

Higher-Order Cheeger Inequality for Partitioning with Buffers

Konstantin Makarychev*
Northwestern

Yury Makarychev†
TTIC

Liren Shan‡
TTIC

Aravindan Vijayaraghavan§
Northwestern

Abstract

We prove a new generalization of the higher-order Cheeger inequality for partitioning with buffers. Consider a graph $G = (V, E)$. The buffered expansion of a set $S \subseteq V$ with a buffer $B \subseteq V \setminus S$ is the edge expansion of S after removing all the edges from set S to its buffer B . An ε -buffered k -partitioning is a partitioning of a graph into disjoint components P_i and buffers B_i , in which the size of buffer B_i for P_i is small relative to the size of P_i : $|B_i| \leq \varepsilon |P_i|$. The buffered expansion of a buffered partition is the maximum of buffered expansions of the k sets P_i with buffers B_i . Let $h_G^{k,\varepsilon}$ be the buffered expansion of the optimal ε -buffered k -partitioning, then for every $\delta > 0$,

$$h_G^{k,\varepsilon} \leq O_\delta(1) \cdot \left(\frac{\log k}{\varepsilon} \right) \cdot \lambda_{\lfloor (1+\delta)k \rfloor},$$

where $\lambda_{\lfloor (1+\delta)k \rfloor}$ is the $\lfloor (1+\delta)k \rfloor$ -th smallest eigenvalue of the normalized Laplacian of G .

Our inequality is constructive and avoids the “square-root loss” that is present in the standard Cheeger inequalities (even for $k = 2$). We also provide a complementary lower bound, and a novel generalization to the setting with arbitrary vertex weights and edge costs. Moreover our result implies and generalizes the standard higher-order Cheeger inequalities and another recent Cheeger-type inequality by Kwok, Lau, and Lee (2017) involving robust vertex expansion.

*The author was supported by the NSF Awards CCF-1955351, CCF-1934931, and ECCS-2216970.

†The author was supported by the NSF Awards CCF-1955173, CCF-1934843, and ECCS-2216899.

‡The author was supported by the NSF Awards CCF-1955351, CCF-1934931, and ECCS-2216970.

§The author was supported by the NSF Awards CCF-1934931, CCF-2154100, and ECCS-2216970.

1 Introduction

Cheeger's inequality is a fundamental result in spectral graph theory that relates the connectivity of a graph to the eigenvalues of the Laplacian matrix associated with the graph. Consider an undirected d -regular graph $G = (V, E)$ on n vertices. Let L_G be the normalized Laplacian of the graph defined by $L_G = I - \frac{1}{d}A$, where A is the adjacency matrix of the graph G . Let $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n \leq 2$ be the eigenvalues of L_G . For every vector $z \in \mathbb{R}^V$ with coordinates $z(u)$ (where $u \in V$),

$$(1.1) \quad z^T L_G z = \frac{1}{d} \sum_{(u,v) \in E} (z(u) - z(v))^2.$$

For a set $S \subseteq V$, let $\delta_G(S, V \setminus S)$ denote the number of edges in the graph crossing the cut $(S, V \setminus S)$. The Cheeger constant or expansion of the graph G is

$$h_G := \min_{S \subseteq V: |S| \leq |V|/2} \phi_G(S), \text{ where } \phi_G(S) := \frac{\delta_G(S, V \setminus S)}{d|S|},$$

is called the expansion of the cut $S, V \setminus S$. Cheeger's inequality by Alon and Milman [AM85, Alo86, Che69] states that

$$(1.2) \quad \frac{\lambda_2}{2} \leq h_G \leq \sqrt{2\lambda_2}.$$

Similar inequalities also hold for graph partitioning into k parts [LRTV12, LGT14]. Here is a *higher order* Cheeger inequality by Lee, Oveis-Gharan and Trevisan [LGT14] (see also the paper [LRTV12] by Louis, Raghavendra, Tetali and Vempala): For every $\delta > 0$,¹

$$(1.3) \quad \frac{\lambda_k}{2} \leq h_G^k \leq O_\delta(\sqrt{\log k}) \cdot \sqrt{\lambda_{\lfloor (1+\delta)k \rfloor}},$$

where λ_i is the i -th smallest eigenvalue of the normalized Laplacian L_G , and

$$h_G^k = \min_{\substack{\text{partitions} \\ P_1, \dots, P_k \text{ of } V}} \max_{i \in [k]} \phi_G(P_i).$$

The upper bounds in (1.2) and (1.3) are constructive, which means that there is a polynomial-time algorithm that finds a partitioning $P_1, \dots, P_k$ using a spectral embedding of G , an embedding of the graph vertices into $\mathbb{R}^{k'}$ based on the first $k' = \lfloor (1+\delta)k \rfloor$ eigenvectors of the Laplacian. Similar spectral algorithms are commonly used in practice [NJW01, McS01]. We refer the reader to examples of applications of Cheeger's inequality to spectral clustering [KVV04, ST07, Spi07], image segmentation [SM00], random sampling and approximate counting [SJ89]. Cheeger's inequality is widely used in combinatorics and graph theory. Higher-order Cheeger inequalities also have connections to the small-set expansion conjecture [RS10, RST12], an important problem in the area of approximation algorithms.

The objective of abovementioned k -way graph partitioning algorithms is to find the Sparsest k -Partition of the graph i.e., a partition $P_1, \dots, P_k$ that *minimizes* the value of $\max_{i \in [k]} \phi_G(P_i)$. Together the lower and upper bounds (1.3) give a bound on the cost of the algorithmic solution in terms of the optimal solution: $\max_{i \in [k]} \phi_G(P_i) \leq O_\delta(\sqrt{\log k} \cdot h_G^{\lfloor (1+\delta)k \rfloor})$. This bound may be good for large values of $h_G^{\lfloor (1+\delta)k \rfloor}$ but can also be really bad for small values of $h_G^{\lfloor (1+\delta)k \rfloor}$. In fact, the approximation factor of such k -way partitioning algorithm may be as large as $\Omega(n)$ even for $k = 2$. It can be so large because the upper bound is non-linear – it has a “square-root loss”. To address this problem, several improved Cheeger inequalities under additional structural assumptions on the graph G have been presented in the literature [KLL⁺13, KLL17].

In this work, we introduce a new type of graph partitioning – partitioning with buffers – and prove a higher-order Cheeger inequality for them. Our inequality avoids the “square-root loss” and provides a constant bi-criteria approximation algorithm for the problems (see below for details). While being a natural problem, in and of itself,

¹The upper bound on h_G^k in [LRTV12] is $O(\sqrt{\log k})\sqrt{\lambda_{ck}}$ where $c > 1$ is an absolute constant.

our results for buffered partitioning also imply the standard higher-order Cheeger inequality (1.3) and a Cheeger-type inequality by Kwok, Lau, and Lee [KLL17] for robust vertex expansion (see Section 1.5). Finally, these Cheeger inequalities can also be extended to a more general setting with arbitrary vertex weights and edge costs: in contrast, we are not aware of such a generalization for the standard Cheeger inequalities i.e., without buffers.

1.1 Cheeger inequality for Buffered Partitions To simplify the exposition, we first present and discuss the setting where G is a d -regular graph. Then, in Section 1.2, we consider non-regular graphs G with arbitrary positive vertex weights and edge costs.

Multi-way Partitioning with Buffers. For every $\varepsilon \in [0, 1)$, an ε -buffered k -partitioning of an undirected graph $G = (V, E)$ is a collection of subsets $P_1, P_2, \dots, P_k \subset V$ and $B_1, B_2, \dots, B_k \subset V$ that satisfy the following conditions:

1. All sets P_i and B_j are pairwise disjoint (i.e., $P_i \cap P_j = \emptyset$, $B_i \cap B_j = \emptyset$, and $P_i \cap B_j = \emptyset$ for all $i, j \in \{1, \dots, k\}$);
2. $\bigcup_{i=1}^k (P_i \cup B_i) = V$;
3. Sets P_i are nonempty;
4. $|B_i| \leq \varepsilon |P_i|$ for all $i \in \{1, \dots, k\}$.

We say that B_i is the *buffer* for P_i . We denote this buffered partition by $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$. Now we define the buffered expansion of a set P with buffer B for d -regular graphs. Later, we will extend this definition to graphs with *arbitrary vertex weights and edge costs*. The buffered expansion of a set P with buffer B

$$\phi_G(P \parallel B) = \frac{\delta_G(P, V \setminus (P \cup B))}{d|P|}.$$

The definition is similar to that of the standard set expansion except we do not count edges from set S to its buffer B . Define the cost $\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ of a buffered partition:

$$(1.4) \quad \phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k) = \max_{i \in \{1, \dots, k\}} \phi_G(P_i \parallel B_i).$$

See Figure 5 on page 38 for an illustration of the edges that contribute towards the expansion $\phi_G(P_i \parallel B_i)$. The ε -buffered expansion of the graph $G = (V, E)$ is defined as the minimum value among all ε -buffered partitions:

$$(1.5) \quad h_G^{k, \varepsilon} = \min_{\substack{\varepsilon\text{-buffered } k\text{-partition} \\ (P_1, \dots, P_k \parallel B_1, \dots, B_k)}} \phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k).$$

Our main result is a new Cheeger-type inequality that relates buffered expansion to the eigenvalues of the Laplacian. We first state it for regular graphs. Consider a d -regular graph G . Let L_G be its normalized Laplacian and $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be its eigenvalues.

THEOREM 1.1. *For every $\delta \in (0, 1)$,*

$$(1.6) \quad h_G^{k, \varepsilon} \leq \frac{c(\delta) \log k}{\varepsilon} \cdot \lambda_{\lfloor (1+\delta)k \rfloor},$$

where $c(\delta)$ is a function that depends only on δ . Furthermore, there is a randomized polynomial-time algorithm that given G finds an ε -buffered k -partitioning $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ with $\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k) \leq \frac{c(\delta) \log k}{\varepsilon} \lambda_{\lfloor (1+\delta)k \rfloor}$.

Our algorithm in Theorem 1.1 uses the top $\lfloor (1+\delta)k \rfloor$ eigenvalues and eigenvectors of the Laplacian matrix L_G . Note that these eigenvalues and eigenvectors can be approximated to the desired precision in polynomial time. However, we ignore numerical/precision issues in this paper and in the remainder of this paper, assume that the algorithm has the exact values of eigenvalues and eigenvectors. As in the standard Cheeger-type inequality (1.3), we upper bound expansion for k -way partitioning in terms of $\lambda_{k'}$, where $k' = \lfloor (1+\delta)k \rfloor$ may be larger than k (depending on the value of $\delta > 0$). However, for every fixed k , we can let $\delta = 1/(k+1)$ and get the following result.

COROLLARY 1.1. For every k , $h_G^{k,\varepsilon} \leq \frac{c_k}{\varepsilon} \cdot \lambda_k$, where c_k depends only on k . Furthermore, there is a randomized polynomial-time algorithm that given G finds an ε -buffered k -partitioning $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ with $\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k) \leq \frac{c_k}{\varepsilon} \lambda_k$.

Theorem 1.2 presented later is a novel generalization of Theorem 1.1 to graphs with vertex weights and edge costs.

Approximation results The spectral graph partitioning algorithm provided by Theorem 1.1 can be seen as an $O_\delta\left(\frac{1}{\varepsilon} \log k\right)$ -pseudo-approximation algorithm for the k -way sparsest partitioning problem. It finds an ε -buffered k -partitioning $(P_1, \dots, P_k, B_1, \dots, B_k)$ with the maximum expansion bounded by $O_{\varepsilon,\delta}(\log k)$ times the cost of the true optimum solution of the non-buffered $\lfloor(1+\delta)k\rfloor$ -way partitioning problem. That is, the solution produced by our algorithm has an approximation factor of $O_{\varepsilon,\delta}(\log k)$ but (1) uses ε buffers around each set P_i , and (2) has fewer sets than the true optimal solution. This pseudo-approximation algorithm also works for non-regular graphs with vertex weights and edge costs. See Theorem C.1 for details. Applying this pseudo-approximation algorithm recursively, we get an $O(1/\varepsilon)$ -pseudo-approximation algorithm for the Buffered Balanced Cut problem (see Theorem D.1) and an $O(\log^2 k)$ pseudo-approximation algorithm for a buffered variant of the balanced k -partitioning problem (see Corollary D.1).

Let us examine some applications of buffered partitioning and our techniques.

Applications Spectral algorithms are widely used across several application domains because they are very fast and scalable in practice [PSL90, vL07]. For example, a standard off-the-shelf package finds the first 100 eigenvectors of the Twitter graph [LM12] in less than half a minute. This graph has 81 thousand nodes and 1.3 million edges. In contrast, linear programming and semidefinite programming based methods do not scale well and cannot handle such large graphs at the present time. This motivates the design of spectral algorithms for graph partitioning with stronger guarantees. Our work demonstrates that one can achieve very good theoretical guarantees for Buffered Sparsest k -Partitioning.

As mentioned earlier, the algorithms we present in this paper give an $O_{\varepsilon,\delta}(\log k)$ -pseudo-approximation for the Buffered Sparsest k -Partitioning problem, and a $O(1/\varepsilon)$ -pseudo-approximation for the Buffered Balanced Cut problem (see Section D). For constant ε , this corresponds to a constant factor approximation with buffers. For comparison, the best known approximation guarantees for Balanced Cut or Sparsest k -Cut without buffers incur logarithmic factors in the number of vertices n .² Similarly, the best known approximation for Sparsest k -Partitioning is $O_\delta(\sqrt{\log n \log k})$ [LM14]. The caveat is, of course, that our algorithm produces an ε -buffered partitioning but we compare its cost with the cost of the optimal non-buffered partitioning.

In applications of graph partitioning and clustering, relaxing the partitioning using buffers is often benign and even natural. Let us consider the following application of graph partitioning. Suppose we have a graph whose nodes represent user profiles in a social network (like the Twitter graph we mentioned earlier) and edges represent connections between them (friends, followers, etc). We would like to assign these profiles to two machines so that each machine is assigned about the same number of profiles and the number of separated connections is minimized. These are common requirement for graph processing systems. In other words, we need to solve the Balanced Cut problem for the given graph. If we run our algorithm on this graph, we will get two parts S , T and buffer B . We can store S and T on the first and second machines, respectively, and replicate nodes in B on both machines. This way we will separate only nodes located in S and T . Partitioning with buffers can be useful to obtain better solutions for several other applications such as resource allocation and scheduling, where graph partitioning is used.

Moreover, in applications like community detection, it is common for the communities to have small overlaps [YL14, YL12]. Vertices belonging to multiple communities may correspond to influential or well-connected nodes, that would disproportionately affect the cost in a disjoint partition. While there has been much recent interest in detecting overlapping communities, it is challenging to obtain algorithmic guarantees in the overlapping setting (see [KBL16, OATT22] for different formulations and results on this problem); in particular, there are very few theoretical results for spectral algorithms even in average-case models. An ε -buffered partitioning with sets S, T and buffer B can be viewed as two overlapping communities $S' = S \cup B$ and $T' = T \cup B$ with small overlap $|S \cap T| \leq \varepsilon \min\{|S|, |T|\}$. Hence ε -buffered partitions capture overlapping communities and allow us to

²For Balanced Cut without buffers, the best true approximation factor is $O(\log n)$ [AR04, R  c08], and the best pseudo-approximation is $O(\sqrt{\log n})$ [ARV09].

reason about spectral methods even in the overlapping setting (see also footnote 6).

Finally, buffered partitioning is an interesting problem in its own right, it gives a common, versatile generalization that captures important results in spectral graph theory including higher-order Cheeger inequalities and robust vertex expansion as described in the next few sections.

1.2 Graphs with vertex weights and edge costs In the standard Cheeger inequality, the weight of every vertex must be equal to the total weight of edges incident on it. For instance, in d -regular graphs, the weights of all vertices are equal to d .³ Surprisingly, we can generalize our variant of Cheeger's inequality to vertex weighted graphs. We show that the Cheeger inequality for buffered partitions also holds when graph $G = (V, E, w, c)$ has vertex weights $w_u > 0$ and edge costs $c_{uv} > 0$. In that case, we define the non-normalized Laplacian $\tilde{L}_G$ for G as follows. $\tilde{L}_G(u, u)$ is the total cost of all edges incident on u and $\tilde{L}_G(u, v) = -c_{uv}$ for $(u, v) \in E$; all other entries are zero. Then, for any vector $z \in \mathbb{R}^n$, we have

$$(1.7) \quad z^T \tilde{L}_G z = \sum_{(u,v) \in E} c_{uv} (z(u) - z(v))^2.$$

Further, we define the weight matrix D_w as follows: $D_w(u, u) = w_u$ and $D_w(u, v) = 0$ if $u \neq v$ (D_w is a diagonal matrix). Finally, we define the normalized Laplacian $L_G = D_w^{-1/2} \tilde{L}_G D_w^{-1/2}$. Note that

$$z^T L_G z = \sum_{(u,v) \in E} c_{uv} \left(\frac{z(u)}{w_u^{1/2}} - \frac{z(v)}{w_v^{1/2}} \right)^2.$$

Denote the weight of a set of vertices A by $w(A) = \sum_{u \in A} w_u$. We extend the definitions of $\delta_G(A, B)$, $\phi_G(P \parallel B)$, $\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k)$, and $h_G^{k, \varepsilon}$ to graphs with vertex weights and edge costs:

$$\delta_G(A, B) = \sum_{\substack{u \in A, v \in B \\ (u,v) \in E}} c_{uv} \quad \text{and} \quad \phi_G(P \parallel B) = \frac{\delta(P, V \setminus (P \cup B))}{w(P)}$$

Quantities $\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ and $h_G^{k, \varepsilon}$ are given by formulas (1.4) and (1.5), respectively. We say that partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ is ε -buffered if $w(B_i) \leq \varepsilon w(P_i)$ for every $i \in [k]$.

Note that the definitions of L_G , δ_G , ϕ_G , and $h_G^{k, \varepsilon}$ are consistent with those for regular graphs with unit vertex weights and unit edge costs. As a side note, we observe that the definition of L_G coincides with the definition of the normalized Laplacian in the standard Cheeger inequality for non-regular graphs with edge costs. Note that in that inequality, vertex weights are defined as $w_u = \sum_{v: (u,v) \in E} c_{uv}$. In contrast to the standard Cheeger inequality, our variant holds for arbitrary vertex weights and edge costs.

THEOREM 1.2. *Let $G = (V, E, w, c)$ be a graph with positive weights $w_u > 0$ and edge costs $c_{uv} > 0$, $\varepsilon \in [0, 1)$, $\delta \in (0, 1)$, and $k \geq 2$ be an integer. Assume that $\max_u w_u \leq \varepsilon w(V)/(3k)$. Then*

$$(1.8) \quad h_G^{k, \varepsilon} \leq \frac{\kappa(\delta) \log k}{\varepsilon} \cdot \lambda_{\lfloor (1+\delta)k \rfloor}(L_G),$$

where $\kappa(\delta)$ is a function that depends only on δ . Furthermore, there is a randomized polynomial-time algorithm that given G finds an ε -buffered k -partitioning $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ with $\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k) \leq \frac{\kappa(\delta) \log k}{\varepsilon} \lambda_{\lfloor (1+\delta)k \rfloor}(L_G)$.

This new generalization with vertex weights and edge costs is crucial for the pseudoapproximation guarantees for the buffered versions of Balanced Cut (Theorem D.1) and Balanced k -way partitioning (Theorem D.1) that were mentioned earlier.

³Admittedly, if vertex weights w_u and edge costs c_{uv} are arbitrary, we can apply the standard Cheeger inequality as follows. First, let us rescale all w_u by letting $w'_u = \alpha w_u$ where $\alpha = \max_u \frac{\sum_{v: (u,v) \in E} c_{uv}}{w_u}$. This ensures that $w'_u \geq \sum_{v: (u,v) \in E} c_{uv}$. Then, introduce loops (u, u) of cost $c_{uu} = w_u - \sum_{v: (u,v) \in E} c_{uv}$. After this step, $w'_u = \sum_{v: (u,v) \in E} c_{uv}$ and thus we can apply the standard Cheeger inequality. While this approach yields a Cheeger-type inequality, we lose a factor of $\sqrt{\alpha}$ in the approximation guarantee for h_G and h_G^k ; e.g., we get $h_G \leq \sqrt{\alpha} \cdot \sqrt{2\lambda_2}$ when $k = 2$. Note that α can be arbitrarily large. Further, if $w_u \ll \sum_{v: (u,v) \in E} c_{uv}$ even for a single vertex u , then $\alpha \gg 1$.

1.3 Buffered Cheeger's inequality for $k = 2$ For $k = 2$, we provide an alternative slightly simpler variant of buffered Cheeger's inequality. We give a polynomial-time algorithm that partitions V into three disjoint sets: parts S , T , and buffer B , satisfying $S \cup T \cup B = V$ and $|B| \leq \varepsilon \min(|S|, |T|)$. The buffered expansion of S and T , defined as $\delta(S, T) / \min(w(S), w(T))$ is at most $O(\lambda_2 / \varepsilon)$ (see Proposition 2.1 for details).

