\widehat{\mu}_i||_2^2 \leq \frac{20\sqrt{\epsilon}}{\varphi} ||\widehat{\mu}_i||_2^2$ which implies

$$||\widehat{\Pi}\widehat{\mu}_i||_2^2 \geq \left(1 - 20\frac{\sqrt{\epsilon}}{\varphi}\right) ||\widehat{\mu}_i||_2^2.$$

Thus in order to complete the proof we need to show $||(I - \widehat{\Pi})\widehat{\mu}_i||_2^2 \leq \frac{20\sqrt{\epsilon}}{\varphi} ||\widehat{\mu}_i||_2^2$. Let $S' = \{\widehat{\mu}_1, \dots, \widehat{\mu}_k\} \setminus \{\widehat{\mu}_i\}$. Let $\widehat{\Pi}'$ denote the orthogonal projection matrix into *span*$(S')^\perp$. Note that $S \subseteq S'$, hence *span*(S) is a subspace of *span*(S'), therefore we have $||(I - \widehat{\Pi})\widehat{\mu}_i||_2^2 \leq ||(I - \widehat{\Pi}')\widehat{\mu}_i||_2^2$. Thus it suffices to prove $||(I - \widehat{\Pi}')\widehat{\mu}_i||_2^2 \leq \frac{20\sqrt{\epsilon}}{\varphi} ||\widehat{\mu}_i||_2^2$. Let $\widehat{H} = [\widehat{\mu}_1, \widehat{\mu}_2, \dots, \widehat{\mu}_{i-1}, \widehat{\mu}_{i+1}, \dots, \widehat{\mu}_k]$ denote a matrix such that its columns are the vectors in S'. Let $\widehat{W} \in \mathbb{R}^{(k-1) \times (k-1)}$ denote a diagonal matrix such that for all $j < i$ we have $\widehat{W}(j,j) = \sqrt{|C_j|}$ and for all $j \geq i$ we have $\widehat{W}(j,j) = \sqrt{|C_{j+1}|}$. Let $\widehat{Z} = \widehat{H}\widehat{W}$. Then the orthogonal projection matrix onto the span of S' is defined as $(I - \widehat{\Pi}') = \widehat{Z}(\widehat{Z}^T\widehat{Z})^{-1}\widehat{Z}^T$. By Lemma 30 item (2), $(\widehat{Z}^T\widehat{Z})^{-1}$ is spectrally close to I, hence, $(\widehat{Z}^T\widehat{Z})^{-1}$ exists. Therefore we have

$$\begin{aligned} ||(I - \widehat{\Pi}')\widehat{\mu}_i||_2^2 &= \widehat{\mu}_i^T \widehat{Z}(\widehat{Z}^T\widehat{Z})^{-1}\widehat{Z}^T\widehat{\mu}_i \\ &= \widehat{\mu}_i^T \widehat{Z}((\widehat{Z}^T\widehat{Z})^{-1} - I)\widehat{Z}^T\widehat{\mu}_i + \widehat{\mu}_i^T \widehat{Z}\widehat{Z}^T\widehat{\mu}_i \end{aligned} \tag{192}$$

By Lemma 30 item (2) we have

$$\left| \widehat{\mu}_i^T \widehat{Z} \left((\widehat{Z}^T\widehat{Z})^{-1} - I \right) \widehat{Z}^T\widehat{\mu}_i \right| \leq \frac{5\sqrt{\epsilon}}{\varphi} ||\widehat{Z}^T\widehat{\mu}_i||_2^2 \tag{193}$$

Thus we get

$$\begin{aligned} ||(I - \widehat{\Pi}')\widehat{\mu}_i||_2^2 &\leq \widehat{\mu}_i^T \widehat{Z}((\widehat{Z}^T\widehat{Z})^{-1} - I)\widehat{Z}^T\widehat{\mu}_i + \widehat{\mu}_i^T \widehat{Z}\widehat{Z}^T\widehat{\mu}_i && \text{By (192)} \\ &\leq \left(\frac{5\sqrt{\epsilon}}{\varphi} + 1\right) ||\widehat{Z}^T\widehat{\mu}_i||_2^2 && \text{By (193)} \\ &\leq 2 \cdot ||\widehat{Z}^T\widehat{\mu}_i||_2^2 && \text{For small enough } \frac{\epsilon}{\varphi^2} \end{aligned}$$

By Lemma 42 we have

$$||\widehat{Z}^T\widehat{\mu}_i||_2^2 = \widehat{\mu}_i^T \widehat{Z}\widehat{Z}^T\widehat{\mu}_i \leq \frac{10\sqrt{\epsilon}}{\varphi} \cdot ||\widehat{\mu}_i||_2^2$$

Therefore we get

$$||(I - \widehat{\Pi})\widehat{\mu}_i||_2^2 \leq ||(I - \widehat{\Pi}')\widehat{\mu}_i||_2^2 \leq 2||\widehat{Z}^T\widehat{\mu}_i||_2^2 \leq \frac{20\sqrt{\epsilon}}{\varphi} ||\widehat{\mu}_i||_2^2 \tag{194}$$

Hence,

$$||\widehat{\Pi}\widehat{\mu}_i||_2^2 \geq \left(1 - 20\frac{\sqrt{\epsilon}}{\varphi} ||\widehat{\mu}_i||_2^2\right).$$

Proof of item (2)**:** Note that

$$\langle \widehat{\mu}_i, \widehat{\mu}_j \rangle = \langle (I - \widehat{\Pi})\widehat{\mu}_i + \widehat{\Pi}\widehat{\mu}_i, (I - \widehat{\Pi})\widehat{\mu}_j + \widehat{\Pi}\widehat{\mu}_j \rangle = \langle (I - \widehat{\Pi})\widehat{\mu}_i, (I - \widehat{\Pi})\widehat{\mu}_j \rangle + \langle \widehat{\Pi}\widehat{\mu}_i, \widehat{\Pi}\widehat{\mu}_j \rangle$$

Thus by triangle inequality we have

$$|\langle \widehat{\Pi}\widehat{\mu}_i, \widehat{\Pi}\widehat{\mu}_j \rangle| \le |\langle \widehat{\mu}_i, \widehat{\mu}_j \rangle| + |\langle (I - \widehat{\Pi})\widehat{\mu}_i, (I - \widehat{\Pi})\widehat{\mu}_j \rangle|$$

By Cauchy-Schwarz we have

$$\begin{aligned}
|\langle (I - \widehat{\Pi})\widehat{\mu}_i, (I - \widehat{\Pi})\widehat{\mu}_j \rangle| &\le ||(I - \widehat{\Pi})\widehat{\mu}_i||_2 ||(I - \widehat{\Pi})\widehat{\mu}_i||_2 \\
&\le \frac{20\sqrt{\epsilon}}{\varphi} ||\widehat{\mu}_i||_2 ||\widehat{\mu}_j||_2 && \text{By (194)} \\
&\le \frac{40\sqrt{\epsilon}}{\varphi} \cdot \frac{1}{\sqrt{|C_i||C_j|}} && \text{By Lemma 7 and } ||\widehat{\mu}_i - \mu_i||_2 \le \zeta ||\mu_i||_2
\end{aligned}$$

Also for any $i, j \in [k]$ we have

$$\begin{aligned}
&|\langle \widehat{\mu}_i, \widehat{\mu}_j \rangle - \langle \mu_i, \mu_j \rangle| \\
&= |\langle \mu_i + (\widehat{\mu}_i - \mu_i), \mu_j + (\widehat{\mu}_j - \mu_j) \rangle - \langle \mu_i, \mu_j \rangle| \\
&\le |\langle \widehat{\mu}_i - \mu_i, \widehat{\mu}_j - \mu_j \rangle| + |\langle \widehat{\mu}_i - \mu_i, \mu_j \rangle| + |\langle \widehat{\mu}_j - \mu_j, \mu_i \rangle| && \text{By triangle inequality} \\
&\le ||\widehat{\mu}_i - \mu_i||_2 ||\widehat{\mu}_j - \mu_j||_2 + ||\widehat{\mu}_i - \mu_i||_2 ||\mu_j||_2 + ||\widehat{\mu}_j - \mu_j||_2 ||\mu_i||_2 && \text{By Cauchy-Schwarz} \\
&\le (\zeta^2 + 2\zeta) (||\mu_i||_2 ||\mu_j||_2) && \text{Since } ||\widehat{\mu}_i - \mu_i||_2 \le \zeta ||\mu_i||_2 \text{ for all } i \\
&\le 6 \cdot \zeta \cdot \frac{1}{\sqrt{|C_i||C_j|}} && \text{By Lemma 7 } ||\mu_i||_2^2 \le \frac{2}{|C_i|} \text{ for all } i
\end{aligned} \tag{195}$$

Note that

$$\begin{aligned}
|\langle \widehat{\mu}_i, \widehat{\mu}_j \rangle| &\le |\langle \mu_i, \mu_j \rangle| + |\langle \mu_i, \mu_j \rangle - \langle \widehat{\mu}_i, \widehat{\mu}_j \rangle| && \text{By triangle inequality} \\
&\le \frac{8\sqrt{\epsilon}}{\varphi} \cdot \frac{1}{\sqrt{|C_i||C_j|}} + 6\zeta \cdot \frac{1}{\sqrt{|C_i||C_j|}} && \text{By Lemma 7 and (195)} \\
&\le \frac{10\sqrt{\epsilon}}{\varphi} \frac{1}{\sqrt{|C_i||C_j|}} && \text{Since } \zeta \le \frac{\sqrt{\epsilon}}{20 \cdot k \cdot \varphi}
\end{aligned}$$

Therefore we get

$$|\langle \widehat{\Pi}\widehat{\mu}_i, \widehat{\Pi}\widehat{\mu}_j \rangle| \le |\langle \widehat{\mu}_i, \widehat{\mu}_j \rangle| + |\langle (I - \widehat{\Pi})\widehat{\mu}_i, (I - \widehat{\Pi})\widehat{\mu}_j \rangle| \le \frac{50\sqrt{\epsilon}}{\varphi} \cdot \frac{1}{\sqrt{|C_i||C_j|}}.$$

□

6.5 Partitioning scheme works with *approximate* cluster means & dot products

In Section 6.3 we showed that the partitioning scheme works if we have access to real centers (i.e. $\mu_1, \ldots, \mu_k$), to exact dot product evaluations (i.e $\langle \cdot, \cdot \rangle$) and OuterConductance is precise.

In this section we show that approximations to all above is enough for the partitioning scheme to work. More precisely we show that if we have access only to $\langle \cdot, \cdot \rangle_{apx} \approx \langle \cdot, \cdot \rangle$, the search procedure finds $\hat{\mu}_i$'s that are only approximately equal to μ_i's and OuterConductance is only approximately correct then FindCenters still succeeds with high probability.

In order to prove such a statement we first show a technical Lemma (*Lemma 43*), that relates the approximate dot product with approximate centers to the dot product with the actual cluster centers.

Note that the following Lemma 43 works for any $S \subset \{\mu_1, \ldots, \mu_k\}$ and the corresponding $\widehat{S}$. This is useful for application in Lemma 45 because it allows to reason about candidate sets $\widehat{C}^(T_1, \ldots, T_b)_{\widehat{\mu}}$, after we associate $\bigcup_{i \in [b]} T_i$ with $\widehat{S}$.