We provide a self-contained proof of this simpler result for $k = 2$ in Section 2. We remark that this result coupled with Lemma 5.1 from this paper and Theorem 4.6 from the paper by Lee, Oveis-Gharan, and Trevisan [LGT14] already yields weak versions of our main results (Theorems 1.1 and 1.2) where $O(\log k)$ is replaced with $O(\log^2 k)$. This extra logarithmic factor is a large loss in the context of graph partitioning problems, and this is analogous to the weaker higher order Cheeger inequality obtained in [LGT14] by combining Theorem 4.6 of [LGT14] with the standard Cheeger inequality for $k = 2$.⁴ To get a tight bound of $O(\frac{1}{\varepsilon} \log k)$, we design a new algorithm (see the next section for why our result is tight in both k and ε). We give an overview of new techniques in Section 1.7.

1.4 Our result generalizes higher-order Cheeger inequalities Our main result (Theorem 1.1) can be seen as a generalization of Cheeger's inequality (1.2) and the higher-order Cheeger inequality (1.3). To obtain these results, we apply Theorem 1.1 with $\varepsilon = \sqrt{\lambda_{\lfloor (1+\delta)k \rfloor} \log k}$. We find the largest set P_t among $P_1, \dots, P_k$. We may assume that P_t contains at least $\Omega(\delta n)$ vertices (see Section B for the details). Then we include all buffers in set P_t ; that is, we let $P'_t = P_t \cup \bigcup_i B_i$. We obtain a non-buffered partition of G . Using that $|B_i| \leq \varepsilon |P_i|$ and $\delta(P_i, B_i) \leq d|B_i|$ (since the graph is d -regular), we get for $i \neq t$ (here $k' = \lfloor (1+\delta)k \rfloor$),

$$\phi_G(P_i) = \phi_G(P_i \parallel B_i) + \frac{\delta(P_i, B_i)}{d|P_i|} \leq \frac{c(\delta) \log k}{\sqrt{\lambda_{k'} \log k}} \lambda_{k'} + \frac{d \cdot \sqrt{\lambda_{k'} \log k} |P_i|}{d|P_i|} = (c(\delta) + 1) \sqrt{\lambda_{k'} \log k}.$$

We bound $\phi_G(P'_t)$ (the expansion of the updated set P'_t) as follows,

$$\phi(P'_t) = \frac{\sum_{i \neq t} \delta(P_i, P'_t)}{d|P'_t|} \leq \frac{\sum_{i \neq t} \phi_G(P_i) \cdot |P_i| \cdot d}{\delta n \cdot d} \leq \frac{(c(\delta) + 1) \sqrt{\lambda_{k'} \log k}}{\Omega(\delta n)} \sum_{i \neq t} |P_i| \leq \frac{c(\delta) + 1}{\Omega(\delta)} \sqrt{\lambda_{k'} \log k}.$$

Hence Theorem 1.1 provides an alternate proof of (1.3). Furthermore, this proof suggests that the factor of $O(\frac{1}{\varepsilon} \log k)$ in the upper bound of Theorem 1.1 cannot be improved. It also shows that our inverse dependence on ε is tight even for $k = 2$ (as otherwise we would be able to strengthen Cheeger's inequality, which is known to be tight).

1.5 Connection to Robust Expansion Theorem 1.1 also generalizes the Cheeger-type inequality by Kwok, Lau, and Lee [KLL17] that gives a bound for λ_2 in terms of *robust expansion* [KLM06]. Let $\eta \in (0, 1)$. For $S \subseteq V$, define

$$(1.9) \quad N_\eta(S) = \min \left\{ |T| : T \subseteq V \setminus S, \delta_G(S, T) \geq (1 - \eta) \delta_G(S, V \setminus S) \right\}$$

$$(1.10) \quad \phi_\eta^V(S) = \frac{N_\eta(S)}{|S|} \quad \text{and} \quad \phi_\eta^V(G) = \min_{S: |S| \leq |V|/2} \phi_\eta^V(S)$$

In other words, $\phi_\eta^V(S)$ is the vertex expansion of set S after we remove an η fraction of the edges leaving S in the optimal way (which minimizes the vertex expansion of S in the remaining graph). Quantity $\phi_\eta^V(S)$ is less sensitive to additions of a small number of edges to graph G than the standard vertex expansion. For that reason, $\phi_\eta^V(S)$ is called the *robust vertex expansion* of G . Kwok, Lau, and Lee [KLL17] proved the following result for $\eta = 1/2$.

THEOREM 1.3. (SEE THEOREM 1 IN [KLL17]) $\lambda_2 = \Omega(h_G \cdot \phi_{1/2}^V(G))$.

The following generalization of Theorem 1.3 is an immediate corollary of Theorem 1.1 (see Appendix A for a proof).

COROLLARY 1.2. For every $\eta \in (0, 1)$ we have $\lambda_2 = \Omega(\eta \cdot h_G \cdot \phi_\eta^V(G))$.

We remark that Theorem 1.3 is related to the case $k = 2$ in Theorem 1.1.

⁴The stronger bound of Theorem 4.1 in [LGT14] avoids Theorem 4.6.

1.6 Lower Bounds We also prove a lower bound on $h_G^{k,\varepsilon}$, which is linear in λ_k .

THEOREM 1.4. *For every d -regular graph G , integer $k \geq 2$, and $\varepsilon > 0$, we have,*

$$h_G^{k,\varepsilon} \geq \frac{\lambda_k - \varepsilon}{2}.$$

We remark that the additive dependence on ε in the above lower bound (Theorem 1.4) is unavoidable even when $k = 2$.⁵ This is useful to derive a lower bound on the optimal buffered expansion $h_G^{k,\varepsilon}$; moreover in conjunction with the upper bound (applied with a larger ε'), one can also get a bicriteria approximation for buffered k -way partitioning.⁶

1.7 Overview and Organization We start with proving a weaker version of our main result (Theorem 1.1) for $k = 2$ in Section 2. This proof is significantly simpler than the general proof but nevertheless illustrates why we get a linear dependence on λ_k rather than a square-root dependence in our Cheeger-type inequality. In the proof, we use the thresholding idea from the proof of the standard Cheeger inequality but add an extra twist – use two thresholds instead of one. First, we compute the eigenvector u corresponding to the second smallest eigenvalue λ_2 of the normalized Laplacian L_G of G . Let $u(i)$ be the i -th coordinate of u . Recall that in the proof of Cheeger’s inequality, we put each vertex i either in S or in T , depending on whether $u(i)^2 \geq \tau$ or $u(i)^2 < \tau$ for an appropriately chosen threshold τ . To prove our inequality for $k = 2$, we use two thresholds τ and $(1 + \varepsilon)\tau$ and, loosely speaking (see Section 2 for the precise description), put i in T , B , S depending on whether $u(i)^2$ lies in $(-\infty, \tau]$, $(\tau, (1 + \varepsilon)\tau)$, or $[(1 + \varepsilon)\tau, \infty)$, respectively.

In the subsequent sections, we prove the main result i.e., Theorem 1.1 for arbitrary k . Recall the definition of the spectral embedding of graph G , which we use in our proof. Let $x_1, \dots, x_{k'}$ be the eigenvectors of L_G corresponding to the $k' = \lfloor (1 + \delta)k \rfloor$ smallest eigenvalues. Note that the coordinates of vectors x_i are indexed by vertices u ; denote the coordinate with index u by $x_i(u)$. The spectral embedding maps vertex u to vector $\bar{u} \in \mathbb{R}^{k'}$ with coordinates $x_1(u), \dots, x_{k'}(u)$. We compute the spectral embedding. And now our goal is to partition vectors $\bar{u}$ (so that the corresponding buffered partition satisfies the desired properties). To do so, we introduce a new technical tool – *orthogonal separators with buffers* – for partitioning sets of vectors.

Given a set of unit vectors, the orthogonal separator procedure generates three (disjoint) random sets – set X (called an orthogonal separator) and its two buffers Y and Z – such that

1. if $u \in X$ and v is close to u then v is in $X \cup Y \cup Z$ with high probability
2. if vectors u and v are far apart, then it is unlikely that both of them are in X
3. $|Y|, |Z|$ are at most $\varepsilon|X|$ in expectation

(See Theorems 3.1 and 3.2 for details.) Orthogonal separators with buffers provide a basic building block for constructing buffered partitionings. We repeatedly apply the orthogonal separator procedure to normalized vectors $\psi(\bar{u}) = \frac{\bar{u}}{\|\bar{u}\|}$ and obtain subsets X_t and their buffers Y_t, Z_t . Merging the obtained sets and filtering/thresholding them based on the lengths of vectors $\bar{u}$, we obtain a *partial* buffered partitioning. This partitioning has all the desired properties except that it does not necessarily cover the entire vertex set V . We show the desired properties of the partial buffered partitioning by using the properties of the orthogonal separator. We use item 1 to argue that the buffered expansion of each set P_i is small, item 2 to argue that the obtained sets are not too large and thus there are at least k sets in the partitioning, and item 3 to argue that $|B_i| \leq \varepsilon|P_i|$.

Note that orthogonal separators with buffers generalize (non-buffered) orthogonal separators introduced by Chlamtac, Makarychev, and Makarychev [CMM06] and used in a number of SDP-based approximation algorithms for graph partitioning problems. An analog of Theorem 3.2 for (non-buffered) orthogonal separators was first proved by Bansal, Feige, Krauthgamer, Makarychev, Nagarajan, Naor, and Schwartz [BFK⁺14] (see also [LM14]).

⁵For the tight example, consider two cliques on vertex sets A and B of size $(1 + \varepsilon)n/2$ each, with overlap of $|A \cap B| = \varepsilon n$ vertices and with no edges between $A \setminus B$ and $B \setminus A$. Some of the edges incident on $A \cap B$ are resampled to ensure (approximate) regularity. While $h_G^{2,\varepsilon} = 0$, it is easy to show that $\lambda_2 = \Omega(\varepsilon)$.

⁶Specifically, for any $\varepsilon \in [0, 1]$, $\delta \in (0, 1)$, and $\varepsilon' > \varepsilon$, our algorithm given a graph G finds an ε' -buffered k -partitioning $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ with $\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k) \leq c(\delta) \log k \cdot (h_G^{\lfloor k(1+\delta) \rfloor, \varepsilon} + \varepsilon)/\varepsilon'$, where $c(\delta) > 0$ is a constant that only depends on δ .

Our high level approach follows the paper by Louis and Makarychev [LM14]. However, our algorithm and its analysis substantially differ from theirs because we need to use *orthogonal separators with buffers* and keep track of the buffers between clusters. Also, our algorithm uses a spectral embedding while the algorithm by Louis and Makarychev [LM14] uses an embedding obtained from an SDP relaxation, which imposes additional constraints on vectors.

We prove some useful claims about the spectral embedding in Section 6. We define orthogonal separators with buffers and present the main theorem about them (Theorem 3.2) in Section 3. We prove Theorem 3.2 in Section 7. We show how to obtain a partial buffered clustering in Section 4. Finally, in Section 5, we show how to obtain a true buffered partitioning.

The proof of the Cheeger inequality for graphs with arbitrary vertex weights and edge costs (Theorem 1.2) is almost identical to that of Theorem 1.1. In order to simplify the exposition, we only present the proof of Theorem 1.1. The same proof with minimal changes works in the general case. Instead of presenting essentially the same proof again, we give a black box reduction from Theorem 1.1 to Theorem 1.2 in Appendix E. The reduction however may significantly increase the running time of the algorithm. We stress that the algorithm from Theorem 1.1 also works with weighted graphs.

The other sections and appendices are organized as follows. In Section A, we show that Theorem 1.1 implies Corollary 1.2, which we discussed in Section 1.5. In Section B, we prove a technical claim about ε -buffered partitions. In Section C, we prove a lower bound on h_G^k for unbuffered partitions of graphs G with vertex weights and edge costs. Combining this lower bound with Theorem 1.2, we get a pseudo-approximation algorithm for the Sparsest k -way Partitioning problem (Theorem C.1). In Section D, we present our pseudo-approximation algorithm for the Buffered Balanced Cut problem. In Section F, we prove Theorem 1.4 (a lower bound on $h_G^{k,\varepsilon}$ discussed above). In Section G, we give a few useful estimates on the Gaussian distribution, which we use throughout the paper.

Other related work. Clustering with vertex deletion and duplication has been studied in other context as well. We refer the reader to the following recent results: Filtser and Le [FL21], Haeupler, Hershkowitz, and Zuzic [HHZ21], Filtser [Fil22].

2 Warm up: Cheeger's Inequality with a Buffer for $k = 2$

As a warmup, we provide a self-contained proof of a weaker version of Theorem 1.1 for $k = 2$. Here, we will consider cuts (S, T) with a common buffer B (instead of disjoint buffers for S and T). Such cuts consist of three disjoint sets S , T , and B that partition the set of vertices V into three groups. We will refer to such a partition as $(S, T \parallel B)$. While there are many new ideas needed to obtain Theorem 1.1 in full generality, this simpler setting already demonstrates how one can leverage buffers to obtain an improved upper bound.

PROPOSITION 2.1. *Let $\varepsilon \in (0, 1/4)$. Consider any graph $G = (V, E)$ with positive vertex weights $w_u > 0$ and edge costs $c_{uv} > 0$. Let λ_G be the second smallest eigenvalue of $L_G = D_w^{-1/2} \tilde{L}_G D_w^{-1/2}$, the normalized Laplacian of G . Then, in polynomial time we can find three disjoint sets S, B, T with $S \cup B \cup T = V$, $w(S) \leq w(T)$ and $w(B) \leq \varepsilon w(S)$ such that*

$$\phi_G(S, T \parallel B) = \frac{|\delta(S, T)|}{w(S)} \leq 4 \left(1 + \frac{2}{\varepsilon}\right) \lambda_G.$$

Proof. The proof follows the same general strategy as the standard proof of the Cheeger inequality. We show how to find a distribution over (buffered) partitions (S, B, T) in the graph G , by thresholding the second eigenvector of L_G , such that:

$$(I) \quad \mathbf{E} |\delta(S, T)| \leq (1 + 1/\varepsilon) \lambda_G \cdot \mathbf{E}[w(S)] \quad \text{and} \quad (II) \quad \mathbf{E}[w(B)] \leq \varepsilon \mathbf{E}[w(S)].$$

The first condition gives an upper bound on the expected number of (non-buffered) edges crossing the cut, while the second condition gives a bound on the expected size of the buffer. A simple probabilistic argument (see Lemma 2.1) allows us to conclude that there exists a single buffered threshold cut that simultaneously satisfies both the properties (with some slack).

Consider the spectrum of matrix $L_G = D_w^{-1/2} \tilde{L}_G D_w^{-1/2}$. The first eigenvector of the non-normalized Laplacian $\tilde{L}_G$ is the vector of all ones denoted by $\mathbf{1}$. Its eigenvalue is 0. In other words, $\tilde{L}_G \mathbf{1} = 0$. Consequently,

$L_G(D_w^{1/2}\mathbf{1}) = D_w^{-1/2}\tilde{L}_G\mathbf{1} = 0$. Hence, $D_w^{1/2}\mathbf{1}$ is the first eigenvector of L_G . Let y be an eigenvector of L_G corresponding to the second eigenvalue $\lambda_G = \lambda_2$ of L_G . Then, $y \perp D_w^{1/2}\mathbf{1}$ and

$$(2.11) \quad \langle y, L_G y \rangle = \langle y, D_w^{-1/2}\tilde{L}_G D_w^{-1/2}y \rangle = \lambda_G \|y\|^2.$$

Let $v = D_w^{-1/2}y$. Then, we have $v \perp D_w\mathbf{1}$ (because $\langle v, D_w\mathbf{1} \rangle = \langle y, D_w^{1/2}\mathbf{1} \rangle = 0$) and

$$(2.12) \quad \langle v, \tilde{L}_G v \rangle = \langle D_w^{-1/2}y, \tilde{L}_G D_w^{-1/2}y \rangle = \lambda_G \|y\|^2 = \lambda_G \|D_w^{1/2}v\|^2.$$

Step 1. Splitting the vector. For technical reasons, we need to split vector v into two vectors v_+ and v_- such that the vertex weight of non-zero coordinates in each vector is at most $w(V)/2$,

$$w(\{i : v_+(i) > 0\}) \leq w(V)/2; \quad w(\{i : v_-(i) > 0\}) \leq w(V)/2.$$

We do this by following a standard trick that is often used in the proof of Cheeger's inequality. Let z denote the smallest coordinate value in the vector v such that the total vertex weight of coordinates with a value greater than z in vector v is at most $w(V)/2$, i.e.

$$w(\{i : v(i) > z\}) \leq w(V)/2; \quad w(\{i : v(i) < z\}) \leq w(V)/2.$$

Then we shift the entire vector v by z and get $v' = v - z\mathbf{1}$. Since $\tilde{L}_G\mathbf{1} = 0$ and $v \perp D_w\mathbf{1}$, we have

$$\langle v', \tilde{L}_G v' \rangle = \langle v, \tilde{L}_G v \rangle - \underbrace{2z\langle v, \tilde{L}_G \mathbf{1} \rangle}_{=0} + \underbrace{z^2\langle \mathbf{1}, \tilde{L}_G \mathbf{1} \rangle}_{=0} \stackrel{\text{by (2.12)}}{=} \lambda_G \|D_w^{1/2}v\|^2 \leq \lambda_G \|D_w^{1/2}v'\|^2.$$

The last inequality holds because

$$\|D_w^{1/2}v'\|^2 = \|D_w^{1/2}v\|^2 + z^2\|D_w^{1/2}\mathbf{1}\|^2 - 2z\langle D_w^{1/2}v, D_w^{1/2}\mathbf{1} \rangle = \|D_w^{1/2}v\|^2 + z^2 \underbrace{\|D_w^{1/2}\mathbf{1}\|^2}_{\geq 0} - 2z \underbrace{\langle v, D_w\mathbf{1} \rangle}_{=0}.$$

We now split the vector v' into two vectors v_+, v_- with disjoint supports as follows:

$$v_+(i) = \begin{cases} v(i) - z, & \text{if } v(i) \geq z; \\ 0, & \text{otherwise,} \end{cases} \quad v_-(i) = \begin{cases} 0, & \text{if } v(i) \geq z; \\ v(i) - z, & \text{otherwise.} \end{cases}$$

CLAIM 2.1. For $u = v_+$ or $u = v_-$, we have $u \neq 0$ and $\langle u, \tilde{L}_G u \rangle \leq \lambda_G \|D_w^{1/2}u\|^2$.

Proof. Vectors $D_w^{1/2}v_+$ and $D_w^{1/2}v_-$ are orthogonal because their supports are disjoint (note: $D_w^{1/2}$ is a diagonal matrix). All coordinates of $D_w^{1/2}v_+$ are non-negative, and all coordinates of $D_w^{1/2}v_-$ are non-positive. Thus, $\|D_w^{1/2}v_+\|^2 + \|D_w^{1/2}v_-\|^2 = \|D_w^{1/2}(v_+ + v_-)\|^2 = \|D_w^{1/2}v'\|^2$ and

$$\langle v', \tilde{L}_G v' \rangle = \langle v_+, \tilde{L}_G v_+ \rangle + \langle v_-, \tilde{L}_G v_- \rangle + \underbrace{2\langle v_-, \tilde{L}_G v_+ \rangle}_{\geq 0} \geq \langle v_+, \tilde{L}_G v_+ \rangle + \langle v_-, \tilde{L}_G v_- \rangle.$$

The last inequality holds because all off diagonal entries in $\tilde{L}_G$ are non-positive; $v_+(i)v_-(j) \leq 0$ for all $i \neq j$; and $v_+(i)v_-(i) = 0$. We have

$$\langle v_+, \tilde{L}_G v_+ \rangle + \langle v_-, \tilde{L}_G v_- \rangle \leq \langle v', \tilde{L}_G v' \rangle \leq \lambda_G \|D_w^{1/2}v'\|^2 = \lambda_G (\|D_w^{1/2}v_+\|^2 + \|D_w^{1/2}v_-\|^2).$$

Thus, for $u = v_+$ or $u = v_-$ the desired inequality holds. $\square$

Let u be as above. We assume without loss of generality that $\|u\|_\infty = \max_u |u(i)| = 1$ (if $\|u\|_\infty \neq 1$, we divide u by $\|u\|_\infty$). Next, we show that there exists an ε -buffered partition with small expansion by thresholding on this vector u .