Lemma 43. *Let $k \geq 2$, $\varphi \in (0,1)$, $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Then conditioned on the success of the spectral dot product oracle the following conditions hold.*

Let $\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_k$ be such that for all $i \in [k]$ $\|\widehat{\mu}_i - \mu_i\|^2 \leq 10^{-12} \cdot \frac{\epsilon}{\varphi^2 \cdot k^2} \|\mu_i\|^2$. Let $i \in [k]$ and $S \subseteq \{\mu_1, \ldots, \mu_k\} \setminus \{\mu_i\}$ and $\widehat{S} \subseteq \{\hat{\mu}_1, \ldots, \hat{\mu}_k\} \setminus \{\hat{\mu}_i\}$ be the corresponding subset to S. Let Π be the orthogonal projection onto span$(S)^\perp$ and $\widehat{\Pi}$ be the orthogonal projection onto span$(\widehat{S})^\perp$. Let also $\pi_i : \mathbb{R}^k \to \mathbb{R}^k$ be the projection onto the subspace spanned by $\Pi\mu_i$ and $\widehat{\Pi}\widehat{\mu}_i$. Then if $\|\Pi_i f_x\|^2 \leq \frac{10^4}{\min_{p \in [k]} |C_p|}$ then:

$$\left| \frac{\langle f_x, \Pi\mu_i \rangle}{\|\Pi\mu_i\|^2} - \frac{\left\langle f_x, \widehat{\Pi}\widehat{\mu}_i \right\rangle_{apx}}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|^2_{apx}} \right| \leq 0.02$$

Furthermore if $\widehat{\mu}_i$'s are averages of s points, then $\frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i \rangle_{apx}}{\|\widehat{\Pi}\widehat{\mu}_i\|^2_{apx}}$ can be computed in $\widetilde{O}_\varphi\left(s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ time with preprocessing time of $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ and space $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$

Proof. First we prove the runtime guarantee and then we show correctness.

Runtime. We first bound the running time. If we set the precision parameter of Algorithm 6 to $\xi = 10^{-6} \cdot \frac{\sqrt{\epsilon}}{\varphi}$ then by Theorem 2 the preprocessing time takes $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ time, $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ space, and by Corollary 1 computing $\frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i \rangle_{apx}}{\|\widehat{\Pi}\widehat{\mu}_i\|^2_{apx}}$ takes $\widetilde{O}_\varphi\left(s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ time.

Correctness. Now we show that we also obtain a good approximation. We will show it in two steps:

1. $\left| \frac{\langle f_x, \Pi\mu_i \rangle}{\|\Pi\mu_i\|^2} - \frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i \rangle}{||\widehat{\Pi}\widehat{\mu}_i||^2} \right| \leq 0.01$

2. $\left| \frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i \rangle}{\|\widehat{\Pi}\widehat{\mu}_i\|^2} - \frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i \rangle_{apx}}{\|\widehat{\Pi}\widehat{\mu}_i\|^2_{apx}} \right| \leq 0.01$

If we are able to prove 1 and 2 then the claim of the Lemma follows from triangle inequality. Before we present the two proofs we show a useful fact:

$$\begin{aligned}
||\widehat{\Pi}\widehat{\mu}_i - \Pi\mu_i|| &\leq ||\widehat{\Pi}\widehat{\mu}_i - \hat{\mu}_i|| + ||\Pi\mu_i - \mu_i|| + ||\widehat{\mu}_i - \mu_i|| && \text{By triangle inequality} \\
&\leq \frac{20\epsilon^{1/4}}{\sqrt{\varphi}}||\hat{\mu}_i|| + \frac{16\epsilon^{1/4}}{\sqrt{\varphi}}||\mu_i|| + 10^{-6} \cdot \frac{\sqrt{\epsilon}}{\varphi \cdot k}\|\mu_i\| && \text{By Lemma 41, 12 and the bound on } \|\widehat{\mu}_i - \mu_i\|^2 \\
&\leq \frac{40\epsilon^{1/4}}{\sqrt{\varphi}}||\mu_i|| && \text{As } \|\widehat{\mu}_i - \mu_i\|^2 \leq 10^{-12} \cdot \frac{\epsilon}{\varphi^2 \cdot k^2}\|\mu_i\|^2
\end{aligned} \tag{196}$$

Proof of 1: Notice that

$$\begin{aligned}
\left|\frac{\langle f_x, \Pi\mu_i\rangle}{||\Pi\mu_i||^2} - \frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i\rangle}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right| &= \left|\left\langle f_x, \frac{\Pi\mu_i}{||\Pi\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right\rangle\right| \\
&= \left|\left\langle \Pi_i f_x, \frac{\Pi\mu_i}{||\Pi\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right\rangle\right| && \text{By definition of } \pi_i \\
&\leq ||\Pi_i f_x|| \left|\left|\frac{\Pi\mu_i}{||\Pi\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right|\right| && \text{By Cauchy-Schwarz} \qquad (197)
\end{aligned}$$

First we will upper bound $\left|\left|\frac{\Pi\mu_i}{||\Pi\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right|\right|$. We split it into two cases:

Case 1. If $\frac{\Pi\mu_i}{||\Pi\mu_i||^2} \geq \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}$ then we have:

$$\begin{aligned}
\left|\left|\frac{\Pi\mu_i}{||\Pi\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right|\right| &\leq \left|\left|\frac{\Pi\mu_i}{(1-\frac{16\sqrt{\epsilon}}{\varphi})||\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right|\right| && \text{By Lemma 12} \\
&\leq \left|\left|\frac{\Pi\mu_i}{(1-\frac{16\sqrt{\epsilon}}{\varphi})||\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{(1+\frac{20\sqrt{\epsilon}}{\varphi})(1+10^{-12}\cdot\frac{\epsilon}{\varphi^2\cdot k^2})||\mu_i||^2}\right|\right| && \text{Lemma 41, assumptions} \\
&\leq \frac{2}{||\mu_i||^2}\left|\left|\Pi\mu_i - \left(1-\frac{1600\sqrt{\epsilon}}{\varphi}\right)\widehat{\Pi}\widehat{\mu}_i\right|\right| \\
&\leq \frac{2}{||\mu_i||^2}\left(\left|\left|\frac{1600\sqrt{\epsilon}}{\varphi}\Pi\mu_i\right|\right| + \left(1-\frac{1600\sqrt{\epsilon}}{\varphi}\right)||\widehat{\Pi}\widehat{\mu}_i - \Pi\mu_i||\right) && \text{By triangle inequality} \\
&\leq \frac{12800\sqrt{\epsilon}}{\varphi}\frac{1}{||\mu_i||} && \text{By (196) and Lemma 12}
\end{aligned}$$

Case 2. If $\frac{\Pi\mu_i}{||\Pi\mu_i||^2} < \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}$ then we have:

$$\begin{aligned}
\left|\left|\frac{\Pi\mu_i}{||\Pi\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right|\right| &\leq \left|\left|\frac{\Pi\mu_i}{(1+\frac{16\sqrt{\epsilon}}{\varphi})||\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right|\right| && \text{By Lemma 12} \\
&\leq \left|\left|\frac{\Pi\mu_i}{(1+\frac{16\sqrt{\epsilon}}{\varphi})||\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{(1-\frac{20\sqrt{\epsilon}}{\varphi})(1-10^{-12}\cdot\frac{\epsilon}{\varphi^2\cdot k^2})||\mu_i||^2}\right|\right| && \text{Lemma 41, assumptions} \\
&\leq \frac{2}{||\mu_i||^2}\left|\left|\Pi\mu_i - \left(1+\frac{1600\sqrt{\epsilon}}{\varphi}\right)\widehat{\Pi}\widehat{\mu}_i\right|\right| \\
&\leq \frac{2}{||\mu_i||^2}\left(\left|\left|\frac{1600\sqrt{\epsilon}}{\varphi}\Pi\mu_i\right|\right| + \left(1+\frac{1600\sqrt{\epsilon}}{\varphi}\right)||\widehat{\Pi}\widehat{\mu}_i - \Pi\mu_i||\right) && \text{By triangle inequality} \\
&\leq \frac{12800\sqrt{\epsilon}}{\varphi}\frac{1}{||\mu_i||} && \text{By (196) and Lemma 12}
\end{aligned}$$

Combining the two cases we get:

$$\left|\left|\frac{\Pi\mu_i}{||\Pi\mu_i||^2} - \frac{\widehat{\Pi}\widehat{\mu}_i}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right|\right| \leq \frac{12800\sqrt{\epsilon}}{\varphi}\frac{1}{||\mu_i||}.$$

Substituting into (197) we get:

$$\begin{aligned}
\left|\frac{\langle f_x, \Pi\mu_i\rangle}{||\Pi\mu_i||^2} - \frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i\rangle}{||\widehat{\Pi}\widehat{\mu}_i||^2}\right| &\leq ||\Pi_i f_x|| \cdot \frac{12800\sqrt{\epsilon}}{\varphi}\frac{1}{||\mu_i||} \\
&\leq \frac{100}{\sqrt{\min_{p\in[k]}|C_p|}} \cdot \frac{12800\sqrt{\epsilon}}{\varphi}\frac{1}{||\mu_i||} && \text{By assumption of the Lemma} \\
&\leq 0.005\frac{1}{\sqrt{\max_{p\in[k]}|C_p|}\cdot||\mu_i||} && \text{As } \frac{\epsilon}{\varphi^2} \text{ is sufficiently small and } \frac{\max_{p\in[k]}|C_p|}{\min_{p\in[k]}|C_p|} = O(1) \\
&\leq 0.01 && \text{By Lemma 7}
\end{aligned}$$

Proof of 2:

$$
\begin{aligned}
\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2 &\geq ||\widehat{\Pi}\widehat{\mu}_i||^2 - 10^{-6}\cdot\frac{\sqrt{\epsilon}}{\varphi}\cdot n^{-1} && \text{By Corollary 1, setting of } \xi \text{ and assumptions}\\
&\geq \left(1-\frac{20\sqrt{\epsilon}}{\varphi}\right)\cdot||\widehat{\mu}_i||^2 - 0.01\cdot n^{-1} && \text{By Lemma 41 and } \frac{\epsilon}{\varphi^2} \text{ small}\\
&\geq \left(1-10^{-12}\frac{\epsilon}{\varphi^2\cdot k}\right)\cdot 0.99\cdot||\mu_i||^2 - 0.01\cdot n^{-1} && \text{By } ||\widehat{\mu}_i-\mu_i||^2\leq 10^{-12}\frac{\epsilon}{\varphi^2\cdot k}||\mu_i||^2 \text{ and } \frac{\epsilon}{\varphi^2} \text{ small}\\
&\geq \left(1-\frac{4\sqrt{\epsilon}}{\varphi}\right)\cdot 0.98\cdot n^{-1} - 0.01\cdot n^{-1} && \text{By Lemma 7, } |C_i|\leq n,\ \frac{\epsilon}{\varphi^2} \text{ small}\\
&\geq 0.5\cdot n^{-1} && \text{As } \frac{\epsilon}{\varphi^2} \text{ small} \qquad (198)
\end{aligned}
$$

Next notice that:

$$
\begin{aligned}
\left|\frac{\langle f_x,\widehat{\Pi}\widehat{\mu}_i\rangle}{||\widehat{\Pi}\widehat{\mu}_i||^2} - \frac{\left\langle f_x,\widehat{\Pi}\widehat{\mu}_i\right\rangle_{apx}}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right| &\leq \left|\frac{\langle f_x,\widehat{\Pi}\widehat{\mu}_i\rangle}{||\widehat{\Pi}\widehat{\mu}_i||^2} - \frac{\left\langle f_x,\widehat{\Pi}\widehat{\mu}_i\right\rangle}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right| + \left|\frac{10^{-6}\cdot\frac{\sqrt{\epsilon}}{\varphi}\cdot n^{-1}}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right| && \text{By Corollary 1}\\
&\leq \left|\left\langle f_x,\widehat{\Pi}\widehat{\mu}_i\right\rangle\left(\frac{1}{||\widehat{\Pi}\widehat{\mu}_i||^2} - \frac{1}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right)\right| + \left|\frac{10^{-6}\cdot n^{-1}}{0.5\cdot n^{-1}}\right| && \text{By (198) and } \frac{\epsilon}{\varphi^2} \text{ small}\\
&\leq \left|\left\langle f_x,\widehat{\Pi}\widehat{\mu}_i\right\rangle\right|\left|\frac{1}{||\widehat{\Pi}\widehat{\mu}_i||^2} - \frac{1}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right| + 10^{-5} && (199)
\end{aligned}
$$

Now we will separately bound $\left|\left\langle f_x,\widehat{\Pi}\widehat{\mu}_i\right\rangle\right|$ and $\left|\frac{1}{||\widehat{\Pi}\widehat{\mu}_i||^2} - \frac{1}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right|$ from (199). As $|\langle a,b\rangle|\leq ||a||\cdot||b||$ we get:

$$
\left|\left\langle f_x,\widehat{\Pi}\widehat{\mu}_i\right\rangle\right| \leq ||\Pi_i f_x||\cdot||\widehat{\Pi}\widehat{\mu}_i|| \qquad (200)
$$

Now we bound the second term from (199):

$$
\begin{aligned}
\left|\frac{1}{||\widehat{\Pi}\widehat{\mu}_i||^2} - \frac{1}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right| &= \left|\frac{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2 - ||\widehat{\Pi}\widehat{\mu}_i||^2}{||\widehat{\Pi}\widehat{\mu}_i||^2\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right|\\
&\leq \left|\frac{10^{-6}\cdot\frac{\sqrt{\epsilon}}{\varphi}\cdot n^{-1}}{||\widehat{\Pi}\widehat{\mu}_i||^2\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right| && \text{Corollary 1, setting of } \xi \text{ and assumptions}\\
&\leq 10^{-5}\cdot\frac{\sqrt{\epsilon}}{\varphi}\cdot\left|\frac{0.5\cdot n^{-1}}{||\widehat{\Pi}\widehat{\mu}_i||^2\cdot 0.5\cdot n^{-1}}\right| && \text{By (198)}\\
&\leq 10^{-5}\cdot\frac{\sqrt{\epsilon}}{\varphi}\cdot\left|\frac{1}{||\widehat{\Pi}\widehat{\mu}_i||\cdot(||\Pi\mu_i|| - \frac{40\epsilon^{1/4}}{\sqrt{\varphi}}||\mu_i||)}\right| && \text{By (196)}\\
&\leq 10^{-4}\cdot\frac{\sqrt{\epsilon}}{\varphi}\cdot\frac{1}{||\widehat{\Pi}\widehat{\mu}_i||\cdot||\mu_i||} && \text{By Lemma 12 and } \frac{\epsilon}{\varphi^2} \text{ small}
\end{aligned}
$$
(201)

Substituting (200) and (201) in (199) we get:

$$
\begin{aligned}
\left| \frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i \rangle}{\|\widehat{\Pi}\widehat{\mu}_i\|^2} - \frac{\left\langle f_x, \widehat{\Pi}\widehat{\mu}_i \right\rangle_{apx}}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2} \right| &\le 10^{-5} + 10^{-4} \cdot \frac{\sqrt{\epsilon}}{\varphi} \cdot \frac{||\Pi_i f_x||}{||\mu_i||} \\
&\le 10^{-5} + 10^{-4} \cdot \frac{\sqrt{\epsilon}}{\varphi} \cdot \frac{100}{\sqrt{\min_{p\in[k]} |C_p|}} \cdot \frac{1}{||\mu_i||} && \text{By assumption} \\
&\le 10^{-5} + 10^{-3} \frac{1}{\sqrt{\max_{p\in[k]} |C_p|} \cdot ||\mu_i||} && \text{As } \frac{\epsilon}{\varphi^2} \text{ small, } \frac{\max_{p\in[k]} |C_p|}{\min_{p\in[k]} |C_p|} = O(1) \\
&\le 0.01 && \text{By Lemma 7}
\end{aligned}
$$

□

Now we are ready to show that there exist an algorithm (Algorithm 11) that can estimate accurately the size of candidate clusters of the form $\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}$ and then, if the size is not too small, estimate outer-conductance of all candidate clusters. The proof of correctness of the algorithm is based on applications of standard concentration bounds.

Algorithm 11 $\textsc{OuterConductance}(G, \widehat{\mu}, (T_1, T_2, \ldots, T_b), S, s_1, s_2)$
▷ T_i's are sets of $\widehat{\mu}_j$ where $\widehat{\mu}_j$'s are given as sets of points
▷ see Section 5.6 for the reason of such representation
▷ s_1 is # sampled points for size estimation
▷ s_2 is # sampled points for outer-conductance estimation

```
cnt := 0
for t = 1 to s_1 do
    x ~ UNIFORM{1..n}                    ▷ Sample a random vertex and test if it belongs to the cluster
    if ISINSIDE(x, μ̂, (T_1, T_2, ..., T_b), S) then
        cnt := cnt + 1
if (n/s_1) · cnt < min_{p∈[k]} |C_p|/2 then
    return ∞                             ▷ If the estimated size is too small return ∞
e := 0, a := 0
for t = 1 to s_2 do
    x ~ UNIFORM{1..n}
    y ~ UNIFORM{w ∈ N(u)}                ▷ N(u) = neighbors of u in G
    if ISINSIDE(x, μ̂, (T_1, T_2, ..., T_b), S) then
        a := a + 1
        if ¬ISINSIDE(y, μ̂, (T_1, T_2, ..., T_b), S) then
            e = e + 1
return e/a
```

Lemma 44. *Let $k \ge 2$, $\varphi, \epsilon, \gamma \in (0,1)$. Let $G = (V,E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$.*

For a set of approximate centers $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$, where each $\widehat{\mu}_i$ is represented as an average of at most s embedded vertices (i.e f_x's), an ordered partial partition $(T_1, \ldots, T_b)$ of $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$ and $\widehat{\mu} \in \{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\} \setminus \bigcup_{j\in[b]} T_i$ the following conditions hold.

If Algorithm 11 is invoked with $(G, \widehat{\mu}, (T_1, \ldots, T_b), \{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\} \setminus \bigcup_{j\in[b]} T_i, s_1, s_2)$ then it runs in $\widetilde{O}_\varphi\left((s_1+s_2) \cdot s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ time and if $s_1 = \Theta(k \log(\frac{1}{\eta}))$ and $s_2 = \Theta(\frac{\varphi^2 \cdot k}{\epsilon} \log(\frac{1}{\eta}))$ then with probability $1-\eta$ it returns a value q with the following properties.

- *If $|\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}| \ge \frac{3}{4} \min_{p\in[k]} |C_p|$ then $q \in \left[\frac{1}{2}\phi\left(\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}\right) - \epsilon/\varphi^2, \frac{3}{2}\phi\left(\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}\right) + \epsilon/\varphi^2\right]$,*
- *If $|\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}| < \frac{3}{4} \min_{p\in[k]} |C_p|$ then $q \ge \frac{1}{2}\phi\left(\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}\right) - \epsilon/\varphi^2$.*

Proof. We start with the runtime analysis then follows the correctness analysis.

Runtime. Algorithm 11 has two phases: one from line 1 to line 7 and second from line 8 to line 16.

During the first phase Algorithm 11 calls Algorithm 9 s_1 times and Algorithm 9 runs in $\widetilde{O}_\varphi(s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)})$ time as it computes $k^{O(1)}$ values of the form $\frac{\langle f_x, \hat{\mu}_i\rangle_{apx}}{||\hat{\mu}_i||^2_{apx}}$ which are computed in time $\widetilde{O}_\varphi(s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)})$ by Lemma 43, so in total the runtime of this phase is $\widetilde{O}_\varphi(s_1 \cdot s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)})$.

During the second phase Algorithm 11 calls Algorithm 9 $2s_2$ times so the runtime of this phase is $\widetilde{O}_\varphi(s_2 \cdot s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)})$ in total.

So in total the runtime is $\widetilde{O}_\varphi((s_1+s_2) \cdot s^4 \cdot \left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)})$.

Correctness. For simplicity we denote $\widehat{C}^{(T_1,\ldots,T_b)}_{\widehat{\mu}}$ by $\widehat{C}$ and $\min_{p\in[k]}|C_p|$ by $r_{\min}$ in this proof. Notice that the Algorithm 11 in the first phase computes $\text{cnt} = \sum_{i=1}^{s} X_i$, where X_i's are independent Bernoulli trials with success probability $p = \frac{|\widehat{C}|}{n}$. Let $z := \frac{n}{s_1}\sum_{i=1}^{s_1} X_i$. We introduce notation $x \approx_{\delta,\alpha} y$ to denote $x \in [(1-\delta)y - \alpha, (1+\delta)y+\alpha]$. By Chernoff-Hoeffding bounds we get that there exists a universal constant Γ such that for all $0 < \delta \le 1/2, \alpha > 0$

$$z \approx_{\delta,\alpha\cdot n} |\widehat{C}| \text{ with probability } 1 - 2^{-\Gamma s_1 \alpha\delta}.$$

Setting $\delta = 1/2, \alpha = \frac{r_{\min}}{8n}$ we get that $z \approx_{1/2, r_{\min}/8} |\widehat{C}|$ with probability

$$1 - 2^{-\Gamma s_1 \frac{r_{\min}}{32n}} \ge 1 - 2^{-\Omega(s_1/k)},$$

as $\frac{\max_{p\in[k]}|C_p|}{\min_{p\in[k]}|C_p|} = O(1)$. So if $s_1 = \Theta(k\log(1/\eta))$ then with probability $1-\eta/2$ we have

$$z \approx_{1/2, r_{\min}/8} |\widehat{C}|. \tag{202}$$

Observe that if $\widehat{C} < r_{\min}/4$ then by (202) we have that $z \le (1+1/2)|\widehat{C}| + r_{\min}/8 < r_{\min}/2$, which means that Algorithm 11 returns ∞. Note that it is consistent with the conclusion of the Lemma.