Step 2. Random Thresholding with Buffers. Pick a random threshold $t \in [0, 1]$ uniformly distributed in $[0, 1]$ and define sets S , T , and buffer B as follows:

$$(2.13) \quad S = \{i : u(i)^2 > t\}$$

$$(2.14) \quad T = \{i : u(i)^2 \leq t/(1+\varepsilon)\},$$

$$(2.15) \quad B = V \setminus (S \cup T) = \{i : t/(1+\varepsilon) < u(i)^2 \leq t\}.$$

Note that $B \cup S = \{i : u(i)^2 > t/(1+\varepsilon)\}$. Since t is picked uniformly from $[0, 1]$ and $\|u\|_\infty = 1$, we have

$$\mathbf{E}[w(S)] = \sum_{i=1}^n w_i \Pr\{i \in S\} = \sum_{i=1}^n w_i \cdot u(i)^2 = \|D_w^{1/2} u\|^2,$$

and

$$(2.16) \quad \mathbf{E}[w(B \cup S)] = \sum_{i=1}^n w_i \cdot \min((1+\varepsilon)|u(i)|^2, 1) \leq (1+\varepsilon)\|D_w^{1/2} u\|^2.$$

Thus, $\mathbf{E}[w(B)] \leq \varepsilon\|D_w^{1/2} u\|^2 = \varepsilon\mathbf{E}[w(S)]$, as stated in Equation (II).

By our choice of z , the weight of vertices with positive values in u is at most $w(V)/2$. Since S contains a subset of vertices with positive values in u , we have $w(S) \leq w(V)/2$.

Note that for every edge (i, j) from S to T , we have $u(i)^2 > t > t/(1+\varepsilon) \geq u(j)^2$. Thus, for all edges $(i, j) \in \delta(S, T)$, we have: (a) $i \in S, j \in T$ if $u(i)^2 > u(j)^2$ and (b) $i \in T, j \in S$ if $u(i)^2 < u(j)^2$. Now consider an edge $(i, j) \in E$ with $u(i)^2 > u(j)^2$. The probability that $(i, j) \in \delta(S, T)$ equals

$$\begin{aligned} \Pr\{(i, j) \in \delta(S, T)\} &= \Pr\{i \in S; j \in T\} = \Pr\{t \leq u(i)^2 \text{ \& } t \geq (1+\varepsilon)u(j)^2\} \\ &= \max\{u(i)^2 - (1+\varepsilon)u(j)^2, 0\}. \end{aligned}$$

To bound the right side, we use the following simple claim.

CLAIM 2.2. *For all $\varepsilon > 0$ and all real numbers a and b , we have*

$$a^2 - (1+\varepsilon)b^2 \leq (1+1/\varepsilon)(a-b)^2.$$

Proof. If $b = 0$, then the inequality holds. Assume, that $b \neq 0$. Divide both sides by b^2 and denote $\lambda = a/b$. We need to show that $(1+1/\varepsilon)(\lambda-1)^2 - (\lambda^2 - (1+\varepsilon)) \geq 0$. Write,

$$\begin{aligned} (1+1/\varepsilon)(\lambda-1)^2 - (\lambda^2 - (1+\varepsilon)) &= 1/\varepsilon\lambda^2 - 2(1+1/\varepsilon)\lambda + (\sqrt{\varepsilon} + 1/\sqrt{\varepsilon})^2 \\ &= (\lambda/\sqrt{\varepsilon} - (\sqrt{\varepsilon} + 1/\sqrt{\varepsilon}))^2 \geq 0. \end{aligned}$$

□

Hence from the above Claim 2.2, we have

$$\Pr\{i \in S; j \in T\} \leq (1+1/\varepsilon)(u(i) - u(j))^2.$$

By linearity of expectation,

$$\begin{aligned} \mathbf{E}|\delta(S, T)| &\leq (1+1/\varepsilon) \sum_{\substack{(i,j) \in E \\ u(i)^2 > u(j)^2}} c_{ij}(u(i) - u(j))^2 \stackrel{\text{by (1.7)}}{=} (1+1/\varepsilon)\langle u, \tilde{L}_G u \rangle \leq \\ &\leq (1+1/\varepsilon)\lambda_G \|D_w^{1/2} u\|^2 \leq (1+1/\varepsilon)\lambda_G \cdot \mathbf{E}[w(S)]. \end{aligned}$$

We bounded $\langle u, \tilde{L}_G u \rangle$ using Claim 2.1 (cf. Equation (2.12)). Thus, this distribution over buffered partitions $(S, T \parallel B)$ satisfies Equation (I). Since (I) and (II) both hold, we can use Lemma 2.1 (see below) to conclude that there exists a cut $(\hat{S}, \hat{T})$ with buffer $\hat{B}$ for which

$$|\delta(\hat{S}, \hat{T})| \leq 2(1+1/\varepsilon)\lambda_G \cdot w(\hat{S}), \text{ and } w(\hat{B}) \leq 2\varepsilon \cdot w(\hat{S}).$$

For this cut $(\hat{S}, \hat{T})$ with buffer $\hat{B}$, we have

$$\frac{|\delta(\hat{S}, \hat{T})|}{w(\hat{S})} \leq \frac{2(1 + 1/\varepsilon)\lambda_G \cdot w(\hat{S})}{w(\hat{S})} = 2(1 + 1/\varepsilon)\lambda_G.$$

By (2.13) and (2.14), we have $\hat{S} \subseteq \{i : u(i)^2 > 0\}$ and $\hat{T} \supseteq \{i : u(i)^2 = 0\}$. Thus $w(\hat{T}) \leq w(\{i : u(i)^2 > 0\}) \leq w(V)/2$ and $w(\hat{T}) \geq w(\{i : u(i)^2 = 0\}) = w(V) - w(\{i : u(i)^2 > 0\}) \geq w(V)/2$. Therefore, $w(\hat{T}) \leq w(\hat{S})$. We conclude that

$$\frac{|\delta(\hat{S}, \hat{T})|}{w(\hat{T})} \leq \frac{|\delta(\hat{S}, \hat{T})|}{w(\hat{S})} \leq 2(1 + 1/\varepsilon)\lambda_G.$$

We obtain the desired result for $\varepsilon' = 2\varepsilon$. To finish the proof, it remains to show Lemma 2.1.

□

LEMMA 2.1. *For any $m \geq 2$, consider m arbitrary jointly distributed non-negative random variables $X_1, \dots, X_{m-1}$ and Y . Suppose that for every $i = 1, \dots, m-1$, $\mathbf{E}[X_i] \leq \alpha_i \mathbf{E}[Z]$. Then,*

$$(2.17) \quad \Pr\{X_i \leq (m-1)\alpha_i Y, \quad \forall i \in [m-1]\} > 0.$$

Proof. Consider a new random variable $Z = \sum_{i=1}^{m-1} \frac{X_i}{(m-1)\alpha_i}$. By the linearity of expectation, we have

$$\mathbf{E}[Y] \geq \frac{1}{m-1} \sum_{i=1}^{m-1} \frac{\mathbf{E}[X_i]}{\alpha_i} = \mathbf{E}[Z].$$

This implies that $\Pr\{Y \geq Z\} > 0$; otherwise we would have $\mathbf{E}[Y] < \mathbf{E}[Z]$. If $Y \geq Z$, then we have for every $i = 1, \dots, m-1$, $X_i \leq (m-1)\alpha_i Y$. Therefore, inequality (2.17) holds. □

3 Orthogonal Separators with Buffers

In this section, we introduce orthogonal separators with buffers. We will prove Theorems 3.1, 3.2, and 3.3 in Section 7. In these theorems, we provide randomized procedures to generate orthogonal separators with buffers in a set of unit vectors U in $\mathbb{R}^d$. In the next section, we will use the procedure in Theorem 3.3 to create a partial partitioning. We first use spectral embedding to map each vertex $u \in V$ to a vector $\bar{u} \in \mathbb{R}^k$. We will run this procedure on normalized vectors $\psi(\bar{u}) = \bar{u}/\|\bar{u}\|$ for all vertices $u \in V$. We first give the definition of the orthogonal separator with one buffer.

DEFINITION 3.1. *Consider a finite set U of unit vectors in $\mathbb{R}^d$. A distribution over two disjoint subsets of U is an m -orthogonal separator with an ε -buffer, distortion $\mathcal{D}$, separation radius R , and probability scale α if the following conditions hold for two subsets $X, Y \subseteq U$ chosen according to this distribution:*

1. *For all $\bar{u} \in U$, $\Pr\{\bar{u} \in X\} = \alpha$.*
2. *For all $\bar{u} \in U$, $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$.*
3. *For all $\bar{u}, \bar{v} \in U$ with $\|\bar{u} - \bar{v}\| \geq R$, $\Pr\{\bar{v} \in X \mid \bar{u} \in X\} \leq \frac{1}{m}$.*
4. *For all $\bar{u}, \bar{v} \in U$, $\Pr\{\bar{v} \notin X \cup Y \mid \bar{u} \in X\} \leq \mathcal{D}\|\bar{u} - \bar{v}\|^2$.*

We call X an orthogonal separator and Y its buffer.

In this definition, conditions 1 and 2 restrict the size of an orthogonal separator and its buffer respectively. Condition 3 requires that for every pair of vectors $\bar{u}, \bar{v} \in U$, if $\bar{u}, \bar{v}$ are almost orthogonal, then vectors $\bar{u}, \bar{v}$ are separated by X with high probability. Condition 4 upper bounds the probability that vectors $\bar{u}, \bar{v}$ are separated by the orthogonal separator X with a buffer Y . In the following theorem, we show there exists such an orthogonal separator with one buffer. The construction of the orthogonal separator with one buffer and its proof is in Section 7.

THEOREM 3.1. *There exists a randomized polynomial-time procedure that given a finite set U of unit vectors in $\mathbb{R}^d$ and positive parameters $\varepsilon \in (0, 1)$, $m \geq 3$, $R \in (0, 2)$, returns an m -orthogonal separator with an ε -buffer with distortion $\mathcal{D} = O_R(1/\varepsilon \log m)$, separation radius R , and probability scale $\alpha \geq O_R(1/\text{poly}(m))$.*

In the above theorem, we show that if vectors $\bar{u}$ and $\bar{v}$ are far apart, then they are both contained in X with a small probability. Suppose that every point $\bar{u}$ has a certain weight or measure $\mu(\bar{u})$. We now show that by slightly altering the distribution of X and Y , we can guarantee that the measure of every X is not much larger than the measure of the heaviest ball of radius R (see item 3 below for details).

DEFINITION 3.2. *Consider a finite set U of unit vectors in $\mathbb{R}^d$ equipped with a measure μ . A distribution over two disjoint subsets of U is an δ -orthogonal separator with an ε -buffer, distortion $\mathcal{D}$, separation radius R , and probability scale α if the following conditions hold for two subsets $X, Y \subseteq U$ chosen according to this distribution:*

1. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in X\} = \alpha$.
2. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$.
3. $\min_{\bar{u} \in X} \mu(X \setminus \text{Ball}(\bar{u}, R)) \leq \delta\mu(U)$ (always).
4. For all $\bar{u}, \bar{v} \in U$, $\Pr\{\bar{v} \notin X \cup Y \mid \bar{u} \in X\} \leq \mathcal{D} \|\bar{u} - \bar{v}\|^2$.

THEOREM 3.2. *There exists a randomized procedure that given a finite set U of unit vectors in $\mathbb{R}^d$ equipped with a measure μ and positive parameters $\varepsilon \in (0, 1)$, $\delta \leq 2/3$, $R \in (0, 2)$, returns an δ -orthogonal separator with an ε -buffer with distortion $\mathcal{D} = O_R(1/\varepsilon \log^{1/\delta})$, separation radius R , and probability scale $\alpha \geq O_R(1/\text{poly}(m))$.*

By using the orthogonal separator with one buffer above, we can find a buffered partitioning of the graph with buffered expansion in Theorem 1.1, but buffers B_i may overlap. To get disjoint buffers as in Theorem 1.1, we use the orthogonal separator with two buffers defined as follows.

DEFINITION 3.3. *Consider a finite set U of unit vectors in $\mathbb{R}^d$ equipped with a measure μ . A distribution over three disjoint subsets of U is an δ -orthogonal separator with two ε -buffers, distortion $\mathcal{D}$, separation radius R , and probability scale α if the following conditions hold for three disjoint subsets $X, Y, Z \subseteq U$ chosen according to this distribution:*

1. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in X\} = \alpha$.
2. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$ and $\Pr\{\bar{u} \in Z\} \leq \varepsilon\alpha$.
3. $\min_{\bar{u} \in X} \mu(X \setminus \text{Ball}(\bar{u}, R)) \leq \delta\mu(U)$ (always).
4. For all $\bar{u}, \bar{v} \in U$, $\Pr\{\bar{v} \notin X \cup Y \mid \bar{u} \in X\} \leq \mathcal{D} \|\bar{u} - \bar{v}\|^2$, and $\Pr\{\bar{v} \notin X \cup Y \cup Z \mid \bar{u} \in X \cup Y\} \leq \mathcal{D} \|\bar{u} - \bar{v}\|^2$.

In the following theorem, we slightly modify the procedure above to get orthogonal separators with two buffers.

THEOREM 3.3. *There exists a randomized procedure that given a finite set U of unit vectors in $\mathbb{R}^d$ equipped with a measure μ and positive parameters $\varepsilon \in (0, 1)$, $\delta \leq 2/3$, $R \in (0, 2)$, returns an δ -orthogonal separator with two ε -buffers with distortion $\mathcal{D} = O_R(1/\varepsilon \log^{1/\delta})$, separation radius R , and probability scale $\alpha \geq O_R(1/\text{poly}(m))$.*

4 Partial Partitioning Algorithm

In this section, we give an algorithm for finding a partial ε -buffered partitioning $(P_1, B_1), \dots, (P_{k'}, B_{k'})$ of G . This partitioning satisfies all the properties of the partitioning from Theorem 1.1 except the union of sets P_i does not necessarily cover the entire vertex set of G . For notational convenience, we will use k to denote the index of the eigenvalue that we compare the cost to. Eventually this theorem will be applied with $k = (1 + O(\delta))\hat{k}$, where $\hat{k}$ is the desired number of clusters (which we denoted by k in Theorem 1.1). We obtain this partial partitioning using the Partial Partitioning Algorithm which consists of Steps 1, 2, 3, and 4 provided in Figures 1, 2, 3, and 4.

The Partial Partitioning Algorithm generates this partial partitioning $(P_1, B_1), \dots, (P_{k'}, B_{k'})$ with $k' \geq (1-2\delta)k$ and partitions the uncovered vertices $V \setminus \bigcup_{i \in [k']} P_i \cup B_i$ into disjoint subsets A'_i, A''_i for $i \in [k']$ and R'_B, R'_P . We prove that these subsets P_i, B_i, A'_i, A''_i for $i \in [k']$ and R'_B, R'_P satisfy six properties given in Theorem 4.1 (see below). The first three properties show subsets P_i, B_i forms a partial ε -buffered partitioning. In Section 5, we show how to transform this partial partitioning with k' clusters into a true buffered partitioning of G with $\hat{k}$ clusters. We combine those additional sets A'_i, A''_i, R'_P, R'_B to get a true buffered partitioning. The properties 4, 5, and 6 in Theorem 4.1 are used in Section 5.

Find a spectral embedding for graph G :

- Let L_G be the normalized Laplacian matrix for G .
- Find the top k eigenvalues of L_G and corresponding orthogonal unit eigenvectors $x_1, \dots, x_k \in \mathbb{R}^V$. Denote coordinate $u \in V$ of x_i by $x_i(u)$.
- Embed each vertex $u \in V$ into k -dimensional vector $\bar{u}$ defined as follows: the i -th coordinate of $\bar{u}$ is $x_i(u)$.

Figure 1: Step 1 of the Partial Partitioning Algorithm. At this step, the algorithm maps vertices of G into vectors using the standard spectral embedding.

Let $R = \sqrt{\delta/6}$, $\delta' = \delta/2k$, and $T = 2/\alpha \ln 1/\delta$.

Set $\Sigma_0 = \emptyset$ and $\Gamma_0 = \emptyset$.

For $t = 1, \dots, T$:

- Sample an orthogonal separator X_t with buffers Y_t, Z_t using Theorem 3.3 with parameters ε , R , and δ' . For convenience, we assume that X_t, Y_t , and Z_t contain not vectors but the corresponding vertices of G .
- Let $\tilde{P}_t = X_t \setminus (\bigcup_{i < t} X_i \cup Y_i \cup Z_i)$ and $\Sigma_t = \Sigma_{t-1} \cup \tilde{P}_t$.
- Let $\tilde{B}_t = (X_t \cup Y_t) \setminus (\Sigma_t \cup \Gamma_{t-1})$ and $\Gamma_t = \Gamma_{t-1} \cup \tilde{B}_t$.
- Let $R_P = V \setminus (\bigcup_{t=1}^T X_t \cup Y_t \cup Z_t)$ and $R_B = V \setminus (\Sigma_T \cup \Gamma_T \cup R_P)$.

Figure 2: Step 2 of the Partial Partitioning Algorithm. At this step, the algorithm finds a *crude* partial partitioning $\{(P_t, \tilde{B}_t)\}_t$ of V .

Let $R'_P = R_P$ and $R'_B = R_B$.

For $t = 1, \dots, T$:

- Find r_t that minimizes $\phi_G(P_t \parallel B_t)$ subject to the constraints $|B_t| \leq C'_{4.1}(\delta)\varepsilon|P_t|$, $|A''_t| \leq 10\varepsilon|P_t|$, $\delta(A'_t, P_t \cup B_t) \leq C''_{4.1}(\delta)/\varepsilon \cdot \lambda_k \log k \cdot d|P_t|$, and $\delta_G(P_t \cup B_t, (\Sigma_T \cup R_P) \setminus \tilde{P}_t) \leq C''_{4.1}(\delta)/\varepsilon \cdot \lambda_k \log k \cdot d|P_t|$ where

$$P_t = \{u \in \tilde{P}_t : \|\bar{u}\|^2 \geq r_t\}$$

$$B_t = \{u \in \tilde{B}_t : \|\bar{u}\|^2 \geq r_t/(1+\varepsilon)\} \cup \{u \in \tilde{P}_t : \|\bar{u}\|^2 \in [r_t/(1+\varepsilon), r_t]\}$$

$$A'_t = \{u \in \tilde{P}_t : \|\bar{u}\|^2 \leq r_t/(1+\varepsilon)^2\}$$

$$A''_t = \{u \in \tilde{P}_t : \|\bar{u}\|^2 \in (r_t/(1+\varepsilon)^2, r_t/(1+\varepsilon))\}$$

Note that it suffices to consider r in $\{\|\bar{u}\|^2 : u \in \tilde{P}_t \cup \tilde{B}_t\}$. If no such r_t exists, we let $P_t = \emptyset$, $B_t = \emptyset$, $A'_t = \emptyset$, and $A''_t = \emptyset$.

- If no such r_t exists, then add $\tilde{P}_t$ to R'_P and add $\tilde{B}_t$ to R'_B . Otherwise, add $\tilde{B}_T \setminus B_t$ to R'_B .

Figure 3: Step 3 of the Partial Partitioning Algorithm. At this step, the algorithm refines the *crude* partial partitioning $\{(\tilde{P}_t, \tilde{B}_t)\}_t$ of V and obtains sets $\{(P_t, B_t, A'_t, A''_t)\}_t$.

For $t = 1, \dots, T$:

- Discard all sets P_t, B_t, A'_t, A''_t if $P_t = \emptyset$, or

$$\phi_G(P_t \parallel B_t) > \frac{C''_{4.1}(\delta)}{\varepsilon} \lambda_k \log k,$$

where $C''_{4.1}(\delta)$ is some function that depends only on δ (see Theorem 4.1).

- If sets P_t, B_t, A'_t, A''_t are discarded, then add $\tilde{P}_t$ to R'_P and add $\tilde{B}_t$ to R'_B .

Figure 4: Step 4 of the Partial Partitioning Algorithm. At this step, the algorithm discards all sets (P_t, B_t) that do not satisfy the conditions of Theorem 4.1.

THEOREM 4.1. *The Partial Partitioning Algorithm is a polynomial-time randomized algorithm that given a d -regular graph $G = (V, E)$, natural $k > 1$, and positive parameters $\varepsilon, \delta \in (0, 1/80)$, finds subsets R'_P, R'_B and P_i, B_i, A'_i, A''_i of V for $i \in [k']$ with $k' \geq (1 - 2\delta)k$ such that*

1. All sets P_i, B_i, A'_i, A''_i and R'_P, R'_B are disjoint and all sets P_i are nonempty, and

$$R'_P \cup R'_B \cup \bigcup_{i=1}^{k'} P_i \cup B_i \cup A'_i \cup A''_i = V;$$

2. $|B_i| \leq C'_{4.1}(\delta) \varepsilon |P_i|$ for all $i \in \{1, \dots, k'\}$; and

3. $\phi_G(P_i \parallel B_i) \leq \frac{C''_{4.1}(\delta)}{\varepsilon} \lambda_k \log k$, for all $i \in [k']$,

4. $|A''_i| \leq 10\varepsilon|P_i|$, for all $i \in [k']$;

5. $|R'_B| \leq 16\varepsilon n$;

$$6. \sum_{j=1}^{k'} \delta_G(A'_j, P_i \cup B_i) + \delta_G(R'_P, P_i \cup B_i) \leq \frac{2C''_{4.1}(\delta)}{\varepsilon} \lambda_k \log k \cdot d|P_i|, \text{ for all } i \in [k'].$$

Remark: We will assume that $\varepsilon \leq \delta$. If that is not the case, we can replace ε with $\varepsilon' = \delta$ and hide the additional factor of ε/ε' in the bound on $\phi_G(P_i \parallel B_i)$ and $\sum_{j=1}^{k'} \delta_G(A'_j, P_i \cup B_i) + \delta_G(R'_P, P_i \cup B_i)$ in the constant $C''_{4.1}(\delta)$. We will also assume that $\delta \geq 1/(3k)$: indeed if $\delta < 1/(3k)$, we can increase it to $1/(3k)$ and we will still have $k' \geq \lceil (1 - 2/(3k))k \rceil = k$, as for the original value of δ .