For the analysis of the second stage we assume that $|\widehat{C}| \ge r_{\min}/4$. We will analyze what value is returned in the second stage. First we will bound the probability that $a \le \frac{s_2 \cdot r_{\min}}{8\cdot n}$. For $i \in [1\ldots s_2]$ let X_i be a binary random variable which is equal 1 iff in $i-th$ iteration of the for loop we increase the a counter. We have that, for every i, $P[X_i = 1] = |\widehat{C}|/n$ and the X_i's are independent. Notice that $a = \sum_{i=1}^{s_2} X_i$. From Chernoff bound we have that for $\delta < 1$:

$$P\left[\left|\sum_{i=1}^{s_2} X_i - \mathbb{E}\left[\sum_{i=1}^{s_2} X_i\right]\right| > \delta \cdot \mathbb{E}\left[\sum_{i=1}^{s_2} X_i\right]\right] \le 2e^{-\frac{\delta^2}{3}\mathbb{E}\left[\sum_{i=1}^{s_2} X_i\right]}, \tag{203}$$

Noticing that $\mathbb{E}\left[\sum_{i=1}^{s_2} X_i\right] = s_2\frac{|\widehat{C}|}{n}$ if we set $\delta = 1/2$ we get that

$$P\left[\left|\sum_{i=1}^{s_2} X_i - s_2\frac{|\widehat{C}|}{n}\right| > s_2\frac{|\widehat{C}|}{2n}\right] \le 2e^{-s_2\frac{|\widehat{C}|}{12n}} \le 2e^{-\frac{s_2\cdot r_{\min}}{48\cdot n}}, \tag{204}$$

So with probability at least $1 - 2e^{-\frac{s_2\cdot r_{\min}}{48\cdot n}} \ge 1 - 2e^{-\Omega(s_2/k)}$ (as $\frac{\max_{p\in[k]}|C_p|}{\min_{p\in[k]}|C_p|} = O(1)$) we have that

$$a = \sum_{i=1}^{s} X_i \ge \frac{1}{2}\cdot s_2 \cdot \frac{|\widehat{C}|}{n} \ge \frac{s_2\cdot r_{\min}}{8\cdot n} \ge \Omega(s_2/k). \tag{205}$$

Now observe that line 14 of OuterConductance is invoked exactly a times. Let Y_j be the indicator random variable that is 1 iff e is increased in the j-th call of line 14. Notice that

$$P[Y_i = 1] = \phi(\widehat{C}) \tag{206}$$

That is because if U_i is a random variable denoting a vertex u sampled in i-th step then U_i is uniform on set $\widehat{C}$ conditioned on $X_i = 1$ and the graph is regular. Now by the Chernoff-Hoeffging bounds we get that for all $0 < \delta \le 1/2, \alpha > 0$ we have:

$$\frac{1}{a}\sum_{i=1}^{a} Y_i \approx_{\delta,\alpha} \phi(\widehat{C}) \text{ with probability } 1 - 2e^{-\Gamma a\alpha\delta}.$$

Setting $\delta = 1/2, \alpha = \frac{\epsilon}{\varphi^2}$ we get that $\frac{1}{a}\sum_{i=1}^{a} Y_i \approx_{1/2,\epsilon/\varphi^2} \phi(\widehat{C})$ with probability:

$$1 - 2e^{-\Gamma a\epsilon/(4\varphi^2)} \geq 1 - 2e^{-\Omega(a\epsilon/\varphi^2)} \tag{207}$$

Now taking the union bound over (205) and (207) we get that if we set $s_2 = \Theta(\frac{\varphi^2 \cdot k}{\epsilon}\log(1/\eta))$ then $\frac{1}{a}\sum_{i=1}^{a} Y_i \approx_{1/2,\epsilon/\varphi^2} \phi(\widehat{C})$ with probability:

$$\begin{aligned} 1 - 2e^{-\Omega(s_2/k)} - 2e^{-\Omega(a\epsilon/\varphi^2)} &\geq 1 - 2e^{-\Omega(s_2/k)} - 2e^{-\Omega(\frac{\epsilon \cdot s_2}{\varphi^2 \cdot k})} && \text{By (205)} \\ &\geq 1 - \eta/2 \end{aligned}$$

To conclude the proof we observe the following.

- If $|\widehat{C}| < \frac{r_{\min}}{4}$ then with probability $1 - \eta/2$ the Algorithm returns ∞,
- If $|\widehat{C}| \in [\frac{r_{\min}}{4}, \frac{3 \cdot r_{\min}}{4})$ then either the Algorithm returns ∞ in the first stage or it reaches the second stage and with probability $1 - \eta$ it returns a value ψ such that $\psi \approx_{1/2,\epsilon/\phi^2} \varphi(\widehat{C})$,
- If $|\widehat{C}| \geq \frac{r_{\min}}{4}$ then by the union bound over the two stages with probability $1 - \eta$ it reaches the second stage and returns a value ψ such that $\psi \approx_{1/2,\epsilon/\varphi^2} \phi(\widehat{C})$.

The above covers all the cases and is consistent with the conclusions of the Lemma.

□

Before we give the statement of the next Lemma we introduce some definitions. In Lemma 44 we proved that for every call to OUTERCONDUCTANCE the value returned by the Algorithm 11 is, in a sense given by the conclusions of Lemma 44, a good approximation to outer-conductance of $\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}$ (where $\widehat{\mu}, (T_1, \ldots, T_b)$ are the parameters of the call) with high probability. What follows is a definition of an event that the values returned by OUTERCONDUCTANCE throughout the run of the final algorithm always satisfy one conclusion of Lemma 44. Later we use Definition 13 in Lemma 45 and then in the proof of Theorem 8 we will lower bound the probability of $\mathcal{E}_{\text{conductance}}$.

Definition 13 (**Event $\mathcal{E}_{\text{conductance}}$**). Let $k \geq 2$, $\varphi, \epsilon, \gamma \in (0,1)$. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$.

We define $\mathcal{E}_{\text{conductance}}$ as an event:

For every call to Algorithm 11 (i.e. OUTERCONDUCTANCE) that is made throughout the run of FINDCENTERS the following is true. If Algorithm 11 is invoked with $(G, \widehat{\mu}, (T_1, \ldots, T_b), \{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\} \setminus \bigcup_{j \in [b]} T_i, s_1, s_2)$ then it returns a value q with the following property.

- If $|\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}| \geq \frac{3}{4}\min_{p \in [k]} |C_p|$ then $q \in \left[\frac{1}{2}\phi\left(\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}\right) - \epsilon/\varphi^2, \frac{3}{2}\phi\left(\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_b)}\right) + \epsilon/\varphi^2\right]$.

The following Lemma is the key part of the corresponding proof of correctness of Algorithm 8 (see Theorem 8 below). It is a generalization of Lemma 37. We show that if $\widehat{\mu}$'s are close to real centers and $\mathcal{E}$ and $\mathcal{E}_{\text{conductance}}$ hold then at every stage of the for loop from line 4 of Algorithm 8 at least half of the candidate clusters:

$$\mathcal{C}_i := \bigcup_{\widehat{\mu} \in S} \{\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_{i-1})}\},$$

pass the test from line 6 of Algorithm 8, which means that they have small outer-conductance and satisfy condition (146).

Lemma 45. *Let $k \geq 2$, $\varphi \in (0,1)$, $\frac{\epsilon}{\varphi^2}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Then conditioned on the success of the spectral dot product oracle there exists an absolute constant Υ such that the following conditions hold.*

If COMPUTEORDEREDPARTITION*$(G, \widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_k, s_1, s_2)$ is called with $(\widehat{\mu}_1, \ldots, \widehat{\mu}_k)$ such that for every $i \in [k]$ we have $\|\widehat{\mu}_i - \mu_i\|^2 \leq 10^{-12} \cdot \frac{\epsilon}{\varphi^2 \cdot k^2}\|\mu_i\|^2$ then the following holds. Assume that at the beginning of the i-th iteration of the for loop from line 4 of Algorithm 8 $|S| = b$ and, up to renaming of $\widehat{\mu}$'s, $S = \{\widehat{\mu}_1, \ldots, \widehat{\mu}_b\}$, the corresponding clusters are $\mathcal{C} = \{C_1, \ldots, C_b\}$ respectively and the ordered partial partition of μ's is equal to $(T_1, \ldots, T_{i-1})$. Then if for every $C \in \mathcal{C}$ we have that $|V^{(T_1,\ldots,T_{i-1})} \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$ then at the beginning of $(i+1)$-th iteration:*

1. $|S| \leq b/2$ *(that is at least half of the remaining cluster means were removed in i-th iteration),*

2. *for every* $\mu \in S$ *the corresponding cluster* C *satisfies* $|V^{(T_1,\ldots,T_i)} \cap C| \geq \left(1 - \Upsilon \cdot (i+1) \cdot \frac{\epsilon}{\varphi^2}\right)|C|$, *where* $(T_1, \ldots, T_i)$ *is the ordered partial partition of* μ*'s created in the first* i *iterations.*

Proof. **Outline of the proof.** We start but defining a subset of vertices called outliers and then we show that the number of them is small. Next we prove that for vertices that are not outliers the evaluations of $\frac{\langle f_x, \widehat{\Pi}\widehat{\mu}_i\rangle_{apx}}{\|\widehat{\Pi}\widehat{\mu}_i\|_{apx}^2}$ are approximately correct (as in Lemma 43). Next we mimic the structure, and on the high level the logic, of the proof of Lemma 37: we first show the first conclusion of the Lemma and then the second one.

For simplicity we will denote $\min_{p\in[k]} |C_p|$ by $r_{\min}$ in this proof. Without loss of generality we can assume $S = \{\hat{\mu}_1, \ldots, \hat{\mu}_b\}$ at the beginning of the i-th iteration of the for loop from line 4 of Algorithm 8 and the corresponding clusters be $C_1, \ldots, C_b$ respectively. Assume that for every $C \in \mathcal{C}$ we have that $|V^{(T_1,\ldots,T_{i-1})} \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$.

Let $\widehat{\Pi}$ be the projection onto the span$(\bigcup_{j<i} T_j)^\perp$. Recall that each T_j is a subset of $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$. For every $j < i$ let

$$T'_j := \bigcup_{\widehat{\mu}\in T_j} \{\mu\}.$$

That is T'_j's are T_j's with $\widehat{\mu}$'s replaced by the corresponding μ's. Now let Π be the projection onto the span$(\bigcup_{j<i} T'_j)^\perp$.

Outliers. First we define a set of outliers, i.e. X, as the set of points with abnormally long projection onto the subspace spanned by $\{\Pi\mu_1, \ldots, \Pi\mu_b, \widehat{\Pi}\widehat{\mu}_1, \ldots, \widehat{\Pi}\widehat{\mu}_b\}$. Then we show that the number of outliers is small.

Let Q be the orthogonal projection onto the span$(\{\Pi\mu_1, \ldots, \Pi\mu_b, \widehat{\Pi}\hat{\mu}_1, \ldots, \widehat{\Pi}\hat{\mu}_b\})$ and let:

$$X := \left\{x \in V : ||Qf_x||^2 > \frac{10^4}{r_{\min}}\right\}$$

By Lemma 33 we get that

$$\sum_{x\in V} ||Qf_x - Q\mu_x||^2 \leq O\left(b \cdot \frac{\epsilon}{\varphi^2}\right). \tag{208}$$

Moreover for every $x \in X$:

$$\begin{aligned}
||Qf_x - Q\mu_x|| &\geq ||Qf_x|| - ||Q\mu_x|| && \text{By triangle inequality}\\
&\geq ||Qf_x|| - ||\mu_x|| && \text{As projection can only decrease the norm}\\
&> \frac{10^2}{\sqrt{r_{\min}}} - \left(1 + O\left(\frac{\sqrt{\epsilon}}{\varphi}\right)\right)\frac{1}{\sqrt{r_{\min}}} && \text{By Lemma 7 and Definition of } X\\
&\geq \frac{90}{\sqrt{r_{\min}}} && \text{For } \frac{\epsilon}{\varphi^2} \text{ small enough}
\end{aligned} \tag{209}$$

Combining (208) and (209) we get:

$$|X| \leq O\left(b \cdot \frac{\epsilon}{\varphi^2}\right) \cdot r_{\min} \leq O\left(b \cdot \frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k} \tag{210}$$

Tests performed for non-outliers are approximately correct. Observe that by the fact that spectral dot product succeeds we have by Lemma 43 that for all $x \in V \setminus X$ and for all $i \in [k]$:

$$\left|\frac{\langle f_x, \Pi\mu_i\rangle}{\|\Pi\mu_i\|^2} - \frac{\left\langle f_x, \widehat{\Pi}\widehat{\mu}_i\right\rangle_{apx}}{\left\|\widehat{\Pi}\widehat{\mu}_i\right\|_{apx}^2}\right| \leq 0.02, \tag{211}$$

as $||Qf_x||^2 \leq \frac{10^4}{r_{\min}}$ and the norm in any subspace can only be smaller and thus the assumption of Lemma 43 is satisfied.

1. At least half of the cluster means is removed from S. Now we proceed with proving that most of the candidate clusters $\widehat{C}_{\widehat{\mu}}^{(T_1,\ldots,T_{i-1})}$ have small outer-conductance and thus the corresponding $\widehat{\mu}$'s are removed from set S (see line 6 of ComputeOrderedPartition). For brevity we will refer to $(T_1,\ldots,T_{i-1})$ as P in this proof.

Let $\mu \in S$. Let

$$I := \bigcup_{\mu',\mu''\in\{\mu_1,\ldots,\mu_d\}} C_{\Pi\mu',0.9} \cap C_{\Pi\mu'',0.9}.$$

By Lemma 36 we have that

$$|I| \leq O\left(b \cdot \frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k}. \tag{212}$$

So by (210) and (212) and Markov inequality we get that there exists a subset of clusters $\mathcal{R} \subseteq \mathcal{C}$ such that $|\mathcal{R}| \geq b/2$ and for every $C \in \mathcal{R}$ we have that:

$$|C \cap (I \cup X)| \leq O\left(\frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k} \tag{213}$$

We will argue that for any order of the for loop from line 4 of Algorithm 8 it is true that for every $C \in \mathcal{R}$ with corresponding means $\mu, \widehat{\mu}$ the candidate cluster $\widehat{C}_{\widehat{\mu}}^{P}$ satisfies the if statement from line 6 of Algorithm 8. Recall that as per Definition 12:

$$\widehat{C}_{\widehat{\mu}}^{P} = \left\{ x \in V : \textsc{IsInside}\left(x, \widehat{\mu}, P, \{\widehat{\mu}_1,\ldots,\widehat{\mu}_k\} \setminus \bigcup_{j\in[i-1]} T_j\right) = \textsc{True} \right\}.$$

First note that behavior of the algorithm is independent of the order of the for loop from line 4 of Algorithm 8 as by definition $\widehat{C}_{\widehat{\mu}}^{P}$'s for $\widehat{\mu} \in S$ are pairwise disjoint. Now let $C \in \mathcal{R}$, $\mu, \widehat{\mu}$ be the means corresponding to C and $\widehat{C}_{\widehat{\mu}}^{P}$ be the candidate cluster corresponding to $\widehat{\mu}$ with respect to $P = (T_1,\ldots,T_{i-1})$.

Now the goal is to show:

$$|\widehat{C}_{\widehat{\mu}}^{P} \triangle C| \leq O\left(\frac{\epsilon}{\varphi^2}\cdot \log(k)\right) \cdot |C|,$$

from which we will later conclude that the outer-conductance of the candidate set $\widehat{C}_{\widehat{\mu}}^{P}$ is small. Intuitively we would like to argue that

$$C_{\Pi\mu,0.96} \,\widetilde{\subseteq}\, C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} \,\widetilde{\subseteq}\, C_{\Pi\mu,0.9}, \tag{214}$$

and then use Lemmas from Section 6.2. The equation (214) *is true up to the outliers as Lemma 43 guarantees a bound of* 0.02 *for the test computations for vertices of small norm.*

Now we give a formal proof, which is split into 2 parts:

Showing $|\widehat{C}_{\widehat{\mu}}^{P} \cap C| \geq \left(1 - O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right)\right)|C|$. First we note that by (211) $C_{\Pi\mu,0.96}$ is mostly contained in $C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93}$. Recall that (see Definition 9 and Definition 8) we have:

$$C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} = \left\{ x \in V : \left\langle f_x, \widehat{\Pi}\widehat{\mu}\right\rangle_{apx} \geq 0.93 \left\|\widehat{\Pi}\widehat{\mu}\right\|^2_{apx} \right\},$$

$$C_{\Pi\mu,0.96} = \left\{ x \in V : \langle f_x, \Pi\mu\rangle \geq 0.96 ||\Pi\mu||^2 \right\}.$$

And (211) gives us that the errors for non-outliers are bounded by 0.02, so formally we get:

$$C_{\Pi\mu,0.96} \setminus C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} \subseteq X \tag{215}$$

Similarly, also by (211) we get that the intersections of candidate clusters $C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93}$ lie mostly in I. Formally:

$$C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} \cap \bigcup_{\hat{\mu}'\neq\hat{\mu}} C^{apx}_{\widehat{\Pi}\hat{\mu}',0.93} \subseteq I \cup X \tag{216}$$

By Lemma 31 we get that

$$|C \cap C_{\Pi\mu,0.96}| \geq \left(1 - O\left(\frac{\epsilon}{\varphi^2}\right)\right)|C| \tag{217}$$

Note that having two thresholds (0.9 and 0.96) is very important here (see Remark 7). Intuitively we need some slack to show $C_{\Pi\mu,0.96} \widetilde{\subseteq} C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} \widetilde{\subseteq} C_{\Pi\mu,0.9}$ as there is always some error in computation of $\frac{\left\langle f_x, \widehat{\Pi}\widehat{\mu}_i \right\rangle_{apx}}{\|\widehat{\Pi}\widehat{\mu}_i\|^2_{apx}}$.

Now combining inductive assumption $|V^P \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$, (213), (215), (216) and (217) we get that:

$$|\widehat{C}^P_{\widehat{\mu}} \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C| - O\left(\frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k} - O\left(\frac{\epsilon}{\varphi^2}\right) \cdot |C|$$
$$\geq \left(1 - O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)\right)|C| \tag{218}$$

Showing $|\widehat{C}^P_{\widehat{\mu}} \cap (V^P \setminus C)| \leq O\left(\frac{\epsilon}{\varphi^2}\right)|C|$. Recall that as per Definition 12 we have:

$$V^P = V \setminus \bigcup_{j<i} \bigcup_{\widehat{\mu} \in T_j} \widehat{C}^{(T_1,\ldots,T_{j-1})}_{\widehat{\mu}}$$

By Lemma 32 we get that:

$$|C_{\Pi\mu,0.9} \cap (V^P \setminus C)| \leq |C_{\Pi\mu,0.9} \cap (V \setminus C)| \leq O\left(\frac{\epsilon}{\varphi^2}\right)|C| \tag{219}$$

By (211) we get:

$$C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} \setminus C_{\Pi\mu,0.9} \subseteq X \tag{220}$$

Let π' be the projection onto the span of $\{\Pi\mu, \widehat{\Pi}\hat{\mu}\}$. Moreover let:

$$X' := \left\{x \in V : \|\pi' f_x\|^2 > \frac{10^4}{r_{\min}}\right\}.$$

Note that by Lemma 33 we have:

$$\sum_{x \in V} ||\pi' f_x - \pi' \mu_x||^2 \leq O\left(\frac{\epsilon}{\varphi^2}\right) \tag{221}$$

Moreover for every $x \in X'$ we have:

$$\begin{aligned}
||\pi' f_x - \pi' \mu_x|| &\geq ||\pi' f_x|| - ||\pi' \mu_x|| && \text{By } \triangle \text{ inequality} \\
&\geq \frac{10^2}{\sqrt{r_{\min}}} - \frac{2}{\sqrt{r_{\min}}} && \text{By Lemma 7} \\
&\geq \frac{90}{\sqrt{r_{\min}}}
\end{aligned} \tag{222}$$

Combining (221) and (222) we get that:

$$|X'| \leq O\left(\frac{\epsilon}{\varphi^2}\right) \cdot r_{\min} \leq O\left(\frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k} \tag{223}$$

Then similarly to the analysis of (211) by Lemma 43 and the fact that spectral dot product succeeds we have that for every $x \in V \setminus X'$:

$$\left| \frac{\langle f_x, \Pi\mu \rangle}{\|\Pi\mu\|^2} - \frac{\left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx}}{\left\|\widehat{\Pi}\widehat{\mu}\right\|^2_{apx}} \right| \leq 0.02$$

Thus we get:

$$C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} \setminus C_{\Pi\mu,0.9} \subseteq X', \tag{224}$$

as for points not belonging to X' the error in the tests performed by the Algorithm is upper bounded by 0.02. Combining (219) and (224) we have:

$$|\widehat{C}^P_{\widehat{\mu}} \cap (V^P \setminus C)| \leq O\left(\frac{\epsilon}{\varphi^2}\right)|C| \tag{225}$$

And finally putting (218) and (225) together we have:

$$|\widehat{C}^P_{\widehat{\mu}} \triangle C| \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right) \cdot |C| \tag{226}$$

Outer-conductance of $\widehat{C}^P_{\widehat{\mu}}$ is small. Now we want to argue that $\widehat{C}^P_{\widehat{\mu}}$ passes the outer-conductance test from line 6 in Algorithm 8. From the definition of outer-conductance:

$$\begin{aligned}
\phi(\widehat{C}^P_{\widehat{\mu}}) &\leq \frac{E(C, V \setminus C) + d|\widehat{C}^P_{\widehat{\mu}} \triangle C|}{d(|C| - |\widehat{C}^P_{\widehat{\mu}} \triangle C|)} \\
&\leq \frac{E(C, V \setminus C) + d \cdot O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)|C|}{d(|C| - O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)|C|)} && \text{from (226)} \\
&\leq \frac{O\left(\frac{\epsilon}{\varphi^2}\right) + O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)}{1 - O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right)} && \text{because } \frac{E(C, V \setminus C)}{d|C|} \leq O\left(\frac{\epsilon}{\varphi^2}\right) \\
&\leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right) && \text{for sufficiently small } \frac{\epsilon}{\varphi^2} \cdot \log(k)
\end{aligned}$$

and it follows that

$$\phi(\widehat{C}^P_{\widehat{\mu}}) \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right),$$

To conclude we notice that by (226) we have $|\widehat{C}^P_{\widehat{\mu}}| > \frac{3 \cdot r_{\min}}{4}$, so as $\mathcal{E}_{\text{conductance}}$ is true we get that the candidate cluster $\widehat{C}^P_{\widehat{\mu}}$ passes the test.

2. Clusters corresponding to unremoved $\widehat{\mu}$'s satisfy condition 2. Now we prove that for every $\widehat{\mu}$ that was not removed from set S only small fraction of its corresponding cluster is removed.

Let $\widehat{\mu} \in S$ be such that it is not removed in the i-th step and let μ be the corresponding real center. Let $C \in \mathcal{C}$ be the cluster corresponding to μ. By assumption $|V^P \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$, where recall that $P = (T_1, \ldots, T_{i-1})$.