Proof. Our algorithm consists of four steps. First, we embed the vertex set V into a k dimensional space using the standard spectral embedding (see Section 6 for details). We denote the image of vertex u by $\bar{u}$. We also let $\psi(\bar{u}) = \bar{u}/\|\bar{u}\|$ (that is, $\psi(\bar{u})$ is the normalized $\bar{u}$) and $\mu(u) = \|\bar{u}\|^2$ (note: $\bar{u} \neq 0$ by Claim 6.1). At the second step, we obtain a *crude* partial partitioning $\tilde{P}_1, \dots, \tilde{P}_{k''}$ with buffers $\tilde{B}_1, \dots, \tilde{B}_{k''}$ using a new technical tool, which we introduced in Section 3. We call this tool *orthogonal separators with buffers* (see Theorem 3.3). Finally, we refine the crude partitioning at the third step and discard some sets at the fourth step. We get subsets P_i, B_i, A'_i, A''_i for $i \in [k']$ and two extra subsets R'_P, R'_B . We provide the pseudocode for Steps 1, 2, 3 and 4 in Figures 1, 2, 3, and 4. We now analyze our algorithm.

Before we proceed to the proof, we set some notation. Let $\text{Ball}(u, R)$ be the ball of radius R around u in the metric $\rho(u, v) = \|\psi(\bar{u}) - \psi(\bar{v})\|$:

$$\text{Ball}(u, R) = \{v \in V : \|\psi(\bar{u}) - \psi(\bar{v})\| \leq R\}.$$

We define measure μ on V as follows: for every $S \subseteq V$,

$$\mu(S) = \sum_{u \in S} \mu(u).$$

Step 1: Spectral Embedding. In Section 6, we remind the reader the standard definition of a spectral embedding of G into $\mathbb{R}^k$. We then prove two claims about this embedding. First, we note that $\mu(V) = k$. This is a known fact (see e.g., [LRTV12]). Then, in Lemma 6.1, we show that for $R < 1/\sqrt{2}$, for any vertex $u \in V$,

$$(4.18) \quad \mu(\text{Ball}(u, R)) \leq \frac{1}{1 - 2R^2}.$$

We will use this bound with $R = \sqrt{\delta/6}$.

Step 2: Crude Partial Partitioning. We now analyze the second step of the algorithm described in Figure 2. Let $\{(\tilde{P}_t, \tilde{B}_t)\}_{t=1}^T$ be the crude partial partitioning obtained at this step. Define function

$$(4.19) \quad \eta(u, v) = \begin{cases} \|\bar{u}\|^2, & \text{if } u \in \tilde{P}_t, v \notin \tilde{P}_t \cup \tilde{B}_t \text{ for some } t; \\ 1/\varepsilon \|\bar{u} - \bar{v}\|^2, & \text{if } u \in \tilde{P}_t, v \in \tilde{P}_t \cup \tilde{B}_t \text{ for some } t; \\ 0, & \text{otherwise.} \end{cases}$$

Later, we will use the following sum as an estimate of the size of the edge boundary of set P_t :

$$(4.20) \quad \eta(\tilde{P}_t) = \sum_{\substack{u \in \tilde{P}_t; v \in V; \\ s.t. (u, v) \in E}} \eta(u, v).$$

Note that function $\eta(u, v)$ is not symmetric. If u and v are in $\tilde{P}_t$, then the sum above includes both terms $\eta(u, v)$ and $\eta(v, u)$. Depending on the argument, we will use η to denote the cost of an edge as in Equation (4.19) or the cost of all edges incident on vertices in $\tilde{P}_t$ as in Equation (4.20).

Note that sets $\tilde{P}_t, \tilde{B}_t$ are contained in $X_t \cup Y_t \cup Z_t \setminus \Sigma_{t-1}$, where X_t, Y_t, Z_t are orthogonal separator and its two buffers and Σ_{t-1} are vertices covered by previous $\tilde{P}_i$ for $i < t$. We define another cost function as follows:

$$(4.21) \quad \tilde{\eta}(u, v) = \begin{cases} \|\bar{u}\|^2, & \text{if } u \in \tilde{P}_t \cup \tilde{B}_t, v \notin (X_t \cup Y_t \cup Z_t) \setminus \Sigma_{t-1} \text{ for some } i; \\ 0, & \text{otherwise.} \end{cases}$$

We will use this cost function to bound the total cost of edges from each part in the partial partitioning P_i and B_i to the uncovered part R'_B and R'_P . The cost of all edges incident on vertices in $\tilde{P}_t \cup \tilde{B}_t$ for function $\tilde{\eta}$ is denoted as

$$(4.22) \quad \tilde{\eta}(\tilde{P}_t \cup \tilde{B}_t) = \sum_{\substack{u \in \tilde{P}_t \cup \tilde{B}_t; v \in V; \\ s.t. (u,v) \in E}} \tilde{\eta}(u, v).$$

We prove the following lemma for all sets generated after Step 2.

LEMMA 4.1. *The crude partial partitioning $\{(\tilde{P}_t, \tilde{B}_t)\}_{t=1}^T$ and subsets R_B, R_P obtained at Step 2 of the algorithm satisfies the following properties:*

1. $\mu(\tilde{P}_t) \leq 1 + \delta$ for all t ;
2. $\frac{1}{k} \sum_{t=1}^T \mathbf{E}[\mu(\tilde{P}_t)] \geq 1 - 5\delta$;
3. $\frac{1}{k} \sum_{t=1}^T \mathbf{E}[\mu(\tilde{B}_t)] \leq 4\epsilon$;
4. $\frac{1}{k} \sum_{t=1}^T \mathbf{E}[\eta(\tilde{P}_t)] \leq \frac{C_\delta}{\epsilon} \cdot \lambda_k d \log k$;
5. $\sum_{t=1}^T \mathbf{E}|\tilde{B}_t| + \mathbf{E}|R_B| \leq 4\epsilon n$;
6. $\frac{1}{k} \sum_{t=1}^T \mathbf{E}[\tilde{\eta}(\tilde{P}_t \cup \tilde{B}_t)] \leq \frac{C_\delta}{\epsilon} \cdot \lambda_k d \log k$.

Here, the expectation is taken over the random decisions made by the algorithm at Step 2 (all other steps of the algorithm are deterministic).

Proof. We will use Theorem 3.3 to analyze Step 2 of the algorithm. We first show item (1). Observe that $\tilde{P}_t \subset X_t$ and for every $u \in X_t$, $X_t = \text{Ball}(u, R) \cup (X_t \setminus \text{Ball}(u, R))$. Thus,

$$\mu(\tilde{P}_t) \leq \mu(\text{Ball}(u, R)) + \mu(X_t \setminus \text{Ball}(u, R)).$$

By Lemma 6.1 (see Equation (4.18)), $\mu(\text{Ball}(u, R)) \leq 1/(1 - \delta/3) \leq 1 + \delta/2$ for all u . By Theorem 3.3,

$$\min_{u \in X_t} \mu(X_t \setminus \text{Ball}(u, R)) \leq \frac{\delta \mu(V)}{2k} = \frac{\delta}{2}.$$

Thus, $\mu(\tilde{P}_t) \leq 1 + \delta$.

We now prove item (2). Consider a vertex u . Observe that if u gets assigned to set Σ_t at iteration t , then it remains in the set $\Sigma_{t'}$ in the future iterations $t' > t$. That is, $\Sigma_t \subset \Sigma_{t+1}$. Let $\Xi_t = \bigcup_{i < t} X_i \cup Y_i \cup Z_i$. Then, similarly, we have $\Xi_t \subset \Xi_{t+1}$. If u is not in Ξ_t , then at step $(t+1)$, it is assigned to $\tilde{P}_{t+1}$ with probability at least $\alpha/2$ and to $\Xi_{t+1} \setminus \tilde{P}_{t+1}$ with probability at most $2\epsilon\alpha$ (see Theorem 3.3). Thus,

$$\Pr\{u \in \Sigma_t \mid u \in \Xi_t\} \geq \frac{\alpha/2}{\alpha/2 + 2\epsilon\alpha} = \frac{1}{1 + 4\epsilon}.$$

Also,

$$1 - (1 - \alpha(1 + 2\epsilon))^t \geq \Pr\{u \in \Xi_t\} \geq 1 - (1 - \alpha/2)^t.$$

Therefore (since $T = \lceil 2/\alpha \ln 1/\delta \rceil$ and $\epsilon < \delta < 1/48$),

$$(4.23) \quad \Pr\{u \in \Sigma_T\} \geq \frac{1 - (1 - \alpha/2)^T}{1 + 4\epsilon} \geq \frac{1 - \delta}{1 + 4\delta} \geq 1 - 5\delta.$$

Item (2) follows from the bound above because sets $\tilde{P}_t$ are disjoint and $\Sigma_T = \bigcup_{t=1}^T \tilde{P}_t$.

We then prove items (3) and (5). Note that the remaining parts $R_P = V \setminus \Xi_T$ and $R_B = V \setminus (R_P \cup \Sigma_T \cup \Gamma_T) = \Xi_T \setminus (\Sigma_T \cup \Gamma_T)$. Since all sets $\tilde{B}_t$ are disjoint and $\Gamma_T = \bigcup_{t=1}^T \tilde{B}_t$, we upper bound probabilities $\Pr\{u \in \Gamma_T\}$ and $\Pr\{u \in \Gamma_T \cup R_B\}$. Since $R_B = \Xi_T \setminus (\Sigma_T \cup \Gamma_T)$, we have $\Gamma_T \cup R_B = \Xi_T \setminus \Sigma_T$. Similar to bound (4.23), we have

$$(4.24) \quad \Pr\{u \in \Gamma_T\} \leq \Pr\{u \in \Gamma_T \cup R_B\} \leq \Pr\{u \in \Xi_T \setminus \Sigma_T\} \leq \frac{4\varepsilon}{1+4\varepsilon} \cdot (1 - (1 - \alpha(1+2\varepsilon))^T) \leq 4\varepsilon,$$

where the last inequality is due to $\Pr\{u \in \Xi_T \setminus \Sigma_T \mid u \in \Xi_T\} \leq 4\varepsilon/(1+4\varepsilon)$ and $\Pr\{u \in \Xi_T\} \leq 1 - (1 - \alpha(1+2\varepsilon))^T$. Then, item (3) follows from $\Pr\{u \in \Gamma_T\} \leq 4\varepsilon$ and item (5) follows from $\Pr\{u \in \Gamma_T \cup R_B\} \leq 4\varepsilon$.

We now prove the item (4). Consider an edge (u, v) . We bound the probability of the event $\{\eta(u, v) = \|\bar{u}\|^2\}$. If $\eta(u, v) = \|\bar{u}\|^2$, then $u \in \tilde{P}_t$, and $v \notin \tilde{P}_t \cup \tilde{B}_t$ for some t . We first assume that $v \notin \Sigma_{t'} \cup \Gamma_{t'}$ with $t' \leq t-1$ or, in other words, $v \notin \Sigma_{t-1} \cup \Gamma_{t-1}$. Then, $u \in X_t \setminus \Xi_{t-1}$ and $v \notin X_t \cup Y_t$ for some t (otherwise, if v was in $(X_t \cup Y_t) \setminus \Sigma_{t-1} \cup \Gamma_{t-1}$, v would also be in $\tilde{P}_t$ or $\tilde{B}_t$). If $v \in \tilde{P}_{t'} \cup \tilde{B}_{t'}$ and $u \in \tilde{P}_t$ with $t' < t$, then $v \in (X_{t'} \cup Y_{t'}) \setminus (\Sigma_{t'-1} \cup \Gamma_{t'-1})$ and $u \notin X_{t'} \cup Y_{t'} \cup Z_{t'}$ for some t' . Write,

$$(4.25) \quad \Pr\{\eta(u, v) = \|\bar{u}\|^2\} \leq \underbrace{\sum_{t=1}^T \Pr\{u \in X_t \setminus \Xi_{t-1} \text{ and } v \notin X_t \cup Y_t\}}_{(*)}$$

$$(4.26) \quad + \underbrace{\sum_{t=1}^T \Pr\{v \in (X_t \cup Y_t) \setminus (\Sigma_{t-1} \cup \Gamma_{t-1}) \text{ and } u \notin X_t \cup Y_t \cup Z_t\}}_{(**)}.$$

We upper bound the first term. Two events $\{u \in X_t; v \notin X_t \cup Y_t\}$ and $\{u \notin \Xi_{t-1}\}$ are independent for every t . Thus,

$$\begin{aligned} (*) &\leq \sum_{t=1}^T \Pr\{u \in X_t \text{ and } v \notin X_t \cup Y_t\} \cdot \Pr\{u \notin \Xi_{t-1}\} \\ &= \sum_{t=1}^T \Pr\{v \notin X_t \cup Y_t \mid u \in X_t\} \cdot \Pr\{u \in X_t\} \cdot \Pr\{u \notin \Xi_{t-1}\} \\ &= \sum_{t=1}^T \Pr\{v \notin X_t \cup Y_t \mid u \in X_t\} \cdot \Pr\{u \in X_t \setminus \Xi_{t-1}\}. \end{aligned}$$

By Theorem 3.3,

$$\Pr\{v \notin X_t \cup Y_t \mid u \in X_t\} \leq \mathcal{D} \|\psi(\bar{u}) - \psi(\bar{v})\|^2,$$

where $\mathcal{D} = O(1/\varepsilon \log k/\delta) = O_\delta(1/\varepsilon \log k)$. Observe that events $\{u \in X_t \setminus \Xi_{t-1}\}$ for $t \in \{1, \dots, T\}$ are mutually exclusive. Thus,

$$(*) \leq \mathcal{D} \|\psi(\bar{u}) - \psi(\bar{v})\|^2 \cdot \underbrace{\sum_{t=1}^T \Pr\{u \in X_t \setminus \Xi_{t-1}\}}_{\leq 1} \leq \mathcal{D} \|\psi(\bar{u}) - \psi(\bar{v})\|^2.$$

The same bound holds for $(**)$ in Equation (4.26). We now bound $\mathbf{E}[\eta(u, v)]$:

$$\begin{aligned} \mathbf{E}[\eta(u, v)] &= \Pr\{\eta(u, v) = \|\bar{u}\|^2\} \cdot \|\bar{u}\|^2 + \Pr\{\eta(u, v) = 1/\varepsilon \|\bar{u} - \bar{v}\|^2\} \cdot 1/\varepsilon \|\bar{u} - \bar{v}\|^2 \\ &\leq 2\mathcal{D} \|\psi(\bar{u}) - \psi(\bar{v})\|^2 \cdot \|\bar{u}\|^2 + 1/\varepsilon \|\bar{u} - \bar{v}\|^2. \end{aligned}$$

By Claim 4.1 (see below), $\mathbf{E}[\eta(u, v)] \leq 8\mathcal{D} \|\bar{u} - \bar{v}\|^2 + 1/\varepsilon \|\bar{u} - \bar{v}\|^2 = O_\delta(1/\varepsilon \log k) \|\bar{u} - \bar{v}\|^2$.

CLAIM 4.1. Consider two vertices $u, v \in V$ and the corresponding nonzero vectors $\bar{u}, \bar{v}$. We have

$$\|\bar{u}\|^2 \cdot \|\psi(\bar{u}) - \psi(\bar{v})\|^2 \leq 4\|\bar{u} - \bar{v}\|^2.$$

Remark: This is a known inequality. See e.g., [CMM06] and [LGT14].

Proof. Write,

$$\|\bar{u}\|^2 \cdot \|\psi(\bar{u}) - \psi(\bar{v})\|^2 = \|\bar{u}\|^2 \cdot \left\| \frac{\bar{u}}{\|\bar{u}\|} - \frac{\bar{v}}{\|\bar{v}\|} \right\|^2 = \left\| \bar{u} - \frac{\|\bar{u}\|}{\|\bar{v}\|} \bar{v} \right\|^2.$$

We now use the *relaxed triangle inequality* for squared Euclidean distance $\|x - z\|^2 \leq 2\|x - y\|^2 + 2\|y - z\|^2$. We have

$$\|\bar{u}\|^2 \cdot \|\psi(\bar{u}) - \psi(\bar{v})\|^2 \leq 2\|\bar{u} - \bar{v}\|^2 + 2\left\| \bar{v} - \frac{\|\bar{u}\|}{\|\bar{v}\|} \bar{v} \right\|^2 \leq 4\|\bar{u} - \bar{v}\|^2.$$

Here, we used that $\bar{v}$ and $\frac{\|\bar{u}\|}{\|\bar{v}\|} \bar{v}$ are collinear vectors and, thus,

$$\left\| \frac{\|\bar{u}\|}{\|\bar{v}\|} \bar{v} - \bar{v} \right\| = \left| \left\| \frac{\|\bar{u}\|}{\|\bar{v}\|} \bar{v} \right\| - \|\bar{v}\| \right| = \left| \|\bar{u}\| - \|\bar{v}\| \right| \leq \|\bar{u} - \bar{v}\|.$$

□

We can now finish the proof of Lemma 4.1,

$$\frac{1}{k} \sum_{t=1}^T \mathbf{E}[\eta(\tilde{P}_t)] = \frac{1}{k} \sum_{(u,v) \in E} \mathbf{E}[\eta(u,v)] + \mathbf{E}[\eta(v,u)] = O_\delta(1/\varepsilon \log k) \frac{1}{k} \sum_{(u,v) \in E} \|\bar{u} - \bar{v}\|^2.$$

By Claim 6.2, the right hand side is upper bounded by $O_\delta(1/\varepsilon \log k) d\lambda_k$.

Finally, we prove item (6). Similar to the analysis of item (4), for any edge (u, v) , we bound the probability that $\tilde{\eta}(u, v) = \|\bar{u}\|^2$. If $\tilde{\eta}(u, v) = \|\bar{u}\|^2$, then we have $u \in \tilde{P}_t \cup \tilde{B}_t$ and $v \notin (X_t \cup Y_t \cup Z_t) \setminus \Sigma_{t-1}$ for some t . We also first assume that when u is contained in $\tilde{P}_t \cup \tilde{B}_t$, vertex v is not contained in Σ_{t-1} . Then, we must have $v \notin X_t \cup Y_t \cup Z_t$. If v is covered by $\tilde{P}_t$ for some t before u is covered, then we must have $u \notin X_t \cup Y_t$ (otherwise u is contained in $\tilde{P}_t \cup \tilde{B}_t$). Thus, we have

$$\begin{aligned} \Pr\{\tilde{\eta}(u, v) = \|\bar{u}\|^2\} &\leq \sum_{t=1}^T \Pr\{u \in (X_t \cup Y_t) \setminus \Xi_{t-1} \text{ and } v \notin X_t \cup Y_t \cup Z_t\} \\ &\quad + \sum_{t=1}^T \Pr\{v \in X_t \setminus \Xi_{t-1} \text{ and } u \notin X_t \cup Y_t\}. \end{aligned}$$

By Theorem 3.3, we have $\Pr\{\tilde{\eta}(u, v) = \|\bar{u}\|^2\} \leq 2\mathcal{D} \|\psi(\bar{u}) - \psi(\bar{v})\|^2$. By Claim 6.2, we get

$$\frac{1}{k} \sum_{t=1}^T \mathbf{E}[\tilde{\eta}(\tilde{P}_t \cup \tilde{B}_t)] = \frac{1}{k} \sum_{(u,v) \in E} \mathbf{E}[\tilde{\eta}(u, v)] + \mathbf{E}[\tilde{\eta}(v, u)] = O_\delta(1/\varepsilon \log k) d\lambda_k.$$

□

By item (5) in Lemma 4.1 and Markov's inequality, we have $|R_B| + \sum_{t=1}^T |\tilde{B}_t| \leq 16\varepsilon n$ holds with probability at least $3/4$. In the following analysis, we assume this always holds.