Now the goal is to show:

$$|C \cap (V^{(T_1,\ldots,T_{i-1})} \setminus V^{(T_1,\ldots,T_i)})| \leq O\left(\frac{\epsilon}{\varphi^2}\right)|C| + O\left(\frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k} \leq O\left(\frac{\epsilon}{\varphi^2}\right)|C|,$$

that is, that there is only a small number of vertices that were removed in the i-th stage and belong to C at the same time. Intuitively we want to show that:

$$(V^{(T_1,\ldots,T_{i-1})} \setminus V^{(T_1,\ldots,T_i)}) \cap C_{\Pi\mu,0.96} \approx \emptyset,$$

and then use Lemmas from Section 6.2. The equation above is true up to the outliers as Lemma 43 guarantees a bound of 0.02 *for the test computations for vertices of small norm.*

Now we give a formal proof. Let $x \in V^{(T_1,\ldots,T_{i-1})} \setminus V^{(T_1,\ldots,T_i)} = V^P \setminus V^{(T_1,\ldots,T_i)}$, where $(T_1, \ldots, T_i)$ is the partial partition of $\widehat{\mu}$'s created in the first i steps of the for loop of ComputeOrderedPartition. Then there exists $\widehat{\mu}' \in \{\widehat{\mu}_1, \ldots, \widehat{\mu}_b\}$ such that $x \in \widehat{C}^P_{\widehat{\mu}'}$ (recall that $\widehat{C}^P_{\widehat{\mu}'}$ is the candidate cluster corresponding to $\widehat{\mu}'$ with respect to $P = (T_1, \ldots, T_{i-1})$). Recall (Definition 12) that $\widehat{C}^P_{\widehat{\mu}'}$ is defined as:

$$\widehat{C}^P_{\widehat{\mu}'} = \left\{x \in V : \textsc{IsInside}\left(x, \widehat{\mu}', P, \{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\} \setminus \bigcup_{j \in [i-1]} T_j\right) = \textsc{True}\right\}.$$

This in particular means (see line 8 of Algorithm IsInside) that:

$$\widehat{C}^P_{\widehat{\mu}'} \subseteq C^{apx}_{\widehat{\Pi}\widehat{\mu}',0.93} \setminus \bigcup_{\widehat{\mu}'' \in S\setminus\{\widehat{\mu}'\}} C^{apx}_{\widehat{\Pi}\widehat{\mu}'',0.93},$$

which, as $\widehat{\mu} \in S \setminus \{\widehat{\mu}'\}$, gives us that:

$$\widehat{C}^P_{\widehat{\mu}'} \cap C^{apx}_{\widehat{\Pi}\widehat{\mu},0.93} = \emptyset,$$

which using Definition 8 gives that:

$$\left\langle f_x, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx} < 0.93 \left\| \widehat{\Pi}\widehat{\mu} \right\|^2_{apx}. \tag{227}$$

We define X' similarly as in point 1. Let π' be the projection onto the span of $\{\Pi\mu, \widehat{\Pi}\widehat{\mu}\}$. Moreover let:

$$X' := \left\{ x \in V : \|\pi' f_x\|^2 > \frac{10^4}{r_{\min}} \right\}.$$

Similarly to the proof of (223) we get

$$|X'| \le O\left(\frac{\epsilon}{\varphi^2}\right) \cdot r_{\min} \le O\left(\frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k} \tag{228}$$

Again similarly to the analysis of (211) we note that by Lemma 43 and the fact that spectral dot product succeeds:

$$\text{for every } y \in V \setminus X' \text{ we have } \left| \frac{\langle f_y, \Pi\mu \rangle}{\|\Pi\mu\|^2} - \frac{\left\langle f_y, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx}}{\left\| \widehat{\Pi}\widehat{\mu} \right\|^2_{apx}} \right| \le 0.02 \tag{229}$$

Combining (229) and (227) we get that if $x \in V \setminus X'$ then

$$\begin{aligned} \frac{\langle f_x, \Pi\mu \rangle}{\|\Pi\mu\|^2} &\le \frac{\left\langle f_y, \widehat{\Pi}\widehat{\mu} \right\rangle_{apx}}{\left\| \widehat{\Pi}\widehat{\mu} \right\|^2_{apx}} + 0.02 \\ &< 0.93 + 0.02 \\ &< 0.96 \end{aligned}$$

which also means that $x \notin C_{\Pi\mu,0.96}$. This means that:

$$(V^{(T_1,\ldots,T_{i-1})} \setminus V^{(T_1,\ldots,T_i)}) \cap C_{\Pi\mu,0.96} \subseteq X' \tag{230}$$

But by Lemma 31:

$$|\{x \in C : \langle \Pi f_x, \Pi\mu \rangle < 0.96 \|\Pi\mu\|_2^2\}| \le O\left(\frac{\epsilon}{\varphi^2}\right) \cdot |C| \tag{231}$$

Combining (230), (228) and (231) we get that:

$$|C \cap (V^{(T_1,\ldots,T_{i-1})} \setminus V^{(T_1,\ldots,T_i)})| \le O\left(\frac{\epsilon}{\varphi^2}\right)|C| + O\left(\frac{\epsilon}{\varphi^2}\right) \cdot \frac{n}{k} \le O\left(\frac{\epsilon}{\varphi^2}\right)|C|. \tag{232}$$

By assumption that $|V^{(T_1,\ldots,T_{i-1})} \cap C| \ge \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right)|C|$ and (232) we get that:

$$|V^{(T_1,\ldots,T_i)} \cap C| \ge \left(1 - \Upsilon \cdot (i+1) \cdot \frac{\epsilon}{\varphi^2}\right)|C|,$$

provided that Υ is bigger than the constant hidden under O notation in (232).

□

The following Lemma is a generalization of Theorem 7 that uses Lemma 45 as an inductive step to show that if ComputeOrderedPartition is called with $\widehat{\mu}$'s that are good approximations to μ's then it returns an ordered partition that induces a good collection of clusters.

Lemma 46. *Let $k \geq 2$, $\varphi \in (0,1)$ and $\frac{\epsilon}{\varphi^2} \cdot \log(k)$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Then conditioned on the success of the spectral dot product oracle the following conditions hold.*

If ComputeOrderedPartition*$(G, \widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_k, s_1, s_2)$ is called with $(\hat{\mu}_1, \ldots, \hat{\mu}_k)$ such that for every $i \in [k]$ we have $\|\widehat{\mu}_i - \mu_i\|^2 \leq 10^{-12} \cdot \frac{\epsilon}{\varphi^2 \cdot k^2} \|\mu_i\|^2$ then* ComputeOrderedPartition *returns* $(\text{True}, (T_1, \ldots, T_b))$ *such that $(T_1, \ldots, T_b)$ induces a collection of clusters $\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ such that there exists a permutation π on k elements such that for all $i \in [k]$:*

$$\left|\widehat{C}_{\widehat{\mu}_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|$$

and

$$\phi(\widehat{C}_{\widehat{\mu}_i}) \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right).$$

Proof. Note that for $i = 0$ in the for loop in line 2 of ComputeOrderedPartition S and clusters $\{C_1, \ldots, C_k\}$ trivially satisfy assumptions of Lemma 45. So using Lemma 45 and induction we get that for every $i \in [0..\lceil\log(k)\rceil]$ at the beginning of the i-th iteration:

- $|S| \leq k/2^i$,
- for every $\widehat{\mu} \in S$ with corresponding μ and corresponding cluster C we have $|V^{(T_1, \ldots, T_{i-1})} \cap C| \geq \left(1 - \Upsilon \cdot i \cdot \frac{\epsilon}{\varphi^2}\right) |C|$ (where Υ is the constant from the statement of Lemma 45).

In particular this means that after $O(\log(k))$ iterations set S becomes empty. This also means that ComputeOrderedPartition returns in line 10, so it returns True and the ordered partial partition $(T_1, \ldots, T_b)$ is in fact an ordered partition of $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$.

Note that by definition (see Definition 10) all the approximate clusters $\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ are pairwise disjoint and moreover for every constructed cluster $\widehat{C} \in \{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ we have:

$$\phi(\widehat{C}) \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right),$$

as it passed the test in line 6 of ComputeOrderedPartition. So by Lemma 16 it means that there exists a permutation π on k elements such that for all $i \in [k]$:

$$\left|\widehat{C}_{\widehat{\mu}_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|.$$

Recall Remark 9 for why the proof follows this framework of first arguing about outer-conductance and only after that, using Lemma 16, reasoning about symmetric difference. □

Now we present the final Theorem of this section which shows that FindCenters with high probability returns an ordered partition that induces a good collection of clusters. The proof is a careful union bound of error probabilities.

Theorem 8. *Let $k \geq 2$, $\varphi \in (0,1)$, $\frac{\epsilon \log(k)}{\varphi^3}$ be smaller than a sufficiently small constant. Let $G = (V, E)$ be a d-regular graph that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$. Then Algorithm 10 with probability $1 - \eta$ returns an ordered partition $(T_1, \ldots, T_b)$ such that $(T_1, \ldots, T_b)$ induces a collection of clusters $\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ such that there exists a permutation π on k elements such that for all $i \in [k]$:*

$$\left|\widehat{C}_{\widehat{\mu}_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|$$

and

$$\phi(\widehat{C}_{\widehat{\mu}_i}) \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right).$$

Moreover

- *Algorithm 10 (*FINDCENTERS*) runs in time*

$$\widetilde{O}_\varphi\left(\log^2(1/\eta)\cdot 2^{O(\frac{\varphi^2}{\epsilon}\cdot k^4\log^2(k))}\cdot n^{1/2+O(\epsilon/\varphi^2)}\right),$$

and uses $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ *space,*

- *Algorithm 7 (*HYPERPLANEPARTITIONING*) called with* $(T_1,\ldots,T_b)$ *as a parameter runs in time* $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ *per one evaluation.*

Proof. We first prove the runtime guarantee and then we show correctness.

Runtime. The first step of FINDCENTERS (Algorithm 10) is to call INITIALIZEORACLE$(G,1/2)$ (Algorithm 4) which by Lemma 43 runs in time $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ and uses $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ space (It's the preprocessing time in the statement of Lemma 43). Then Algorithm 10 repeats the following procedure $O(\log(1/\eta))$ times.

It tests all partitions of a set of sampled vertices of size $s=O(\frac{\varphi^2}{\epsilon}\cdot k^4\log(k))$ into k sets. There is at most $k^s=2^{O(\frac{\varphi^2}{\epsilon}\cdot k^4\log^2(k))}$ of them. Notice that for each partition each $\widehat{\mu}_i$ is defined as

$$\widehat{\mu}_i := \frac{1}{|P_i|}\sum_{x\in P_i} f_x,$$

so as the number of sampled points is $O(\frac{\varphi^2}{\epsilon}\cdot k^4\log(k))$ then each $\widehat{\mu}_i$ is an average of at most $O(\frac{\varphi^2}{\epsilon}\cdot k^4\log(k))$ points. To analyze the runtime notice that:

- For each partition Algorithm 10 runs Algorithm 8,
- Algorithm 8 invokes Algorithm 11 (OUTERCONDUCTANCE) $k^{O(1)}$ times,
- OUTERCONDUCTANCE takes, by Lemma 44, $(s_1+s_2)\cdot\frac{1}{\varphi^2}\cdot s^4\cdot\left(\frac{\varphi^2}{\epsilon}k\right)^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\log^2(n)$ time,
- $s_1=\Theta(\frac{\varphi^2}{\epsilon}k^5\log^2(k)\log(1/\eta))$ and $s_2=\Theta(\frac{\varphi^4}{\epsilon^2}k^5\log^2(k)\log(1/\eta))$.