Steps 3 & 4. Our algorithm (the Partial Partitioning Algorithm) refines the crude partial partitioning $\{\tilde{P}_t, \tilde{B}_t\}_{t=1}^T$ at Step 3 and obtains a set of tuples $\{(P_t, B_t, A'_t, A''_t)\}_{t=1}^T$. Then, it removes some of the sets (P_t, B_t, A'_t, A''_t) from the partial partitioning at Step 4. In the analysis of the algorithm, it will be more convenient for us to identify those sets $(\tilde{P}_t, \tilde{B}_t)$ that remain in the solution first and only then find their refinements (P_t, B_t, A'_t, A''_t) . Let

$$(4.27) \quad \mathcal{I} = \left\{ i : \tilde{P}_i \neq \emptyset, \mu(\tilde{B}_i) \leq C'_\delta \varepsilon \mu(\tilde{P}_i), \text{ and } \max\{\eta(\tilde{P}_i), \tilde{\eta}(\tilde{P}_i \cup \tilde{B}_i)\} \leq C''_\delta / \varepsilon \cdot \lambda_k d \log k \mu(\tilde{P}_i) \right\},$$

where $C'_\delta = 192/\delta$ and $C''_\delta = 48C_\delta/\delta$. We will now prove that $\Pr\{|\mathcal{I}| \geq (1 - 2\delta)|k|\} \geq 1/2$. In the next section, we show that for each $i \in \mathcal{I}$, the tuple (P_i, B_i, A'_i, A''_i) satisfies all constraints at Step 3 and 4. Thus, all sets

(P_i, B_i, A'_i, A''_i) with $i \in \mathcal{I}$ remain in the solution after Step 4 and, consequently, the algorithm succeeds with probability at least $1/4$ (We assume $|R_B| + \sum_{t=1}^T |\tilde{B}_t| \leq 16\epsilon n$ at Step 2, which holds with probability at least $3/4$).

Lemma 4.1 gives us upper bounds on the expected values of $k - \sum_t \mu(\tilde{P}_t)$, $\sum_t \mu(\tilde{B}_t)$, $\sum_t \eta(\tilde{P}_t)$, and $\sum_t \tilde{\eta}(\tilde{P}_t \cup \tilde{B}_t)$. These four random variables are non-negative. Thus, by Markov's inequality, with probability at least $1/2$, the following four inequalities hold simultaneously:

$$\begin{aligned} \frac{1}{k} \sum_{t=1}^T \mu(\tilde{P}_t) &\geq 1 - 40\delta; \\ \frac{1}{k} \sum_{t=1}^T \mu(\tilde{B}_t) &\leq 32\epsilon; \\ \frac{1}{k} \sum_{t=1}^T \eta(\tilde{P}_t) &\leq 8C_\delta/\epsilon \lambda_k d \log k. \\ \frac{1}{k} \sum_{t=1}^T \tilde{\eta}(\tilde{P}_t \cup \tilde{B}_t) &\leq 8C_\delta/\epsilon \lambda_k d \log k. \end{aligned}$$

Denote the event that all above inequalities hold by $\mathcal{E}$. We know that $\Pr(\mathcal{E}) \geq 1/2$. Let us assume that $\mathcal{E}$ occurs. Since $\delta < 1/80$, we have

$$\begin{aligned} \sum_{t=1}^T \mu(\tilde{B}_t) &\leq 64\epsilon \sum_{t=1}^T \mu(\tilde{P}_t); \\ \sum_{t=1}^T \eta(\tilde{P}_t) &\leq 16C_\delta/\epsilon \lambda_k d \log k \sum_{t=1}^T \mu(\tilde{P}_t); \\ \sum_{t=1}^T \tilde{\eta}(\tilde{P}_t \cup \tilde{B}_t) &\leq 16C_\delta/\epsilon \lambda_k d \log k \sum_{t=1}^T \mu(\tilde{P}_t). \end{aligned}$$

Let $w_i = \mu(\tilde{P}_i) / \sum_{t=1}^T \mu(\tilde{P}_t)$. We rewrite the inequalities above as follows:

$$\begin{aligned} \sum_{i=1}^T w_i \frac{\mu(\tilde{B}_i)}{\mu(\tilde{P}_i)} &\leq 64\epsilon; \\ \sum_{i=1}^T w_i \frac{\eta(\tilde{P}_i)}{\mu(\tilde{P}_i)} &\leq 16C_\delta/\epsilon \lambda_k d \log k; \\ \sum_{i=1}^T w_i \frac{\tilde{\eta}(\tilde{P}_i \cup \tilde{B}_i)}{\mu(\tilde{P}_i)} &\leq 16C_\delta/\epsilon \lambda_k d \log k. \end{aligned}$$

In the expressions above, we ignore the terms with $w_i = 0$. Note that $\sum_i w_i = 1$. Suppose that we pick i in $\{1, \dots, T\}$ randomly with probability w_i . Then, the above inequalities give bounds on the expected values of $\mu(\tilde{B}_i)/\mu(\tilde{P}_i)$ and $\eta(\tilde{P}_i)/\mu(\tilde{P}_i)$. By Markov's inequality,

$$\Pr_{i \sim w} \{i \in \mathcal{I}\} = \Pr_{i \sim w} \left\{ \mu(\tilde{B}_i) \leq C'_\delta \epsilon \mu(\tilde{P}_i) \text{ and } \max\{\eta(\tilde{P}_i), \tilde{\eta}(\tilde{P}_i \cup \tilde{B}_i)\} \leq C''_\delta/\epsilon \lambda_k d \log k \mu(\tilde{P}_i) \right\} \geq 1 - \delta,$$

where $C'_\delta = 192/\delta$ and $C''_\delta = 48C_\delta/\delta$. Therefore, $\sum_{i \in \mathcal{I}} w_i \geq 1 - \delta$. We have

$$\sum_{i \in \mathcal{I}} \mu(\tilde{P}_i) \geq (1 - \delta) \sum_{i=1}^T \mu(\tilde{P}_i) \geq (1 - \delta)k.$$

We now recall that $\mu(\tilde{P}_i) \leq 1 + \delta$. Consequently,

$$|\mathcal{I}| \geq \frac{1-\delta}{1+\delta}k \geq (1-2\delta)k.$$

We just showed that if event $\mathcal{E}$ occurs, then $|\mathcal{I}| \geq (1-2\delta)k$ and $\Pr(\mathcal{E}) \geq 1/2$. Hence, $\Pr\{|\mathcal{I}| \geq (1-2\delta)k\} \geq 1/2$.

Step 3: Refined Partial Partitioning. At Step 3 of the algorithm, we refine the crude partitioning obtained at Step 2. To this end, we pick a threshold $r_i \in (0, 1)$ for every pair $(\tilde{P}_i, \tilde{B}_i)$ with $i \in \mathcal{I}$. We define the refined partitioning sets to be

- $P_i = \{u \in \tilde{P}_i : \mu(u) \geq r_i\},$
- $B_i = \{u \in \tilde{B}_i : \mu(u) \geq r_i/(1+\varepsilon)\} \cup \{u \in \tilde{P}_i : \mu(u) \in [r_i/(1+\varepsilon), r_i]\},$
- $A'_i = \{u \in \tilde{P}_i : \mu(u) \leq r_i/(1+\varepsilon)^2\},$
- $A''_i = \{u \in \tilde{P}_i : \mu(u) \in (r_i/(1+\varepsilon)^2, r_i/(1+\varepsilon))\}.$

The threshold r_i must satisfy five conditions: (1) $|B_i| \leq C'_{4.1}(\delta)\varepsilon|P_i|$; (2) $\phi_G(P_i \parallel B_i) \leq C''_{4.1}(\delta)/\varepsilon \lambda_k \log k$; (3) $|A''_i| \leq 10\varepsilon|P_i|$, and (4) $\delta_G(A'_i, P_i \cup B_i) \leq C''_{4.1}(\delta)/\varepsilon \lambda_k \log k \cdot d|P_i|$; (5) $\delta_G(P_i \cup B_i, (\Sigma_T \cup R_P) \setminus \tilde{P}_i) \leq C''_{4.1}(\delta)/\varepsilon \lambda_k \log k \cdot d|P_i|$. At Step 4, we drop sets (P_i, B_i, A'_i, A''_i) for which we could not find such threshold. We now show that for every $i \in \mathcal{I}$ such threshold r_i exists (set $\mathcal{I}$ is defined in Equation (4.27)). We use the probabilistic method.

LEMMA 4.2. *Consider $i \in \mathcal{I}$. Suppose, we select elements in sets P_i and B_i using a random threshold r_i , which is uniformly distributed in $(0, 1)$. Then*

1. $\mathbf{E}_{r_i} |B_i| \leq 2C'_\delta \varepsilon \mathbf{E}_{r_i} |P_i|;$
2. $\mathbf{E}_{r_i} \left[\delta_G(P_i, V \setminus (P_i \cup B_i)) \right] \leq \frac{C''_\delta}{\varepsilon} \lambda_k \log k \cdot d \mathbf{E}_{r_i} |P_i|;$
3. $\mathbf{E}_{r_i} |A''_i| \leq 2\varepsilon \mathbf{E}_{r_i} |P_i|;$
4. $\mathbf{E}_{r_i} \left[\delta_G(A'_i, P_i \cup B_i) \right] \leq \frac{C''_\delta}{\varepsilon} \lambda_k \log k \cdot d \mathbf{E}_{r_i} |P_i|.$
5. $\mathbf{E}_{r_i} \left[\delta_G(P_i \cup B_i, (\Sigma_T \cup R_P) \setminus \tilde{P}_i) \right] \leq \frac{C''_\delta}{\varepsilon} \lambda_k \log k \cdot d \mathbf{E}_{r_i} |P_i|.$

Proof. Denote

$$B'_i = \{u \in \tilde{B}_i : \mu(u) \geq r_i/(1+\varepsilon)\} \text{ and } B''_i = \{u \in \tilde{P}_i : \mu(u) \in [r_i/(1+\varepsilon), r_i]\}.$$

Then, $B_i = B'_i \cup B''_i$. Write,

$$\mathbf{E}_{r_i} |P_i| = \sum_{u \in \tilde{P}_i} \Pr_{r_i} \{u \in P_i\} = \sum_{u \in \tilde{P}_i} \Pr_{r_i} \{r_i \leq \mu(u)\} = \mu(\tilde{P}_i).$$

Here, we used that $\mu(u) \leq 1$ for all u (see Claim 6.1). Similarly, $\mathbf{E} |B'_i| \leq (1+\varepsilon)\mu(\tilde{B}_i)$. Then,

$$\mathbf{E} |B''_i| = \sum_{u \in \tilde{P}_i} \Pr_{r_i} \left\{ \mu(i) \in [r_i/(1+\varepsilon), r_i] \right\} = \sum_{u \in \tilde{P}_i} \Pr_{r_i} \left\{ r_i \in [\mu(i), (1+\varepsilon)\mu(i)] \right\} \leq \varepsilon \mu(\tilde{P}_i).$$

Thus, using the definition of set $\mathcal{I}$, we get

$$\mathbf{E} |B_i| \leq \mathbf{E} |B'_i| + \mathbf{E} |B''_i| = (1+\varepsilon)\mu(\tilde{B}_i) + \varepsilon \mu(\tilde{P}_i) \leq ((1+\varepsilon)C'_\delta + 1)\varepsilon \mu(\tilde{P}_i) = 2C'_\delta \varepsilon \mathbf{E} |P_i|.$$

This proves the first claim of Lemma 4.2.

We assign all vertices $u \in \tilde{P}_i$ with $\mu(u) \in (r_i/(1+\varepsilon)^2, r_i/(1+\varepsilon))$ to set A_i'' . Then, we have

$$\begin{aligned} \mathbf{E} |A_i''| &= \sum_{u \in \tilde{P}_i} \Pr_{r_i} \left\{ \mu(i) \in (r_i/(1+\varepsilon)^2, r_i/(1+\varepsilon)) \right\} = \\ &= \sum_{u \in \tilde{P}_i} \Pr_{r_i} \left\{ r_i \in [(1+\varepsilon)\mu(i), (1+\varepsilon)^2\mu(i)] \right\} = \sum_{u \in \tilde{P}_i} (\varepsilon + \varepsilon^2)\mu(i) < 2\varepsilon\mu(\tilde{P}_i). \end{aligned}$$

Since $\mu(\tilde{P}_i) = \mathbf{E}_{r_i} |P_i|$, we get the third claim.

To show claims 2 and 4 of Lemma 4.2, we bound the expected number of edges from set P_i to set $V \setminus (P_i \cup B_i)$, and the expected number of edges from set A_i' to set $P_i \cup B_i$.

CLAIM 4.2. Consider an edge $(u, v) \in E$ with $u \in \tilde{P}_i$. We have

$$\Pr\{u \in P_i; v \notin P_i \cup B_i\} \leq 2\eta(u, v),$$

and

$$\Pr\{u \in A_i'; v \in P_i \cup B_i\} \leq 2\eta(u, v).$$

Proof. Consider two cases. If $v \in \tilde{P}_i \cup \tilde{B}_i$, then

$$\begin{aligned} \Pr\{u \in P_i, v \notin P_i \cup B_i\} &= \Pr\{\mu(u) \geq r_i \text{ and } \mu(v) < r_i/(1+\varepsilon)\} \\ &\leq \Pr\{r_i \in [(1+\varepsilon)\mu(v), \mu(u)]\} \\ &\leq \mu(u) - (1+\varepsilon)\mu(v). \end{aligned}$$

By Claim 2.2,

$$\mu(u) - (1+\varepsilon)\mu(v) = \|\bar{u}\|^2 - (1+\varepsilon)\|\bar{v}\|^2 \leq (1+1/\varepsilon)(\|\bar{u}\| - \|\bar{v}\|)^2 \leq 2(\|\bar{u}\| - \|\bar{v}\|)^2/\varepsilon.$$

Using the triangle inequality $\|\bar{u}\| - \|\bar{v}\| \leq \|\bar{u} - \bar{v}\|$, we conclude that

$$\Pr\{u \in P_i, v \notin P_i \cup B_i\} \leq 2\eta(u, v).$$

Similarly, we have

$$\begin{aligned} \Pr\{u \in A_i' \text{ and } v \in P_i \cup B_i\} &= \Pr\{\mu(u) \leq r_i/(1+\varepsilon)^2 \text{ and } \mu(v) \geq r_i/(1+\varepsilon)\} \\ &\leq \Pr\{r_i \in [(1+\varepsilon)^2\mu(u), (1+\varepsilon)\mu(v)]\} \\ &\leq (1+\varepsilon)(\mu(v) - (1+\varepsilon)\mu(u)) \leq 2(\|\bar{u}\| - \|\bar{v}\|)^2/\varepsilon. \end{aligned}$$

Therefore, we have

$$\Pr\{u \in A_i', v \in P_i \cup B_i\} \leq 2\eta(u, v).$$

If $v \notin \tilde{P}_i \cup \tilde{B}_i$, then $\Pr\{u \in A_i', v \in \tilde{P}_i \cup \tilde{B}_i\} = 0$, and

$$\Pr\{u \in P_i, v \in P_i \cup B_i\} = \Pr\{u \in P_i\} = \|\bar{u}\|^2 = \eta(u, v).$$

□

By Claim 4.2, the expected number of edges from set P_i to set $V \setminus (P_i \cup B_i)$ is at most $2\eta(\tilde{P}_i)$. Also, the expected number of edges from set A_i' to set $P_i \cup B_i$ is at most $2\eta(\tilde{P}_i)$. In other words, $\mathbf{E} [\delta_G(P_i, V \setminus (P_i \cup B_i))] \leq 2\eta(\tilde{P}_i)$ and $\mathbf{E} [\delta_G(A_i', P_i \cup B_i)] \leq 2\eta(\tilde{P}_i)$. Using the definition of set $\mathcal{I}$ (see (4.27)), we get the claims 2 and 4 of Lemma 4.2.

Finally, we prove claim 5 of Lemma 4.2. We have for any edge (u, v) with $u \in \tilde{P}_i \cup \tilde{B}_i$,

$$\Pr\{u \in P_i \cup B_i, v \in (\Sigma_T \cup R_P) \setminus \tilde{P}_i\} \leq \Pr\{u \in P_i \cup B_i\} = (1+\varepsilon)\|\bar{u}\|^2 \leq 2\tilde{\eta}(u, v).$$

Thus, the expected number of edges from $P_i \cup B_i$ to $(\Sigma_T \cup R_P) \setminus \tilde{P}_i$ is at most $2\tilde{\eta}(\tilde{P}_i \cup \tilde{B}_i)$. By the definition of set $\mathcal{I}$ (see (4.27)), we get the conclusion. □

Using Lemma 2.1 with six random variables, we conclude that there exists $r_i \in (0, 1)$ such that inequalities (1) $|B_i| \leq 10C'_\delta \varepsilon |P_i|$, (2) $\delta_G(P_i, V \setminus (P_i \cup B_i)) \leq 5C''_\delta/\varepsilon \lambda_k \log k \cdot d|P_i|$, (3) $|A''_i| \leq 10\varepsilon |P_i|$, (4) $\delta_G(A'_i, P_i \cup B_i) \leq 5C''_\delta/\varepsilon \lambda_k \log k \cdot d|P_i|$, and (5) $\delta_G(P_i \cup B_i, (\Sigma_T \cup R_P) \setminus \tilde{P}_i) \leq 5C''_\delta/\varepsilon \lambda_k \log k \cdot d|P_i|$ hold simultaneously. The second inequality is equivalent to $\phi(P_i \parallel B_i) \leq 5C''_\delta/\varepsilon \lambda_k \log k$. In this theorem, we use the following functions $C''_{4.1}$ and $C'_{4.1}$: $C'_{4.1}(\delta) = 10C'_\delta$ and $C''_{4.1}(\delta) = 5C''_\delta$. Combining the inequalities (4) and (5), we get the property (6) in Theorem 4.1. In the Partial Partitioning Algorithm, all sets P_i, B_i, A'_i, A''_i for $i \in [k']$ and R'_B, R'_P are disjoint and cover the entire graph. Since all tuples (P_i, B_i, A'_i, A''_i) with $P_i = \emptyset$ are discarded at Step 4, all sets P_i returned by the Partial Partitioning Algorithm are nonempty. Note that $R'_B \subseteq R_B \cup \bigcup_{i=1}^T \tilde{B}_i$. Since we assume $|R_B| + \sum_{i=1}^T |\tilde{B}_i| \leq 16\varepsilon n$ at Step 2 (This condition holds with probability at least $3/4$), we have $|R'_B| \leq 16\varepsilon n$. This finishes the proof of Theorem 4.1. $\square$

5 From Disjoint Sets to Partitioning

We now show how to use the partial partitioning given by the Partial Partitioning Algorithm in Section 4 to obtain a true ε -buffered partitioning. We prove the following lemma.

LEMMA 5.1. *Consider a d -regular graph G . Let $\{(P_i, B_i, A'_i, A''_i)\}_{i \in [k']}$ and R'_P, R'_B be a partial ε -buffered partitioning of G given by the Partial Partitioning Algorithm. Then, for every $k \in \{1, \dots, k'\}$ and $\delta' = (k' - k + 1)/k'$, we can convert this partial partitioning into a true $54\varepsilon/\delta'$ -buffered partitioning $P'_1, \dots, P'_k, B'_1, \dots, B'_k$ of G such that*

$$\phi_G(P'_1, \dots, P'_k \parallel B'_1, \dots, B'_k) \leq \frac{4C''_{4.1}(\delta)}{\delta'} \cdot \frac{\log k}{\varepsilon} \lambda_k.$$

Proof. Let us sort all pairs (P_i, B_i, A'_i, A''_i) by size and assume $|P_1| \leq \dots \leq |P_{k'}|$. Now, we generate the true buffered partitioning of the graph. The true buffered partitioning (P'_i, B'_i) contains the pairs (P_i, B_i) for $i \in [k-1]$ in the partial partitioning and a pair of new sets (P'_k, B'_k) . Specifically, we let $P'_i = P_i$ and $B'_i = B_i$ for $i \in [k-1]$ and

$$P'_k = R'_P \cup \bigcup_{j=1}^{k'} A'_j \cup \bigcup_{j=k}^{k'} P_j; \quad B'_k = R'_B \cup \bigcup_{j=1}^{k'} A''_j \cup \bigcup_{j=k}^{k'} B_j.$$

We can think of each set A''_i is the buffer for the set A'_i for $i \in [k']$, and the set R'_B is the buffer for the set R'_P . We also combine these sets and buffers with the largest $k' - k + 1$ pairs (P_i, B_i) for $i = k, k+1, \dots, k'$ in the partial partitioning, respectively.