So in total the runtime of FINDCENTERS is

$$\frac{1}{\varphi^2}\left(\frac{\varphi^2}{\epsilon}k\right)^{O(1)} n^{1/2+O(\epsilon/\varphi^2)}\log^3(n)+\log(1/\eta)2^{O(\frac{\varphi^2}{\epsilon}\cdot k^4\log^2(k))}k^{O(1)}(s_1+s_2)\frac{s^4}{\varphi^2}\left(\frac{\varphi^2}{\epsilon}k\right)^{O(1)} n^{1/2+O(\epsilon/\varphi^2)}\log^2(n)$$

Substituting for s,s_1,s_2 it simplifies to:

$$\frac{1}{\varphi^2}\log^2(1/\eta)\cdot 2^{O(\frac{\varphi^2}{\epsilon}\cdot k^4\log^2(k))}\cdot n^{1/2+O(\epsilon/\varphi^2)}\log^3(n)$$

Runtime of Algorithm 7: Each $\widehat{\mu}_i$ is an average of at most s points, where $s\le O(\frac{\varphi^2}{\epsilon}\cdot k^4\log(k))$, Algorithm 7 performs $k^{O(1)}$ tests $\left\langle f_x,\widehat{\Pi}(\widehat{\mu})\right\rangle_{apx}\ge 0.93||\widehat{\Pi}(\widehat{\mu})||^2$ and by Lemma 43 each test takes $\widetilde{O}_\varphi\left(s^4\cdot\left(\frac{k}{\epsilon}\right)^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ time. So in total the runtime of one invokation of CLASSIFYBYHYPERPLANEPARTIONING$(\cdot,(T_1,\ldots,T_b))$ is in:

$$\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)}\cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$$

Error of OuterConductance algorithm. Now we analyze the error probabilities of OuterConductance across all the iterations of our algorithm. Note that we run the test for each cluster for each partition and for each of the $\log(2/\eta)$ iterations of the algorithm. So in total we run OuterConductance test $2^{O(\frac{\varphi^2}{\epsilon}\cdot k^4\log(k)^2)}k\log\left(\frac{2}{\eta}\right)$ times. By setting s_1 in

$$O\left(k\left(\log(4/\eta)+\log(k\log(1/\eta))+\frac{\varphi^2}{\epsilon}\cdot k^4\log^2(k)\right)\right)\le O\left(\frac{\varphi^2}{\epsilon}\cdot k^5\cdot\log^2(k)\cdot\log(1/\eta)\right),$$

and s_2 in:

$$O\left(\frac{\varphi^2\cdot k}{\epsilon}\left(\log(4/\eta)+\log(k\log(1/\eta))+\frac{\varphi^2}{\epsilon}\cdot k^4\log^2(k)\right)\right)\le O\left(\frac{\varphi^4}{\epsilon^2}\cdot k^5\cdot\log^2(k)\cdot\log(1/\eta)\right),$$

we get by Lemma 44 that the probability that the conclusion of Lemma 44 is not satisfied in a single run is bounded by

$$\frac{\eta}{100\cdot 2^{\Omega\left(\frac{\varphi^2}{\epsilon}\cdot k^4\log^2(k)\right)}k\log\left(\frac{1}{\eta}\right)}$$

So by union bound over the clusters, the partitions and the iterations we conclude that with probability $1-\frac{\eta}{50}$ the algorithm for every invokation returns a value satisfying the statement of Lemma 44. Moreover observe that this also means that $\mathcal{E}_{\text{conductance}}$ is true as conclusions of Lemma 44 are stronger than the property required for event $\mathcal{E}_{\text{conductance}}$ to be true.

W.h.p. every returned ordered partition defines a good clustering. By the lower bound on the error probability of OuterConductance algorithm above we get that with probability $1-\frac{\eta}{50}$ every cluster $\widehat{C}$ that passes the test from line 6 of Algorithm 8 has to satisfy:

$$\phi(\widehat{C})\le O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right),$$

as for $\widehat{C}$ to pass the test the value q returned by OuterConductance has to satisfy $q\le O\left(\frac{\epsilon}{\varphi^2}\cdot\log(k)\right)$ but by Lemma 44 we have $q\ge\frac{1}{2}\phi\left(\widehat{C}_{\widehat{\mu}}^{(T_1,\dots,T_b)}\right)-\epsilon/\varphi^2$. Now by Lemma 16 this implies that if Algorithm 10 returns an ordered partition, then with probability $1-\frac{\eta}{50}$ the collection of clusters it defines satisfies the statement of the Theorem.

Each iteration succeeds with constant probability. In the remaining part of the proof we will show that a clustering is accepted with probability $1-\frac{\eta}{2}$. First note that from the paragraph **Error of OuterConductance algorithm** we know that $\mathcal{E}_{\text{conductance}}$ holds with probability $1-\frac{\eta}{50}$. Next we show that in each iteration of the outermost for loop of Algorithm 10 it succeeds with probability $1/2$ (conditioned on $\mathcal{E}_{\text{conductance}}$). By amplification this will imply our result.

Now consider one iteration. Let S be the set of sampled vertices. Observe that there exists a partition of $S=P_1\cup P_2\cup\cdots\cup P_k$ such that for all $i\in[k]$, $P_i=S\cap C_i$. We set $s=10^{15}\cdot\frac{\varphi^2}{\epsilon}\cdot k^4\log(k)$. Therefore by Lemma 40 with probability at least $\frac{9}{10}$ we have for all $i\in[k]$

$$|S\cap C_i|\ge\frac{0.9\cdot s}{k}\cdot\min_{p,q\in[k]}\frac{|C_p|}{|C_q|}\ge 9\cdot10^{14}\cdot\frac{\varphi^2}{\epsilon}\cdot k^3\log(k).$$

Let $\delta=k^{-50}$ and $\zeta=\frac{10^{-6}\sqrt{\epsilon}}{\varphi\cdot k}$. Therefore, we have

$$|S\cap C_i|\ge 9\cdot10^{14}\cdot\frac{\varphi^2}{\epsilon}\cdot k^3\log(k)\ge c\cdot\left(k\cdot\log\left(\frac{k}{\delta}\right)\cdot\left(\frac{1}{\delta}\right)^{(80\cdot\epsilon/\varphi^2)}\cdot\left(\frac{1}{\zeta}\right)^2\right)^{1/(1-(80\cdot\epsilon/\varphi^2))}$$

where c is the constant from Lemma 39. The last inequality holds since $\frac{\epsilon}{\varphi^2}\log(k)$ is smaller than a sufficiently small constant, hence, $\left(\frac{\varphi^2}{\epsilon}\right)^{(\epsilon/\varphi^2)}\in O(1)$, and $k^{(\epsilon/\varphi^2)}\in O(1)$. Therefore by Lemma 39 for all $i\in[k]$ with probability at least $1-k^{-50}$ we have:

$$\|\widehat{\mu}_i-\mu_i\|_2\le\zeta\cdot\|\mu_i\|_2=\frac{10^{-6}\sqrt{\epsilon}}{\varphi\cdot k}\|\mu_i\|_2.$$

Hence, by union bound over all sets P_i, with probability at least $\frac{9}{10} - k \cdot k^{-50} \geq \frac{7}{8}$ we get $\|\widehat{\mu}_i - \mu_i\|_2 \leq \frac{10^{-6}\sqrt{\epsilon}}{\varphi \cdot k}\|\mu_i\|_2$ for all $i \in [k]$ simultaneously.

Now by Theorem 2 and the union bound we get that spectral dot product oracle succeeds with probability $1-n^{-48}$. So by Lemma 46 and the union bound FINDCENTERS with probability $\frac{7}{8}-n^{-48} \geq \frac{1}{2}$ returns an ordered partition $(T_1, \ldots, T_b)$ which induces a collection of clusters $\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ such that there exists a permutation π on k elements such that for all $i \in [k]$:

$$\left|\widehat{C}_{\widehat{\mu}_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|$$

and

$$\phi(\widehat{C}_{\widehat{\mu}_i}) \leq O\left(\frac{\epsilon}{\varphi^2} \cdot \log(k)\right).$$

□

6.6 LCA

Now we prove the main result of the paper. Recall that a clustering oracle (Definition 4) is a randomized algorithm that when given query access to a d-regular graph $G = (V, E)$ that admits (k, φ, ϵ)-clustering $C_1, \ldots, C_k$ it provides consistent access to a **partition** $\widehat{C}_1, \ldots, \widehat{C}_k$ such that there exists a permutation π on k elements such that for all $i \in [k]$:

$$\left|\widehat{C}_{\widehat{\mu}_i} \triangle C_{\pi(i)}\right| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|. \tag{233}$$

Consistency means that a vertex $x \in V$ is classified in the same way every time it is queried.

First we will show a Proposition (Proposition 3) that shows that it is enough to design an algorithm that returns a **collection of disjoint clusters** (not necessarily a partition) that satisfies (233) to get a clustering oracle. Using this Proposition as a reduction we then show Theorem 3, which is the main Theorem of the paper.

Proposition 3. *If there exists a randomized algorithm $\mathcal{O}$ that when given query access to a d-regular graph $G = (V, E)$ that admits a (k, φ, ϵ)-clustering $C_1, \ldots, C_k$, the algorithm $\mathcal{O}$ provides consistent query access to a* ***collection of disjoint clusters*** *$\mathcal{C} = (\widehat{C}_1, \ldots, \widehat{C}_k)$ of V. The collection $\mathcal{C}$ is determined solely by G and the algorithm's random seed. Moreover, with probability at least $9/10$ over the random bits of $\mathcal{O}$ the collection $\mathcal{C}$ has the following property: for some permutation π on k elements one has for every $i \in [k]$:*

$$|C_i \triangle \widehat{C}_{\pi(i)}| \leq O\left(\frac{\epsilon}{\varphi^3}\right) |C_i|.$$

Then if clusters have equal sizes and $\frac{\epsilon \cdot n}{\varphi^3 \cdot k \log(k)}$ is bigger than a constant then there exists an algorithm $\mathcal{O}'$ that is a (k, φ, ϵ)-clustering oracle with the same running time and space up to constant factors.

Proof. The idea is to assign the points outside $\bigcup_{i \in [k]} \widehat{C}_i$ randomly. That is to assign vertex $x \in V$, $\mathcal{O}'$ works exactly the same like $\mathcal{O}$ but if $\mathcal{O}$ left x unassigned then $\mathcal{O}'$ assigns x to a value chosen from $[k]$ uniformly at random.

Let $R = V \setminus \bigcup_{i \in [k]} \widehat{C}_i$ and for every $i \in [k]$ let $S_i \subseteq R$ be the set of vertices that were randomly assigned to $\widehat{C}_i$. By the fact that for every $i \in [k]$ $|C_i \triangle \widehat{C}_{\pi(i)}| \leq O\left(\frac{\epsilon}{\varphi^3}\right) |C_i|$ we get that there exists a constant C such that:

$$|R| \leq C \cdot \frac{\epsilon}{\varphi^3} \cdot n. \tag{234}$$

Now let $i \in [k]$. By the Chernoff bound we have that for every $\delta \geq 1$:

$$P\left[\left||S_i| - \frac{|R|}{k}\right| \geq \delta \frac{|R|}{k}\right] \leq e^{-\delta \frac{|R|}{3 \cdot k}} \tag{235}$$

Setting $\delta = \frac{C \cdot \epsilon \cdot n}{\varphi^3 \cdot |R|}$ we get:

$$P\left[\left||S_i| - \frac{|R|}{k}\right| \geq C \cdot \frac{\epsilon}{\varphi^3} \cdot \frac{n}{k}\right] \leq e^{-\frac{C \cdot \epsilon \cdot n}{3 \cdot \varphi^3 \cdot k}} \tag{236}$$

Combining (234) and (236) and the assumption that $\frac{\epsilon \cdot n}{\varphi^3 \cdot k \log(k)}$ is bigger than a constant we get that

$$P\left[|S_i| \geq 2C \cdot \frac{\epsilon}{\varphi^3} \cdot \frac{n}{k}\right] \leq \frac{1}{100 \cdot k}$$