By Theorem 4.1, all sets P_i, B_i, A'_i, A''_i and R'_P, R'_B are disjoint and cover the entire graph. Also, all sets P_i and R'_P are nonempty. Thus, all sets P'_i are disjoint and nonempty, and $\bigcup_{i=1}^k P'_i \cup B'_i = V$. Also, for all $i \in [k-1]$, we have $|B_i| \leq \varepsilon |P_i|$ and

$$(5.28) \quad \phi_G(P'_i, B'_i) = \phi_G(P_i, B_i) \leq \frac{C''_{4.1}(\delta)}{\varepsilon} \lambda_k \log k.$$

It remains to verify that the last pair of sets P'_k and B'_k satisfy the required conditions. By items 4 and 5 of Theorem 4.1, we have

$$|B'_k| \leq |R'_B| + \sum_{j=1}^{k'} |A''_j| + \sum_{j=1}^{k'} |B_j| \leq 16\varepsilon n + 11\varepsilon \sum_{j=1}^{k'} |P_j| \leq 27\varepsilon n.$$

Since $|P_1| \leq \dots \leq |P_{k'}|$, we have $\sum_{i=1}^{k-1} |P_i| \leq \frac{k-1}{k'} \sum_{i=1}^{k'} |P_i|$. Thus, we have

$$\begin{aligned} |P'_k| &= |V| - |R'_B| - \sum_{i=1}^{k'} |A''_i| + |B_i| - \sum_{i=1}^{k-1} |P_i| \geq \\ &\geq \left(1 - \frac{k-1}{k'}\right) \cdot \left(|V| - |R'_B| - \sum_{i=1}^{k'} |A''_i| + |B_i|\right) \geq \delta'(n - 27\varepsilon n) \geq \delta'n/2. \end{aligned}$$

Hence, we have $|B'_k| \leq 54\varepsilon/\delta' |P'_k|$.

We now bound the buffered expansion of this last part. By items (3) and (6) of Theorem 4.1, we have

$$\begin{aligned} \phi_G(P'_k \parallel B'_k) &\leq \frac{\sum_{i=1}^{k-1} \delta_G(P'_k, P_i \cup B_i)}{d|P'_k|} \\ &\leq \frac{\sum_{i=1}^{k-1} \sum_{j=1}^{k'} \delta_G(A'_j, P_i \cup B_i) + \delta_G(R'_P, P_i \cup B_i) + \sum_{j=k}^{k'} \delta_G(P_j, P_i \cup B_i)}{d \cdot \delta' n/2} \\ &\leq \frac{2C''_{4.1}(\delta)/\varepsilon \cdot \lambda_k \log k \cdot d \sum_{i=1}^{k-1} |P_i| + \sum_{j=k}^{k'} \delta_G(P_j, V \setminus (P_j \cup B_j))}{d \cdot \delta' n/2} \\ &\leq \frac{4C''_{4.1}(\delta)/\delta'}{\varepsilon} \cdot \lambda_k \log k. \end{aligned}$$

This concludes the proof of Lemma 5.1. $\square$

We now prove the main result of the paper, Theorem 1.1.

Proof. [Proof of Theorem 1.1] Let $\hat{k} = \lfloor (1 + \delta)k \rfloor$ and $\hat{\delta} = \min\{(1 - 1/\sqrt{1 + \delta})/2, 1/80\}$. Let $k' = \lceil (1 - 2\hat{\delta})\hat{k} \rceil$ and $\delta' = (k' - k + 1)/k'$. We first use the Partial Partitioning Algorithm from Section 4 with parameters $\hat{k}$, $\hat{\varepsilon} = \varepsilon\delta'/54$, and $\hat{\delta}$ to obtain a partial $\hat{\varepsilon}$ -buffered partitioning $(P_1, B_1, A'_1, A''_1), \dots, (P_{k'}, B_{k'}, A'_{k'}, A''_{k'})$. By Theorem 4.1, the buffered expansion of each set P_i with buffer set B_i is at most $C''_{4.1}(\delta)/\varepsilon \lambda_{\hat{k}} \log \hat{k}$. Then, we apply Lemma 5.1 to transform this partial partitioning into a true k partitioning. Since $k' = \lceil (1 - 2\hat{\delta})\hat{k} \rceil$, we have $k' \geq \sqrt{1 + \delta}k - 1$. Then, we have $\delta' \geq 1 - 1/\sqrt{1 + \delta}$. By Lemma 5.1, the expansion of this ε -buffered k partitioning is at most $c(\delta)/\varepsilon \lambda_k \log k$, where $c(\delta) = 4C''_{4.1}(\delta)/\delta'$ is a function that only depends on δ . $\square$

6 Spectral Embedding

Consider a d -regular graph G . Let L_G be its normalized Laplacian. Let $x_1, \dots, x_n$ be an orthonormal eigenbasis for L_G and λ_i be the eigenvalue of x_i . Without loss of generality, we assume that $\lambda_1 \leq \dots \leq \lambda_n$. Note that $\lambda_1 = 0$, so we may assume that $x_1 = \mathbf{1}/\sqrt{n}$. Define an $k \times n$ matrix $U = (x_1, \dots, x_k)^T$; that is, the (i, u) entry of U equals $U(i, u) = x_i(u)$ where $i \in [k]$ and $u \in V$. Rows of U are indexed by integers from 1 to k and columns by vertices $u \in V$ of the graph (to simplify notation, we may assume that $V = [n]$). Note that $UU^T = I_k$, since vectors $x_1, \dots, x_k$ are orthonormal. Let $\{e_u\}_{u \in V}$ be the standard orthonormal basis in $\mathbb{R}^V$.

We are ready to define the spectral embedding of G . Let $\bar{u}$ be the column of U indexed by vertex u . The spectral embedding maps vertex u to vector $\bar{u}$.

Define $\psi(u) = u_i/\|u_i\|$. For a subset of vertices $S \subseteq V$, let $\mu(S) = \sum_{u \in S} \|\bar{u}\|^2$ be the measure of set S . Now we will state and prove basic properties of the spectral embedding.

CLAIM 6.1. *For all $u \in V$, we have $0 < \|\bar{u}\| \leq 1$.*

Proof. Since $x_1 = \mathbf{1}/\sqrt{n}$, for all $u \in V$, we have $\bar{u}(1) = 1/\sqrt{n}$ and $\|\bar{u}\| \geq 1/\sqrt{n} > 0$. Further,

$$\|\bar{u}\|^2 = \sum_{i=1}^k \bar{u}(i)^2 = \sum_{i=1}^k x_i(u)^2 = \sum_{i=1}^k \langle x_i, e_u \rangle^2 \leq \sum_{i=1}^n \langle x_i, e_u \rangle^2 = \|e_u\|^2 = 1.$$

$\square$

CLAIM 6.2. *We have*

1. $\sum_{u \in V} \|\bar{u}\|^2 = k$
2. $\sum_{(u,v) \in E} \|\bar{u} - \bar{v}\|^2 \leq kd\lambda_k$

Proof. Note that the (u, v) entry of matrix $U^T U$ equals $\langle \bar{u}, \bar{v} \rangle$, since U has columns $\bar{u}$ for $u \in V$.

1. We have, $\sum_{u \in V} \|\bar{u}\|^2 = \text{tr}(U^T U) = \text{tr}(UU^T) = \text{tr } I_k = k$, as required.

2. We have,

$$\begin{aligned} \sum_{(u,v) \in E} \|\bar{u} - \bar{v}\|^2 &= \sum_{(u,v) \in E} \sum_{i=1}^k \|\bar{u}(i) - \bar{v}(i)\|^2 = \sum_{i=1}^k \sum_{(u,v) \in E} \|x_i(u) - x_i(v)\|^2 \\ &\stackrel{\text{by (1.1)}}{=} d \sum_{i=1}^k x_i^T L_G x_i = d \sum_{i=1}^k \lambda_i \leq dk \lambda_k, \end{aligned}$$

where we used that $\lambda_1 \leq \dots \leq \lambda_k$ in the last inequality. $\square$

We show that the spectral embedding vectors $\{\psi(\bar{v})\}$ satisfy the following spreading property. It is a variant of Lemma 3.2 from the paper by Lee, Oveis-Gharan and Trevisan [LGT14].

LEMMA 6.1. *Assume that we are given a parameter $R \in [0, 1/\sqrt{2}]$. For every vertex u , consider the ball of radius R around u , $\text{Ball}(u, R) = \{v : \|\psi(\bar{u}) - \psi(\bar{v})\| \leq R\}$. Then $\mu(\text{Ball}(u, R)) \leq 1/(1 - 2R^2)$ for every u .*

Proof. Consider a vertex $u \in V$ and $C = \text{Ball}(u, R)$. Let $a_v = \|\bar{v}\|$ for $v \in C$. Then, $\bar{v} = a_v \psi(\bar{v})$ for $v \in C$. We have, $\mu(C) = \sum_{v \in C} a_v^2$. By the definition of C , $\|\psi(\bar{u}) - \psi(\bar{v})\| \leq R$ for $v \in C$ and hence $\|\psi(\bar{v}) - \psi(\bar{w})\| \leq 2R$ for all pairs $v, w \in C$. Therefore,

$$(6.29) \quad \langle \psi(\bar{v}), \psi(\bar{w}) \rangle = 1 - \frac{\|\psi(\bar{v}) - \psi(\bar{w})\|^2}{2} \geq 1 - 2R^2 \quad \text{for all } v, w \in C.$$

Write,

$$\mu(C) = \sum_{v \in C} a_v^2 = \frac{1}{\sum_{v \in C} a_v^2} \sum_{v, w \in C} a_v^2 a_w^2.$$

By inequality (6.29),

$$a_v a_w \leq \frac{a_v a_w \langle \psi(\bar{v}), \psi(\bar{w}) \rangle}{1 - 2R^2} = \frac{\langle \bar{v}, \bar{w} \rangle}{1 - 2R^2}.$$

Thus,

$$\mu(C) \leq \frac{1}{\sum_{v \in C} a_v^2} \sum_{v, w \in C} \frac{a_v a_w \langle \bar{v}, \bar{w} \rangle}{1 - 2R^2}.$$

For any vertex $v \in V$, let $e_v \in \mathbb{R}^V$ be the standard basis vector where $e_v(v) = 1$ and $e_v(u) = 0$ for all $u \neq v$. Let

$$z = \frac{\sum_{v \in C} a_v e_v}{\sqrt{\sum_{v \in C} a_v^2}}.$$

For any standard basis vector e_v , we have $U e_v = \bar{v}$. Therefore,

$$U z = \frac{1}{\sqrt{\sum_{v \in C} a_v^2}} \sum_{v \in C} a_v \bar{v},$$

and

$$\mu(C) \leq \frac{z^T (U^T U) z}{(1 - 2R^2)}.$$

We prove that $\|U z\|^2 = z^T (U^T U) z \leq 1$. To this end, note that z is a unit vector and $\|U z\|^2 \leq \sigma_{\max}(U)^2 = \sigma_{\max}(U^T)^2$, where $\sigma_{\max}(U)$ and $\sigma_{\max}(U^T)$ are the largest singular values of U and U^T , respectively (here, we used the definition of singular values and the fact that matrices U and U^T have the same non-zero singular values). Since $U U^T = I_d$, all singular values of U^T are equal to 1. We conclude that $\|U z\|^2 \leq 1$. $\square$

7 Orthogonal Separators with Buffers – Proofs

In this section, we show the algorithm that generates orthogonal separators with buffers. We prove Theorem 3.1, Theorem 3.2, and Theorem 3.3.

THEOREM 3.1 . *There exists a randomized polynomial-time procedure that given a finite set U of unit vectors in $\mathbb{R}^d$ and positive parameters $\varepsilon \in (0, 1)$, $m \geq 3$, $R \in (0, 2)$, returns an m -orthogonal separator with an ε -buffer with distortion $\mathcal{D} = O_R(1/\varepsilon \log m)$, separation radius R , and probability scale $\alpha \geq O_R(1/\text{poly}(m))$.*

For two disjoint random sets $X, Y \subset U$ chosen from this orthogonal separator distribution, we have the following properties:

1. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in X\} = \alpha$; (for some α that depends on m and R).
2. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$.
3. For all $\bar{u}, \bar{v} \in U$ with $\|\bar{u} - \bar{v}\| \geq R$, $\Pr\{\bar{v} \in X \mid \bar{u} \in X\} \leq \frac{1}{m}$.
4. For all $\bar{u}, \bar{v} \in U$, $\Pr\{\bar{v} \notin X \cup Y \mid \bar{u} \in X\} \leq \mathcal{D} \|\bar{u} - \bar{v}\|^2$, where $\mathcal{D} = O_R(1/\varepsilon \log m)$.

Proof. We use the following procedure to generate orthogonal separators with buffers. We sample a d -dimensional Gaussian vector $g \sim \mathcal{N}(0, I_d)$. For every vector $\bar{u}$ in U , we let $g_u = \langle \bar{u}, g \rangle$ be the projection of vector $\bar{u}$ on the direction g . For a standard gaussian random variable $Z \sim \mathcal{N}(0, 1)$, we use $\bar{\Phi}(t) = \Pr\{Z \geq t\}$ to denote the probability that $Z \geq t$. We pick a threshold t such that $\bar{\Phi}(t) = \alpha$ for some α that we will specify later; our choice of α will guarantee that $t \leq 1$. Let $\varepsilon' = \varepsilon/(e(t + 1/t))$. Then, we construct the orthogonal separator X and the buffer Y as follows:

$$X = \{\bar{u} : g_u \geq t\}; \quad Y = \{\bar{u} : t - \varepsilon' < g_u < t\}.$$

Now we show that this procedure satisfies the required properties.

1. For every vector $\bar{u} \in U$, we have

$$\Pr\{\bar{u} \in X\} = \Pr\{g_u \geq t\} = \bar{\Phi}(t) = \alpha.$$

2. For every vector $\bar{u} \in U$, we have

$$\begin{aligned} \Pr\{\bar{u} \in Y\} &= \Pr\{t - \varepsilon' < g_u < t\} \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\varepsilon')^2}{2}} \cdot \varepsilon' \leq \\ &\leq \frac{\varepsilon' e^{\varepsilon' t}}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = \frac{\varepsilon e^\varepsilon}{e\sqrt{2\pi}(t + 1/t)} e^{-\frac{t^2}{2}} \leq \frac{e^\varepsilon}{e} \cdot \varepsilon \bar{\Phi}(t) \leq \varepsilon\alpha, \end{aligned}$$

where the third inequality is due to Lemma G.1.

3. For every $\bar{u}, \bar{v} \in U$, we have

$$\Pr\{\bar{u} \in X, \bar{v} \in X\} = \Pr\{g_u \geq t, g_v \geq t\} \leq \Pr\{(g_u + g_v)/2 \geq t\}.$$

We know that g_u, g_v are both random Gaussian variables from $\mathcal{N}(0, 1)$. Thus, we have $(g_u + g_v)/2$ is also a Gaussian variable with variance

$$\mathbf{Var} \left[\frac{g_u + g_v}{2} \right] = \frac{1}{4} \mathbf{E}[(g_u + g_v)^2] = \frac{1}{4} (2 + 2\langle \bar{u}, \bar{v} \rangle) = 1 - \frac{\|\bar{u} - \bar{v}\|^2}{4},$$

where the second equality is due to $\mathbf{E}[g_u g_v] = \langle \bar{u}, \bar{v} \rangle$ and the third equality used $\bar{u}, \bar{v}$ are unit vectors. Thus for every $\bar{u}, \bar{v} \in U$ with $\|\bar{u} - \bar{v}\| \geq R$, we have $\mathbf{Var}[(g_u + g_v)/2] \leq 1 - R^2/4$. From Lemma G.2 we get that there exists a constant C such that

$$\Pr \left\{ \frac{g_u + g_v}{2} \geq t \right\} \leq \bar{\Phi} \left(\frac{t}{\sqrt{1 - R^2/4}} \right) \leq \frac{1}{t} (Ct \bar{\Phi}(t))^{\frac{1}{\sqrt{1 - R^2/4}}}.$$

Since $\bar{\Phi}(t) = \alpha$, we have

$$\Pr\{\bar{u} \in X, \bar{v} \in X\} \leq \Pr\left\{\frac{g_u + g_v}{2} \geq t\right\} \leq \alpha \cdot C(Ct\alpha)^{\frac{1}{\sqrt{1-R^2/4}}-1}.$$

By Lemma G.1, we have $t = \Theta(\sqrt{\log 1/\alpha})$. Then we can find some $\alpha \geq 1/\text{poly}(m)$ (for a fixed R) that depends on m and R such that $\Pr\{\bar{u} \in X, \bar{v} \in X\} \leq \alpha/m$. Since $\Pr\{\bar{u} \in X\} = \alpha$, we have

$$\Pr\{\bar{v} \in X \mid \bar{u} \in X\} \leq \frac{1}{m}.$$

4. For every $\bar{u}, \bar{v} \in U$, we have

$$\Pr\{\bar{u} \in X, \bar{v} \notin X \cup Y\} = \Pr\{g_u \geq t, g_v \leq t - \varepsilon'\}.$$

Since g is a standard Gaussian random vector, we have g_u and g_v are jointly Gaussian random variables with distribution $\mathcal{N}(0, 1)$. Since $\varepsilon \leq 1$ and $t = \Theta(\sqrt{\log m})$, we have $\varepsilon' = \varepsilon/(e(t+1/t)) < t$. Using Lemma G.3 on g_u, g_v with parameters $\hat{m} = 1/\alpha$ and $\hat{\varepsilon} = \varepsilon'$, we get

$$\Pr\{g_u \geq t, g_v \leq t - \varepsilon'\} \leq O\left(\frac{\sqrt{\log \hat{m}}}{\varepsilon' \hat{m}}\right) \cdot \|\bar{u} - \bar{v}\|^2 \leq \alpha \mathcal{D} \|\bar{u} - \bar{v}\|^2,$$

where $\mathcal{D} = O_R(1/\varepsilon \log m)$. $\square$

THEOREM 3.2 . *There exists a randomized procedure that given a finite set U of unit vectors in $\mathbb{R}^d$ equipped with a measure μ and positive parameters $\varepsilon \in (0, 1), \delta \leq 2/3, R \in (0, 2)$, returns an δ -orthogonal separator with an ε -buffer with distortion $\mathcal{D} = O_R(1/\varepsilon \log 1/\delta)$, separation radius R , and probability scale $\alpha \geq O_R(1/\text{poly}(m))$.*

For two disjoint random sets $X, Y \subset U$ chosen from this orthogonal separator distribution, we have the following properties:

1. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in X\} \in [\alpha/2, \alpha]$.
2. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$.
3. $\min_{\bar{u} \in X} \mu(X \setminus \text{Ball}(\bar{u}, R)) \leq \delta\mu(U)$ (always).
4. For all $\bar{u}, \bar{v} \in U$, $\Pr\{\bar{v} \notin X \cup Y \mid \bar{u} \in X\} \leq \mathcal{D} \|\bar{u} - \bar{v}\|^2$, where $\mathcal{D} = O_R(1/\varepsilon \log 1/\delta)$.

Proof. We first run the algorithm from Theorem 3.1 with $m = 2/\delta$ and obtain sets X' and Y' . If set X' satisfies the third condition: $\min_{\bar{u} \in X'} \mu(X' \setminus \text{Ball}(\bar{u}, R)) \leq \delta\mu(U)$, we return sets $(X, Y) = (X', Y')$. Otherwise, we return empty sets, $(X, Y) = (\emptyset, \emptyset)$. By Theorem 3.1, $\Pr\{\bar{u} \in X\} \leq \alpha$ and $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$ for all $\bar{u} \in X$. Also, condition (3) always holds (because if X' does not satisfy it, we return $\emptyset$). We now lower bound $\Pr\{\bar{u} \in X\}$:

$$\begin{aligned} \Pr\{\bar{u} \in X\} &= \Pr\{\bar{u} \in X'\} - \Pr\{\bar{u} \in X' \text{ and } X = \emptyset\} \\ &= \Pr\{\bar{u} \in X'\} \cdot (1 - \Pr\{X = \emptyset \mid \bar{u} \in X'\}) \\ &= \alpha(1 - \Pr\{X = \emptyset \mid \bar{u} \in X'\}). \end{aligned}$$

If $X = \emptyset$, then

$$\mu(X' \setminus \text{Ball}(\bar{u}, R)) \geq \min_{\bar{v} \in X'} \mu(X' \setminus \text{Ball}(\bar{v}, R)) > \delta\mu(U).$$

Thus,

$$\Pr\{X = \emptyset \mid \bar{u} \in X'\} \leq \Pr\left\{\mu(X' \setminus \text{Ball}(\bar{u}, R)) > \delta\mu(U) \mid \bar{u} \in X'\right\}.$$

However, by item (3) of Theorem 3.1,

$$\mathbf{E}\left[\mu(X' \setminus \text{Ball}(\bar{u}, R)) \mid \bar{u} \in X'\right] \leq \frac{\mu(U)}{m} = \frac{\delta\mu(U)}{2}.$$

By Markov's inequality,

$$\Pr\{X = \emptyset \mid \bar{u} \in X'\} \leq \frac{1}{2}.$$

Therefore, $\Pr\{\bar{u} \in X\} \geq \alpha(1 - 1/2) = \alpha/2$. Finally,

$$\begin{aligned} \Pr\{\bar{v} \notin X \cup Y \mid \bar{u} \in X\} &= \frac{\Pr\{\bar{v} \notin X \cup Y \text{ and } \bar{u} \in X\}}{\Pr\{\bar{u} \in X\}} \\ &= \frac{\Pr\{\bar{v} \notin X' \cup Y' \text{ and } \bar{u} \in X'\}}{\Pr\{\bar{u} \in X'\}} \cdot \frac{\Pr\{\bar{u} \in X'\}}{\Pr\{\bar{u} \in X\}} \\ &\leq 2 \Pr\{\bar{v} \notin X' \cup Y' \mid \bar{u} \in X'\} \leq 2\mathcal{D} \|\bar{u} - \bar{v}\|^2. \end{aligned}$$

□

THEOREM 3.3 . *There exists a randomized procedure that given a finite set U of unit vectors in $\mathbb{R}^d$ equipped with a measure μ and positive parameters $\varepsilon \in (0, 1)$, $\delta \leq 2/3$, $R \in (0, 2)$, returns an δ -orthogonal separator with two ε -buffers with distortion $\mathcal{D} = O_R(1/\varepsilon \log^{1/\delta})$, separation radius R , and probability scale $\alpha \geq O_R(1/\text{poly}(m))$.*

For three disjoint random sets $X, Y, Z \subset U$ chosen from this orthogonal separator distribution, we have the following properties:

1. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in X\} \in [\alpha/2, \alpha]$.
2. For all $\bar{u} \in U$, $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$, and $\Pr\{\bar{u} \in Z\} \leq \varepsilon\alpha$.
3. $\min_{\bar{u} \in X} \mu(X \setminus \text{Ball}(\bar{u}, R)) \leq \delta\mu(U)$ (always).
4. For all $\bar{u}, \bar{v} \in U$, $\Pr\{\bar{v} \notin X \cup Y \mid \bar{u} \in X\} \leq \mathcal{D} \|\bar{u} - \bar{v}\|^2$, and $\Pr\{\bar{v} \notin X \cup Y \cup Z \mid \bar{u} \in X \cup Y\} \leq \mathcal{D} \|\bar{u} - \bar{v}\|^2$, where $\mathcal{D} = O_R(1/\varepsilon \log^{1/\delta})$.

Proof. We modify the algorithm in Theorem 3.1 to generate three disjoint sets X', Y', Z' as follows. We sample a d -dimensional Gaussian vector $g \sim \mathcal{N}(0, I_d)$. For every vector $\bar{u}$ in U , we let $g_u = \langle \bar{u}, g \rangle$ be the projection of vector $\bar{u}$ on the direction g . We use $\bar{\Phi}(t)$ to denote the probability that a standard gaussian random variable is at least t . We pick a threshold t such that $\bar{\Phi}(t) = \alpha$ for some α that we will specify later; our choice of α will guarantee that $t \leq 1$. Let $\varepsilon' = \varepsilon/(e(t + 1/t))$. Then, we construct the orthogonal separator X' and two buffers Y', Z' as follows:

$$X = \{\bar{u} : g_u \geq t\}; \quad Y = \{\bar{u} : t - \varepsilon' < g_u < t\}; \quad Z = \{\bar{u} : t - 2\varepsilon' < g_u < t - \varepsilon'\}.$$

If set X' satisfies the third condition: $\min_{\bar{u} \in X'} \mu(X' \setminus \text{Ball}(\bar{u}, R)) \leq \delta\mu(U)$, we return sets $(X, Y, Z) = (X', Y', Z')$. Otherwise, we return empty sets, $(X, Y, Z) = (\emptyset, \emptyset, \emptyset)$.

By the similar analysis in Theorem 3.1, we have for all $\bar{u} \in U$, it holds that $\Pr\{\bar{u} \in X\} \leq \alpha$, $\Pr\{\bar{u} \in Y\} \leq \varepsilon\alpha$, and $\Pr\{\bar{u} \in Z\} \leq \varepsilon\alpha$. By Theorem 3.2, we have for all $\bar{u} \in U$, $\Pr\{\bar{u} \in X\} \geq \alpha/2$ and condition (3) always holds. Then, we show that condition (4) holds. The first part of condition (4) is the same as Theorem 3.2. Note that $\alpha \leq \bar{\Phi}(t - \varepsilon') \leq (1 + \varepsilon)\alpha$. Using Lemma G.3 on g_u, g_v with parameters $\hat{m} = 1/\bar{\Phi}(t - \varepsilon')$ and $\hat{\varepsilon} = \varepsilon'$, we have

$$\Pr\{g_u \geq t, g_v \leq t - \varepsilon'\} \leq O\left(\frac{\sqrt{\log \hat{m}}}{\varepsilon' \hat{m}}\right) \cdot \|\bar{u} - \bar{v}\|^2 \leq \alpha \mathcal{D} \|\bar{u} - \bar{v}\|^2,$$

where $\mathcal{D} = O_R(1/\varepsilon \log m)$. □

References

- [Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.
- [AM85] Noga Alon and Vitali D Milman. λ_1 , Isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

- [AR04] Konstantin Andreev and Harald Räcke. Balanced graph partitioning. In *Proceedings of the Symposium on Parallelism in Algorithms and Architectures*, pages 120–124, 2004.
- [ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):1–37, 2009.
- [BFK⁺14] Nikhil Bansal, Uriel Feige, Robert Krauthgamer, Konstantin Makarychev, Viswanath Nagarajan, Joseph Seffi Naor, and Roy Schwartz. Min-max graph partitioning and small set expansion. *SIAM Journal on Computing*, 43(2):872–904, 2014.
- [Che69] Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian. In *Proceedings of the Princeton conference in honor of Professor S. Bochner*, pages 195–199, 1969.
- [CMM06] Eden Chlamtac, Konstantin Makarychev, and Yury Makarychev. How to play unique games using embeddings. In *Proceedings of the Symposium on Foundations of Computer Science*, pages 687–696, 2006.
- [Fil22] Arnold Filtser. Hop-constrained metric embeddings and their applications. In *Proceedings of the Symposium on Foundations of Computer Science*, pages 492–503, 2022.
- [FL21] Arnold Filtser and Hung Le. Clan embeddings into trees, and low treewidth graphs. In *Proceedings of the Symposium on Theory of Computing*, pages 342–355, 2021.
- [HHZ21] Bernhard Haeupler, D Ellis Hershkowitz, and Goran Zuzic. Tree embeddings for hop-constrained network design. In *Proceedings of the Symposium on Theory of Computing*, pages 356–369, 2021.
- [KBL16] Emilie Kaufmann, Thomas Bonald, and Marc Lelarge. A spectral algorithm with additive clustering for the recovery of overlapping communities in networks. In Ronald Ortner, Hans Ulrich Simon, and Sandra Zilles, editors, *Algorithmic Learning Theory*, pages 355–370, Cham, 2016. Springer International Publishing.
- [KLL⁺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. In *Proceedings of the Symposium on Theory of Computing*, pages 11–20, 2013.
- [KLL17] Tsz Chiu Kwok, Lap Chi Lau, and Yin Tat Lee. Improved cheeger’s inequality and analysis of local graph partitioning using vertex expansion and expansion profile. *SIAM Journal on Computing*, 46(3):890–910, 2017.
- [KLM06] Ravi Kannan, László Lovász, and Ravi Montenegro. Blocking conductance and mixing in random walks. *Combinatorics, Probability and Computing*, 15(4):541–570, 2006.
- [KVV04] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 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):1–30, 2014.
- [LM12] Jure Leskovec and Julian McAuley. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25, 2012.
- [LM14] Anand Louis and Konstantin Makarychev. Approximation algorithm for sparsest k -partitioning. In *Proceedings of the Symposium on Discrete Algorithms*, pages 1244–1255, 2014.
- [LR99] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM)*, 46(6):787–832, 1999.
- [LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the Symposium on Theory of Computing*, pages 1131–1140, 2012.
- [McS01] Frank McSherry. Spectral partitioning of random graphs. In *Proceedings of the Symposium on Foundations of Computer Science*, pages 529–537. IEEE, 2001.
- [NJW01] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Proceedings of NeurIPS*, pages 8490–856, 2001.
- [OATT22] Lorenzo Orecchia, Konstantinos Ameranis, Charalampos Tsourakakis, and Kunal Talwar. Practical almost-linear-time approximation algorithms for hybrid and overlapping graph clustering. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 17071–17093. PMLR, 17–23 Jul 2022.
- [PSL90] Alex Pothen, Horst D. Simon, and Kang-Pu Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications*, 11(3):430–452, 1990.
- [Räc08] Harald Räcke. Optimal hierarchical decompositions for congestion minimization in networks. In *Proceedings of the Symposium on Theory of computing*, pages 255–264, 2008.
- [RS10] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the Symposium on Theory of Computing*, pages 755–764, 2010.
- [RST12] Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. In *Proceedings of the Conference on Computational Complexity*, pages 64–73, 2012.
- [SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

- [SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.
- [Spi07] Daniel A Spielman. Spectral graph theory and its applications. In *Proceedings of the Symposium on Foundations of Computer Science*, pages 29–38, 2007.
- [ST97] Horst D Simon and Shang-Hua Teng. How good is recursive bisection? *SIAM Journal on Scientific Computing*, 18(5):1436–1445, 1997.
- [ST07] Daniel A Spielman and Shang-Hua Teng. Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra and its Applications*, 421(2-3):284–305, 2007.
- [vL07] Ulrike von Luxburg. A tutorial on spectral clustering. *CoRR*, abs/0711.0189, 2007.
- [YL12] Jaewon Yang and Jure Leskovec. Community-affiliation graph model for overlapping network community detection. In *2012 IEEE 12th International Conference on Data Mining*, pages 1170–1175, 2012.
- [YL14] Jaewon Yang and Jure Leskovec. Structure and overlaps of ground-truth communities in networks. *ACM Trans. Intell. Syst. Technol.*, 5(2), apr 2014.

A Connection to Robust Expansion

In this section, we prove Corollary 1.2.

Proof. [Proof of Corollary 1.2] Let $\varepsilon^* = \phi_\eta^V(G)$ be the robust vertex expansion of G . If $\varepsilon^* = 0$, then the claim is trivial, because $\lambda_2 \geq 0$. So we assume below that $\varepsilon^* > 0$. Then for every disjoint subsets $S, T \subset V$ with $0 < |S| \leq |V|/2$ and $|T| < \varepsilon^*|S|$, we have

$$(A.1) \quad \delta(S, T) < (1 - \eta)\delta(S, V \setminus S),$$

as otherwise, we would have a contradiction

$$\varepsilon^* = \phi_\eta^V(G) \leq \phi_\eta^V(S) = \frac{N_\eta(S)}{|S|} \leq \frac{|T|}{|S|} < \varepsilon^*.$$

Now we apply Corollary 1.1 of Theorem 1.1 with $k = 2$ and $\varepsilon' = \varepsilon^*/2$. We get an ε' -buffered partition $(P_1, P_2 || B_1, B_2)$ with $\phi_G(P_1, P_2 || B_1, B_2) \leq O(\lambda_2/\varepsilon')$. Assume without loss of generality that $|P_1| \leq n/2$. Note that $|B_1| \leq \varepsilon'|P_1| < \varepsilon^*|P_1|$ and thus by (A.1),

$$\delta(P_1, B_1) < (1 - \eta)\delta(P_1, V \setminus P_1).$$

Therefore,

$$\delta(P_1, V \setminus (P_1 \cup B_1)) = \delta(P_1, V \setminus P_1) - \delta(P_1, B_1) > \eta\delta(P_1, V \setminus P_1).$$

On the other hand,

$$\delta(P_1, V \setminus (P_1 \cup B_1)) \leq d \cdot \phi_G(P_1, P_2 || B_1, B_2) \cdot |P_1| \leq O\left(\frac{d\lambda_2|P_1|}{\varepsilon^*}\right).$$

We conclude that

$$\lambda_2 \geq \Omega(\eta) \cdot \varepsilon^* \cdot \frac{\delta(P_1, V \setminus P_1)}{d|P_1|} = \Omega(\eta \cdot \phi_\eta^V(G) \cdot \phi_G(P_1)) \geq \Omega(\eta \cdot \phi_\eta^V(G) \cdot h_G).$$

□

B Heavy Set P_t in a Buffered Partition

In this section, we argue why we may assume that one of the sets P_t in the buffered partitioning $(P_1, \dots, P_k || B_1, \dots, B_k)$ contains at least $\Omega(\delta n)$ vertices (where $n = |V|$).

COROLLARY B.1. *There exists a buffered partitioning as in Theorem 1.1 (possibly with a different function $c(\delta)$) such that $|P_t| = \Omega(\delta n)$ for some t .*

Proof. Let $\delta' = \sqrt{1 + \delta} - 1 = \Theta(\delta)$ and $k' = \lfloor (1 + \delta')k \rfloor$. Apply Theorem 1.1 with parameters k' and δ' . We get an ε -buffered partitioning $(P_1, \dots, P_{k'} \parallel B_1, \dots, B_{k'})$ with

$$\phi_0 = \phi_G(P_1, \dots, P_{k'} \parallel B_1, \dots, B_{k'}) \leq \frac{c(\delta') \log k'}{\varepsilon} \lambda_{\lfloor (1+\delta)k \rfloor}.$$

Assume without loss of generality that $|P_1| \leq |P_2| \leq \dots \leq |P_{k'}|$. Merge sets $P_k, \dots, P_{k'}$ and sets $B_k, \dots, B_{k'}$. That is, let $P'_k = \bigcup_{i=k}^{k'} P_i$ and $B'_k = \bigcup_{i=k}^{k'} B_i$. We obtain a buffered partitioning $(P_1, \dots, P_{k-1}, P'_k \parallel B_1, \dots, B_{k-1}, B'_k)$. We show that it is ε -buffered and that its buffered expansion is at most ϕ_0 . Clearly, merging does not change the value of $\phi_G(P_i \parallel B_i)$ for $i \in [k-1]$, as it does not change sets P_i and B_i . So it is sufficient to verify that $|B'_k| \leq \varepsilon |P'_k|$ and $\phi_G(P'_k \parallel B'_k) \leq \phi_0$. Indeed,

$$\begin{aligned} |B'_k| &\leq \sum_{i=k}^{k'} |B_i| \leq \sum_{i=k}^{k'} \varepsilon |P_i| = \varepsilon |P'_k|. \\ \phi_G(P'_k \parallel B'_k) &= \frac{\delta_G(P'_k, V \setminus (P'_k \cup B'_k))}{|P'_k|} \leq \frac{\sum_{i=k}^{k'} \delta_G(P_i, V \setminus (P_i \cup B_i))}{|P'_k|} \leq \frac{\sum_{i=k}^{k'} \phi_0 |P_i|}{|P'_k|} = \phi_0. \end{aligned}$$

We used that sets $P_k, \dots, P_{k'}$ are disjoint and thus $|P'_k| = |P_k| + \dots + |P_{k'}|$. Finally, we observe that P'_k is the union of $k' - k + 1 = \Omega(\delta k)$ largest sets out of k' sets that together cover at least $(1 - \varepsilon)n$ vertices. Thus, $|P'_k| \geq \frac{k' - k + 1}{k'} (1 - \varepsilon)n = \Omega(\delta n)$. $\square$

C Lower Bound for k -way Expansion and Pseudo-approximation Algorithm for Sparsest k -way Partitioning

In this section, we present the lower bound for non-buffered k -way expansion h_G^k of graphs with vertex weights and edge costs. The proof is similar to that for graphs without vertex weights shown in [LRTV12, LGT14]. Combined with Theorem 1.2, it gives a pseudo-approximation algorithm for the Sparsest k -way Partitioning problem.

PROPOSITION C.1. *Given any graph $G = (V, E, w, c)$ with vertex weights $w_u > 0$ and edge costs $c_{uv} > 0$, for any integer $k > 1$, the k -way expansion is at least*

$$h_G^k \geq \frac{\lambda_k}{2}.$$

Proof. Let $P_1, P_2, \dots, P_k$ be the optimal solution for k -way expansion. Then, we have for any $i \in [k]$

$$\phi_G(P_i) = \frac{|\delta(P_i, V \setminus P_i)|}{w(P_i)} \leq h_G^k.$$

Let $\mathbf{1}_{P_i}$ be the indicator vector of set P_i for all $i \in [k]$, i.e. $\mathbf{1}_{P_i}(u) = 1$ if $u \in P_i$, otherwise $\mathbf{1}_{P_i}(u) = 0$. Then, we use $x_{P_i} = D_w^{1/2} \mathbf{1}_{P_i}$ to denote the weighted indicator vector. Let $X = \{x_{P_i} : i \in [k]\}$. Since all vectors in X are orthogonal to each other, the span of X has dimension k . By the Courant-Fischer Theorem, we have

$$(C.2) \quad \lambda_k = \min_{S \subset \mathbb{R}^n : \dim(S) = k} \max_{x \in S} \frac{x^T D_w^{-1/2} L_G D_w^{-1/2} x}{x^T x} \leq \max_{x \in \text{span}(X)} \frac{x^T D_w^{-1/2} L_G D_w^{-1/2} x}{x^T x}.$$

Suppose $x \in \text{span}(X)$ is the maximizer of the right-hand side of Equation (C.2). We can write $x = \sum_{i=1}^k \alpha_i x_{S_i}$ for $\alpha_i \in \mathbb{R}$. Then, we have

$$\begin{aligned} x^T D_w^{-1/2} L_G D_w^{-1/2} x &= \left(\sum_{i=1}^k \alpha_i \mathbf{1}_{S_i} \right)^T L_G \left(\sum_{i=1}^k \alpha_i \mathbf{1}_{S_i} \right) = \\ &= \sum_{(u,v) \in E} c_{uv} \left(\sum_{i=1}^k \alpha_i \mathbf{1}_{S_i}(u) - \sum_{i=1}^k \alpha_i \mathbf{1}_{S_i}(v) \right)^2 \leq 2 \sum_{i=1}^k \alpha_i^2 \sum_{(u,v) \in E} c_{uv} (\mathbf{1}_{S_i}(u) - \mathbf{1}_{S_i}(v))^2, \end{aligned}$$

where the last inequality is due to the relaxed triangle inequality, for any edge $(u, v) \in E$ with $u \in S_i$ and $v \in S_j$, $(\alpha_i \mathbf{1}_{S_i}(u) - \alpha_j \mathbf{1}_{S_j}(v))^2 \leq 2\alpha_i^2 \mathbf{1}_{S_i}(u)^2 + 2\alpha_j^2 \mathbf{1}_{S_j}(v)^2$. Taking it into Equation (C.2), we have

$$\lambda_k \leq \frac{2 \sum_{i=1}^k \alpha_i^2 \sum_{(u,v) \in E} c_{uv} (\mathbf{1}_{S_i}(u) - \mathbf{1}_{S_i}(v))^2}{\sum_{i=1}^k \alpha_i^2 \sum_{u \in V} w_u \mathbf{1}_{S_i}(u)} = \frac{2 \sum_{i=1}^k \alpha_i^2 |\delta(P_i, V \setminus P_i)|}{\sum_{i=1}^k \alpha_i^2 w(P_i)} \leq 2h_G^k.$$

□

Plugging the bound on $\lambda_{\lfloor (1+\delta)k \rfloor}(L_G)$ from Proposition C.1 into Theorem 1.2, we get the following $O_{\varepsilon, \delta}(\log k)$ pseudo-approximation algorithm for the Sparsest K -Partitioning problem from

THEOREM C.1. *There exists a polynomial-time algorithm that given a graph $G = (V, E, w, c)$ with vertex weights $w_u > 0$ and edge costs $c_{uv} > 0$, $\varepsilon > 0$, $\delta > 0$, and $k > 1$ such that $\max_{u \in V} w_u \leq \varepsilon w(V)/(3k)$, finds a ε -buffered partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ with*

$$\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k) \leq \frac{\kappa(\delta) \log k}{\varepsilon} h_G^{\lfloor (1+\delta)k \rfloor}.$$

Note that in this theorem, we compare the cost of our ε -buffered k -partition to that of the optimal non-buffered $\lfloor (1+\delta)k \rfloor$ -partition.

D Buffered Balanced Cut

In this section, we present our results for the buffered balanced cut. Consider any graph $G(V, E, w, c)$ with vertex weight $w_u > 0$ and edge cost $c_{uv} > 0$. For any $0 < \gamma \leq 1/2$, the γ -balanced cut of graph G is a partition of graph (L, R) such that $w(L), w(R) \in [\gamma w(V), (1-\gamma)w(V)]$. The γ -balanced cut problem asks to find a γ -balanced cut of a graph to minimize the cut size $\delta(L, R)$. We consider the ε -buffered γ -balanced cut. Given a weighted graph $G(V, E, w, c)$, the ε -buffered γ -balanced cut is a partition of graph G , $(L, R \parallel B)$ such that $w(L), w(R) \in [\gamma w(V), (1-\gamma)w(V)]$ and $w(B) \leq \varepsilon \min(w(L), w(R))$. We show a bi-criteria approximation for the balanced cut problem with an ε -buffered balanced cut.