Using the union bound we get that with probability $9/10 - k \cdot \frac{1}{100 \cdot k} \geq 8/10$ we have that for every $i \in [k]$ $|S_i| \leq 2C \cdot \frac{\epsilon}{\varphi^3} \cdot \frac{n}{k}$. So finally with probability 8/10 for every $i \in [k]$:

$$\begin{aligned}
|C_i \triangle (\widehat{C}_{\pi(i)} \cup S_{\pi(i)})| &\leq |C_i \triangle \widehat{C}_{\pi(i)}| + |S_{\pi(i)}| \\
&\leq O\left(\frac{\epsilon}{\varphi^3}\right) \cdot |C_i| + O\left(\frac{\epsilon}{\varphi^3}\right) \cdot \frac{n}{k} && \text{By definition of } \mathcal{O} \\
&\leq O\left(\frac{\epsilon}{\varphi^3}\right) \cdot |C_i| && \text{As } \frac{\max_{p \in [k]} |C_p|}{\min_{p \in [k]} |C_p|} = O(1),
\end{aligned}$$

which means that $\mathcal{O}'$ is a (k, φ, ϵ)-clustering oracle. □

Theorem 3. *For every integer $k \geq 2$, every $\varphi \in (0,1)$, every $\epsilon \ll \frac{\varphi^3}{\log k}$, every $\delta \in (0, 1/2]$ there exists a (k, φ, ϵ)-clustering oracle that:*

- *has $\widetilde{O}_\varphi\left(2^{O\left(\frac{\varphi^2}{\epsilon} k^4 \log^2(k)\right)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)}\right)$ preprocessing time,*
- *has $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{\delta+O(\epsilon/\varphi^2)}\right)$ query time,*
- *uses $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)}\right)$ space,*
- *uses $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{O(\epsilon/\varphi^2)}\right)$ random bits,*

where O_φ suppresses dependence on φ and $\widetilde{O}$ hides all polylog(n) *factors.*

Proof. By Theorem 8 we get that there exists an algorithm that runs in $\widetilde{O}_\varphi\left(2^{O(\frac{\varphi^2}{\epsilon} \cdot k^4 \log^2(k))} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ time and that with probability 9/10 returns an ordered partition $(T_1, \ldots, T_b)$ of $\{\widehat{\mu}_1, \ldots, \widehat{\mu}_k\}$ such that the induced collection of clusters $\{\widehat{C}_{\widehat{\mu}_1}, \ldots, \widehat{C}_{\widehat{\mu}_k}\}$ satisfies the following. There exists a permutation π on k elements such that for every $i \in [1, \ldots, k]$:

$$|C_{\pi(i)} \triangle \widehat{C}_{\widehat{\mu}_i}| \leq O\left(\frac{\epsilon}{\varphi^3} \cdot \log(k)\right) |C_{\pi(i)}|$$

That algorithm is the preprocessing step of oracle $\mathcal{O}$. Then for each query $x_i \in V$ we run Algorithm 7 which outputs $\widehat{\mu}_j$ such that $x_i \in \widehat{C}_{\widehat{\mu}_j}$ (Note that x_i might not belong to any of $\widehat{C}_{\widehat{\mu}_i}$, see Proposition 3 for how to deal with that). Algorithm 7 by Theorem 8 runs in $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{1/2+O(\epsilon/\varphi^2)}\right)$ time.

Runtime tradeoff. Notice however that by Theorem 2 we can achieve a tradeoff in the preprocessing/query runtime and achieve $\widetilde{O}_\varphi\left(2^{O(\frac{\varphi^2}{\epsilon} \cdot k^4 \log^2(k))} \cdot n^{1-\delta+O(\epsilon/\varphi^2)}\right)$ for preprocessing time and $\widetilde{O}_\varphi((\frac{k}{\epsilon})^{O(1)} \cdot n^{1-\delta+O(\epsilon/\varphi^2)})$ space and $\widetilde{O}_\varphi\left(\left(\frac{k}{\epsilon}\right)^{O(1)} \cdot n^{\delta+O(\epsilon/\varphi^2)}\right)$ for query time.

Random bits. The only thing left to prove is to show that we can implement these two algorithms in LCA model using few random bits. There are couple of places in our Algorithms where we use randomness.

First in InitializeOracle (Algorithm 4) we sample $\widetilde{\Theta}(n^{O(\epsilon/\varphi^2)} \cdot k^{O(1)})$ random points. For that we need $\widetilde{\Theta}(n^{O(\epsilon/\varphi^2)} \cdot k^{O(1)})$ random bits.

For generating random walks in Algorithm 4 and Algorithm 5 we need the following number of random bits. Notice that in all the proofs (see Lemma 26) we only need 4-wise independence of random walks. That means that we can implement generating these random walks using a hash function $h(x)$ that for vertex $x \in V$ generates $O(\log(d) \cdot \frac{1}{\varphi^2} \cdot \log(n))$ bit string that can be interpreted as encoding a random walk of length $O(\frac{1}{\varphi^2} \cdot \log(n))$ (remember that graphs we consider are d-regular so $\log(d)$ bits is enough to encode a neighbour). It's enough for the hash function to be 4-wise independent so it can be implemented using $O(\frac{1}{\varphi^2} \cdot \log(d) \cdot \log(n)) = \widetilde{O}_\varphi(1)$ random bits.

The partitioning scheme (see Algorithm 7) works in $O(\log(k))$ adaptive stages. The stages are adaptive, that is why we use fresh randomness in every stage. For a single stage we observe that in the proof of Lemma 44 we only use Chernoff type bounds. So by [SSS93] we don't need fully independent random variables. In our case it's enough to have $O(\log(n))$-wise independent random variables which can be implemented as hash functions using $O(\log^2(n))$ random bits. This means that in total we need $O(\log(k)\log^2(n)) = \widetilde{O}(1)$ random bits for this.

For sampling set S in Algorithm 10 we can use $O(\frac{\varphi^2}{\epsilon} \cdot k^4 \log(k) \cdot \log(n)) = \widetilde{O}_\varphi(\frac{1}{\epsilon} \cdot k^{O(1)})$ fresh random bits.

So finally the total number of random bits we need is in:

$$\widetilde{O}_\varphi\left(n^{O(\epsilon/\varphi^2)} \cdot k^{O(1)} + 1 + 1 + \frac{1}{\epsilon} \cdot k^{O(1)}\right) \leq \widetilde{O}_\varphi\left(\frac{1}{\epsilon} \cdot n^{O(\epsilon/\varphi^2)} \cdot k^{O(1)}\right)$$

□

Remark 11. *Note that threshold sets $C_{y,\theta}$ (recall Definition 8) are well defined in LCA model because for all $x, y \in V$ whenever we compute $\langle f_x, f_y \rangle_{apx}$ the result is the same as we use consistent randomness (see Definition 4).*

References

[Abb18] Emmanuel Abbe. Community detection and stochastic block models. *Foundations and Trends in Communications and Information Theory*, 14(1-2):1–162, 2018.

[ACL08] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. Local partitioning for directed graphs using pagerank. *Internet Mathematics*, 5(1):3–22, 2008.

[AGPT16] Reid Andersen, Shayan Oveis Gharan, Yuval Peres, and Luca Trevisan. Almost optimal local graph clustering using evolving sets. *J. ACM*, 63(2):15:1–15:31, 2016.

[ALM13] 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*, pages 396–404, 2013.

[ARVX12] Noga Alon, Ronitt Rubinfeld, Shai Vardi, and Ning Xie. Space-efficient local computation algorithms. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 1132–1139, 2012.

[AS12] Pranjal Awasthi and Or Sheffet. Improved spectral-norm bounds for clustering. In Anupam Gupta, Klaus Jansen, José D. P. Rolim, and Rocco A. Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15-17, 2012. Proceedings*, volume 7408 of *Lecture Notes in Computer Science*, pages 37–49. Springer, 2012.

[BJ06] Francis R. Bach and Michael I. Jordan. Learning spectral clustering, with application to speech separation. *J. Mach. Learn. Res.*, 7:1963–2001, 2006.

[BMD+15] Dino Bellugi, David G Milledge, William E Dietrich, Jim A McKean, J Taylor Perron, Erik B Sudderth, and Brian Kazian. A spectral clustering search algorithm for predicting shallow landslide size and location. *Journal of Geophysical Research: Earth Surface*, 120(2):300–324, 2015.

[CKCLL+13] 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. *Proceedings of the Annual ACM Symposium on Theory of Computing*, 01 2013.

[CKK+18] Ashish Chiplunkar, Michael Kapralov, Sanjeev Khanna, Aida Mousavifar, and Yuval Peres. Testing graph clusterability: Algorithms and lower bounds. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 497–508. IEEE, 2018.

[CPS15] Artur Czumaj, Pan Peng, and Christian Sohler. Testing cluster structure of graphs. In *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, 2015.

[CS10] Artur Czumaj and Christian Sohler. Testing expansion in bounded-degree graphs. *Combinatorics, Probability & Computing*, 19(5-6):693–709, 2010.

[DK] Chandler Davis and William Morton Kahan.

[DK70] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

[GR11] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In Oded Goldreich, editor, *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation - In Collaboration with Lidor Avigad, Mihir Bellare, Zvika Brakerski, Shafi Goldwasser, Shai Halevi, Tali Kaufman, Leonid Levin, Noam Nisan, Dana Ron, Madhu Sudan, Luca Trevisan, Salil Vadhan, Avi Wigderson, David Zuckerman*, volume 6650 of *Lecture Notes in Computer Science*, pages 68–75. Springer, 2011.

[HJ90] Roger A. Horn and Charles R. Johnson. *Matrix analysis.* Cambridge University Press, 1990.

[KPS13] Satyen Kale, Yuval Peres, and C. Seshadhri. Noise tolerance of expanders and sublinear expansion reconstruction. *SIAM J. Comput.*, 42(1):305–323, 2013.

[KS08] Satyen Kale and C. Seshadhri. An expansion tester for bounded degree graphs. In Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, editors, *Automata, Languages and Programming, 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July 7-11, 2008, Proceedings, Part I: Tack A: Algorithms, Automata, Complexity, and Games*, volume 5125 of *Lecture Notes in Computer Science*, pages 527–538. Springer, 2008.

[KVV04] Ravi Kannan, Santosh S. Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, 2004.

[LGT14] James R Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM (JACM)*, 61(6):37, 2014.

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 1131–1140, New York, NY, USA, 2012. ACM.

[NJW02] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002.

[NS10] Asaf Nachmias and Asaf Shapira. Testing the expansion of a graph. *Inf. Comput.*, 208(4):309–314, 2010.

[OA14] Lorenzo Orecchia and Zeyuan Allen Zhu. Flow-based algorithms for local graph clustering. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1267–1286, 2014.

[PCS06] Alberto Paccanaro, James A Casbon, and Mansoor AS Saqi. Spectral clustering of protein sequences. *Nucleic acids research*, 34(5):1571–1580, 2006.

[Pen20] Pan Peng. Robust clustering oracle and local reconstructor of cluster structure of graphs. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 2953–2972. SIAM, 2020.

[RTVX11] Ronitt Rubinfeld, Gil Tamir, Shai Vardi, and Ning Xie. Fast local computation algorithms. In *Proceedings of Innovations in Computer Science - ICS 2010, Tsinghua University, Beijing, China, January 7-9, 2011*, pages 223–238, 2011.

[Sin16] Ali Kemal Sinop. How to round subspaces: A new spectral clustering algorithm. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1832–1847, 2016.

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Departmental Papers (CIS)*, page 107, 2000.

[SSS93] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-hoeffding bounds for applications with limited independence. In *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '93, pages 331–340, Philadelphia, PA, USA, 1993. Society for Industrial and Applied Mathematics.

[ST14] 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.

[Tod11] Alexis Akira Toda. Operator reverse monotonicity of the inverse. *The American Mathematical Monthly*, 118(1):82–83, 2011.

[Tro12] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12(4):389–434, 2012.

[VL07] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.