THEOREM D.1. *Let $\varepsilon \in (0, 1/4)$. Consider any weighted graph $G = (V, E, w, c)$ with vertex weight $w_u > 0$ and $c_{uv} > 0$. There is a polynomial-time algorithm that finds three disjoint sets L, B, R with $L \cup B \cup R = V$, $w(L), w(R) \in [1/4 \cdot w(V), 3/4 \cdot w(V)]$, and $w(B) \leq 3\varepsilon \min(w(L), w(R))$ such that*

$$\delta(L, R) \leq O(1/\varepsilon) \cdot \delta(L^*, R^*),$$

where (L^*, R^*) is the optimal $1/3$ -balanced cut. $(L, R \parallel B)$ is a (3ε) -buffered $1/4$ -balanced cut with cut size at most $O(1/\varepsilon)$ times the size of the optimal $1/3$ -balanced cut.

Proof. We first describe our algorithm for buffered balanced cut, which is inspired by the approximation algorithm for balanced cut in [LR99]. The algorithm recursively partitions the graph by using the buffered spectral partitioning algorithm in Section 2. At the beginning, we set the graph $G_1 = G$. Then, we run the ε -buffered spectral partitioning to find a partition $(L_1, R_1 \parallel B_1)$ of the graph G_1 . Suppose $w(L_1) \leq w(R_1)$. If $w(L_1) < w(V)/4$, then we recursively run the ε -buffered spectral partitioning on the subgraph G_2 of G on the set of vertices R_1 . For each call of buffered spectral partitioning, we label the partition $(L_t, R_t \parallel B_t)$ such that $w(L_t) \leq w(R_t)$. We recursively call the ε -buffered spectral partitioning until $\sum_{t=1}^T w(L_t) \geq w(V)/4$. Then, the algorithm returns the partition (L, R, B) of G , where $L = \bigcup_{t=1}^T L_t$, $B = \bigcup_{t=1}^T B_t$, and $R = V \setminus (L \cup B)$.

Then, we show that the partition $(L, R \parallel B)$ returned by this algorithm is a 3ε -buffered $1/4$ -balanced cut. Let $(L_t, R_t \parallel B_t)$ be the buffered partition of graph G_t returned by the t -th call of the buffered spectral partitioning. Then, we have $w(L_t) \leq w(V_t)/2$ and $w(B_t) \leq \varepsilon w(L_t)$. Suppose the algorithm calls the buffered spectral partitioning for T times. Then, we have $w(L) = \sum_{t=1}^T w(L_t) \geq w(V)/4$ and $\sum_{t=1}^{T-1} w(L_t) < w(V)/4$. Since $w(V_T) \leq w(V)$, we have

$$w(L) = \sum_{t=1}^T w(L_t) \leq \sum_{t=1}^{T-1} w(L_t) + w(L_T) \leq w(V)/4 + w(V_T)/2 \leq 3/4 \cdot w(V).$$

Since $w(L) \geq w(V)/4$, we have $w(R) \leq 3/4 \cdot w(V)$. Since $w(L_T) \leq w(V_T)/2$ and $w(B_T) \leq \varepsilon w(L_T)$, we have

$$w(R) = w(V_T) - w(L_T) - w(B_T) \geq \left(1 - \frac{1+\varepsilon}{2}\right) w(V_T).$$

Note that $w(V_T) = w(V) - \sum_{t=1}^{T-1} w(L_t) + w(B_t) \geq (1 - \frac{1+\varepsilon}{4}) w(V)$. Since $\varepsilon \leq 1/4$, we have

$$w(R) \geq \left(1 - \frac{1+\varepsilon}{2}\right) \left(1 - \frac{1+\varepsilon}{4}\right) w(V) \geq \frac{w(V)}{4}.$$

Thus, we have both $w(L)$ and $w(R)$ are in $[w(V)/4, 3w(V)/4]$. Since $w(B_t) \leq \varepsilon w(L_t)$ for all t , we have $w(B) \leq \varepsilon w(L)$ and

$$w(B) \leq \varepsilon w(L) \leq \varepsilon \cdot \frac{3}{4} w(V) \leq 3\varepsilon \cdot w(R).$$

Hence, we have $w(B) \leq 3\varepsilon \cdot \min\{w(L), w(R)\}$.

Next, we bound the size of buffered cut (L, B, R) . For each call of the buffered spectral partitioning, we bound the cut size $\delta(L_t, R_t)$ for the buffered partition (L_t, B_t, R_t) of graph G_t . Let (L^*, R^*) be the optimal non-buffered $1/3$ -balanced partition of graph G . Let $L_t^* = L^* \cap V_t$ and $R_t^* = R^* \cap V_t$. Then, we have $\delta(L_t^*, R_t^*) \leq \delta(L^*, R^*)$. Note that the weight of vertices in $V \setminus V_t$ is at most

$$w(V \setminus V_t) = \sum_{i=1}^{t-1} w(L_i) + w(B_i) \leq (1 + \varepsilon) \cdot \frac{w(V)}{4}.$$

Suppose $w(L_t^*) \geq w(R_t^*)$. Since $w(L^*) \geq w(V)/3$ and $\varepsilon \leq 1/4$, we have

$$w(L_t^*) \geq w(L^*) - w(V \setminus V_t) \geq \left(\frac{1}{3} - \frac{1+\varepsilon}{4}\right) w(V) \geq \frac{1}{48} w(V).$$

By Proposition C.1, we have

$$\frac{\lambda_2(L_{G_t})}{2} \leq \min_{S \subset V_t: w(S) \leq w(V_t)/2} \frac{\delta(S, V_t \setminus S)}{w(S)} \leq \frac{\delta(L_t^*, R_t^*)}{w(L_t^*)}.$$

By Proposition 2.1, we have

$$\begin{aligned} \delta(L_t, R_t) &\leq 4 \left(1 + \frac{8}{\varepsilon}\right) \lambda_2(L_{G_t}) \cdot w(L_t) \leq \\ &\leq 8 \left(1 + \frac{8}{\varepsilon}\right) \cdot \frac{w(L_t)}{w(L_t^*)} \cdot \delta(L_t^*, R_t^*) \leq O\left(\frac{1}{\varepsilon}\right) \cdot \frac{w(L_t)}{w(V)} \cdot \delta(L_t^*, R_t^*). \end{aligned}$$

Combining all cuts edges in $\delta(L_t, R_t)$ for T calls of buffered spectral partitioning, we have

$$\delta(L, R) \leq \sum_{t=1}^T \delta(L_t, R_t) \leq O\left(\frac{1}{\varepsilon}\right) \cdot \sum_{t=1}^T \frac{w(L_t)}{w(V)} \cdot \delta(L_t^*, R_t^*) \leq O\left(\frac{1}{\varepsilon}\right) \delta(L^*, R^*),$$

where the last inequality is due to $w(L) \leq 3/4 \cdot w(V)$. $\square$

We also consider the k -way balanced partition problem. Given a graph $G(V, E, w, c)$, for any $\gamma \geq 1$, we say that $P_1, P_2, \dots, P_k$ is a (γ, k) -balanced partition of G if $w(P_i) \leq \gamma w(V)/k$ for all $i \in [k]$. The (γ, k) -balanced partition problem aims to find a (γ, k) balanced partition to minimize the total cost of edges with two endpoints in different parts. By using the buffered balanced cut algorithm in Theorem D.1 and the recursive bi-section algorithm in [ST97], we show a bi-criteria approximation for the k -way balanced partition.

COROLLARY D.1. *Let $\varepsilon \in (0, 1/4)$. Consider any weighted graph $G = (V, E, w, c)$ with vertex weight $w_u > 0$ and $c_{uv} > 0$. There is a polynomial-time algorithm that finds a ε -buffered $(6, k)$ -balanced partition $P_1, P_2, \dots, P_k, B$ such that $P_1, P_2, \dots, P_k$ and B are disjoint, $w(B) \leq O(\varepsilon)w(V)$, and*

$$\sum_{i < j} \delta(P_i, P_j) \leq O(1/\varepsilon \cdot \log^2 k) \cdot \text{OPT},$$

where OPT is the optimal cost for $(1, k)$ -balanced partition.

E Graphs with Vertex Weights and Edge Costs

In this section, we prove our main results for graphs $G = (V, E, w, c)$ with vertex weights $w_u > 0$ and edge costs $c_{uv} > 0$.

Theorem 1.1 holds for regular graphs with parallel edges but without edge costs and vertex weights. Assume that we have a graph G with edge costs c_{uv} and with vertex weights $w_u = 1$ such that the total cost of all edges incident on a vertex does not depend on the vertex; that is, $C_0 = \sum_{v:(u,v) \in E} c_{uv}$ does not depend on u . If all edge costs are integers, we can simulate edge costs by adding parallel edges – we replace each edge (u, v) with c_{uv} parallel edges. We obtain a C_0 -regular graph G' . Let $L_{G'} = I - \frac{1}{C_0} A_{G'}$ be the normalized Laplacian of G' . Let $L_G = D_w^{-1/2} \tilde{L}_G D_w^{-1/2}$ be the normalized Laplacian of G . It is immediate that $L_G = C_0 L_{G'}$ and $\delta_G(A, B) = \delta_{G'}(A, B)$ for every $A, B \subseteq V$. Let $k' = \lfloor (1 + \delta)k \rfloor$. Then, $\lambda_{k'}(L_{G'}) = \lambda_{k'}(L_G)/C_0$.

By Theorem 1.1, there exists an ε -buffered partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ such that

$$\phi_{G'}(P_i \parallel B_i) = \frac{\delta_{G'}(P_i, V \setminus (P_i \cup B_i))}{C_0 |P_i|} \leq \frac{c(\delta) \log k}{\varepsilon} \cdot \lambda_{k'}(L_{G'})$$

for every $i \in [k]$. Since $\lambda_{k'}(L_{G'}) = \lambda_{k'}(L_G)/C_0$ and $w(P_i) = |P_i|$, we have for all i ,

$$(E.3) \quad \phi_G(P_i \parallel B_i) = \frac{\delta_G(P_i, V \setminus (P_i \cup B_i))}{w(P_i)} \leq \frac{c(\delta) \log k}{\varepsilon} \cdot \lambda_{k'}(L_G).$$

Now if we multiply all edge costs by the same positive number ρ , both the left and right hand side will get multiplied by ρ . Therefore, the inequality holds not only for integer edge costs but also for arbitrary positive rational costs. By continuity, it holds for arbitrary positive edge costs. We get the following corollary.

COROLLARY E.1. *Let G be a graph with positive edge costs c_{uv} and unit vertex weights such that $C_0 = \sum_{v:(u,v) \in E} c_{uv}$ is the same for all vertices u . Then there exists an ε -balanced partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ such that inequality (E.3) holds for all i .*

Now we present a black-box reduction that proves Theorem 1.2. We note that the reduction can significantly increase the running time of the algorithm. However, in fact, we can use the algorithm from Theorem 1.1 to find $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ (the proof of this fact essentially repeats that of Theorem 1.1, and we do not present it here).

THEOREM E.1. *Let $G = (V, E, w, c)$ be a graph with positive weights $w_u > 0$ and edge costs $c_{uv} > 0$, $\varepsilon \in [0, 1)$, $\delta \in (0, 1)$, and $k \geq 2$ be an integer. Assume that $\max_u w_u \leq \varepsilon w(V)/(3k)$. Let $L_G = D_w^{-1/2} \tilde{L}_G D_w^{-1/2}$ be the normalized Laplacian of G . Then*

$$(E.4) \quad h_G^{k, \varepsilon} \leq \frac{\kappa(\delta) \log k}{\varepsilon} \cdot \lambda_{\lfloor (1+\delta)k \rfloor}(L_G),$$

where $\kappa(\delta)$ is a function that depends only on δ .

Proof. Assume first that all vertex weights are integers greater than or equal to 2. Let $W = \sum_{u \in V} w_u$ be the total weight of all vertices. Let $C = \sum_{(u,v) \in E} c_{uv}$ be the total cost of all edges and $B = C \cdot W^2$.

We construct an auxiliary graph G' with unit vertex weights as follows. For each vertex u of G , we create its own “cloud of vertices” Q_u of size w_u ; all vertices $q \in Q_u$ have unit weights. For $(u, v) \in E$, we connect every $q \in Q_u$ with every $q' \in Q_v$ by an edge (q, q') with cost $c'_{qq'} = \frac{c_{uv}}{|Q_u||Q_v|}$. Note that the total cost of all edges between Q_u and Q_v equals c_{uv} . Let $b_u = \sum_{v:(u,v) \in E} \frac{c_{uv}}{|Q_u|}$ be the total cost of edges incident on vertex $q \in Q_u$ (so far). Now we connect every two vertices $q, q' \in Q_u$ by an edge of cost $c'_{qq'} = \frac{B - b_u}{|Q_u| - 1}$. After this step, the total cost of all edges incident on $q \in Q_u$ is exactly B , since q has $|Q_u| - 1$ neighbors in Q_u . We denote the obtained graph by G' .

Properties of $G' = (V', E')$ that we established.

- $|Q_u| = w_u$; all vertices have unit weights in G' .
- The total cost of all the edges between Q_u and Q_v is c_{uv} .

- The total cost of all edges incident on every vertex equals B (and does not depend on u).
- $G'[Q_u]$ is a clique, in which all edges have cost $c'_{qq'} = \frac{B-b_u}{|Q_u|-1} \geq \frac{B-C}{W-1} > CW$.

Now we upper bound $\lambda_{k'}(L_{G'})$ in terms of $\lambda_{k'}(L_G)$.

LEMMA E.1. $\lambda_{k'}(L_{G'}) \leq \lambda_{k'}(L_G)$

Proof. Let $x_1, \dots, x_{k'}$ be the first k' orthogonal unit eigenvectors of L_G . Define vectors $z_1, \dots, z_{k'} \in \mathbb{R}^{|V'|}$ as follows: for $q \in Q_u$, we let $z_i(q) = \frac{x_i(u)}{\sqrt{w_u}}$. First, observe that $z_1, \dots, z_{k'}$ are pairwise orthogonal unit vectors:

$$\langle z_i, z_j \rangle = \sum_{q \in V'} z_i(q) z_j(q) = \sum_{u \in V} \sum_{q \in Q_u} z_i(q) z_j(q) = \sum_{u \in V} |Q_u| \frac{x_i(u) x_j(u)}{w_u} = \langle x_i, x_j \rangle = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases}$$

Further,

$$\begin{aligned} z_i^T L_{G'} z_j &= \sum_{(q,q') \in E'} c'_{qq'} (z_i(q) - z_i(q')) \cdot (z_j(q) - z_j(q')) \\ &= \sum_{(u,v) \in E} \sum_{\substack{q \in Q_u \\ q' \in Q_v \\ (q,q') \in E'}} \frac{c_{uv}}{|Q_u||Q_v|} (z_i(q) - z_i(q')) \cdot (z_j(q) - z_j(q')) \\ &\quad + \sum_{u \in V} \frac{B - b_u}{|Q_u| - 1} \sum_{\substack{q, q' \in Q_u \\ (q,q') \in E'}} (z_i(q) - z_i(q')) \cdot (z_j(q) - z_j(q')) \\ &= \sum_{(u,v) \in E} c_{uv} \left(\frac{x_i(u)}{w_u^{1/2}} - \frac{x_i(v)}{w_v^{1/2}} \right) \cdot \left(\frac{x_j(u)}{w_u^{1/2}} - \frac{x_j(v)}{w_v^{1/2}} \right) = x_i^T L_G x_j. \end{aligned}$$

We conclude that $z_i^T L_{G'} z_j = \lambda_i(L_G)$ if $i = j$, and $z_i^T L_{G'} z_j = 0$, otherwise.

Finally, we use the Courant–Fischer theorem to upper bound $\lambda_{k'}(L_G)$. Let H be the linear span of vectors $z_1, \dots, z_{k'}$. By the Courant–Fischer theorem,

$$\begin{aligned} \lambda_{k'}(L_{G'}) &\leq \max_{z \in H \setminus \{0\}} \frac{z^T L_{G'} z}{\|z\|^2} = \max_{\substack{z = \sum_i \alpha_i z_i \\ \alpha \in \mathbb{R}^{k'} \setminus \{0\}}} \frac{z^T L_{G'} z}{\|z\|^2} = \max_{\alpha \in \mathbb{R}^{k'} \setminus \{0\}} \frac{\sum_{i,j} (\alpha_i \alpha_j) z_i^T L_{G'} z_j}{\|\alpha\|^2} \\ &= \max_{\alpha \in \mathbb{R}^{k'} \setminus \{0\}} \frac{\sum_i \alpha_i^2 \lambda_i(L_G)}{\|\alpha\|^2} = \lambda_{k'}(L_G). \end{aligned}$$

□

Let $\varepsilon' = \varepsilon/10$. We apply Theorem 1.1 to G' and obtain an ε' -buffered partition $(P'_1, \dots, P'_k \parallel B'_1, \dots, B'_k)$ of G' with $\phi_{G'}(P'_1, \dots, P'_k \parallel B'_1, \dots, B'_k) \leq \frac{c(\delta) \log k}{\varepsilon'} \lambda_{k'}(L_{G'}) \leq \frac{c(\delta) \log k}{\varepsilon'} \lambda_{k'}(L_G)$. Observe that if some set Q_u contains a vertex $q \in P'_i$ and a vertex $q' \in P'_j \cup B'_j$ with $j \neq i$ then $\phi_{G'}(P'_i \parallel B'_i)$ is very large

$$\phi_{G'}(P'_i \parallel B'_i) \geq \frac{\delta_{G'}(P'_i, P'_j \cup B'_j)}{w(P'_i)} \geq \frac{c_{qq'}}{W} > C.$$

Then, any partition $(P_1, \dots, P_k \parallel \emptyset, \dots, \emptyset)$ of G satisfies the condition of the theorem:

$$\phi_G(P_i \parallel \emptyset) \leq C/2 < \phi_{G'}(P'_i \parallel B'_i) \leq \frac{c(\delta) \log k}{\varepsilon'} \lambda_{k'}(L_G),$$

as required. So we assume below that if $P'_i \cap Q_u \neq \emptyset$ then $(P'_j \cup B'_j) \cap Q_u = \emptyset$ for every u, i , and $j \neq i$. Then for every u , there are two possibilities: either

1. $Q_u \subseteq P'_i \cup B'_i$ for some i , or
2. $Q_u \subseteq \bigcup_i B'_i$.

Depending on which of the possibilities takes place, we say that u is a vertex of the first or second type, respectively⁷. Now we define an ε -buffered partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ of G . First, we assign every vertex u to one of the sets $P_1, \dots, P_k, B_1, \dots, B_k$ and U , where U is a special set that will be partitioned among $B_1, \dots, B_k$ later. We do that as follows:

1. if $|Q_u \cap P'_i| \geq |Q_u|/2$, we assign u to P_i ;
2. otherwise, if $|Q_u \cap B'_i| \geq |Q_u|/2$, we assign u to B_i ;
3. otherwise, we assign u to U .

Note that each vertex of the first type is necessarily assigned to some P_i or B_i . Each vertex of the second type is assigned to some B_i or U .

Since U consists of the vertices of the second type, we have $\bigcup_{u \in U} Q_u \subset \bigcup_i B'_i$ and thus

$$w(U) = \left| \bigcup_{u \in U} Q_u \right| \leq \left| \bigcup_i B'_i \right| \leq \varepsilon' \left| \bigcup_i P'_i \right|.$$

Here we used that that partition $(P'_1, \dots, P'_k \parallel B'_1, \dots, B'_k)$ is ε' -buffered. We create sets $B''_1, \dots, B''_k$, which are initially empty, and set the capacity of B''_i to $\frac{\varepsilon|P'_i|}{2}$. We distribute vertices from U one-by-one among $B''_1, \dots, B''_k$ so that the total weight assigned to B''_i does not exceed its capacity. We stop when we either assign all the vertices from U or no unassigned vertex in U can be assigned to any B''_i , without violating the capacity requirement for B''_i . We now show that this procedure assigns all the vertices from U . Indeed, assume that some vertex u is not assigned. Then, w_u is greater than the the remaining capacity of every B''_i ; that is, $w_u > \frac{\varepsilon|P'_i|}{2} - w(B''_i)$ for every i . Adding up these inequalities over all i , we get

$$\begin{aligned} kw_u &> \sum_{i=1}^k \left(\frac{\varepsilon|P'_i|}{2} - w(B''_i) \right) \geq \frac{\varepsilon}{2} \left| \bigcup_i P'_i \right| - w(U) \geq \frac{\varepsilon}{2} \left| \bigcup_i P'_i \right| - \left| \bigcup_i B'_i \right| \\ &\geq \frac{\varepsilon}{2} \left| \bigcup_i P'_i \right| - \varepsilon' \left| \bigcup_i P'_i \right| \geq \frac{2\varepsilon}{5} \left| \bigcup_i P'_i \right| \stackrel{*}{\geq} \frac{2\varepsilon(1-\varepsilon')}{5} w(V) \geq \frac{\varepsilon w(V)}{3} \end{aligned}$$

Inequality $\stackrel{*}{\geq}$ above follows from two inequalities: $|\bigcup_i P'_i| + |\bigcup_i B'_i| = w(V)$ and $|\bigcup_i B'_i| \leq \varepsilon' |\bigcup_i P'_i|$. We get that $w_u > \frac{\varepsilon w(V)}{3k}$, which contradicts to the assumption of the theorem. We conclude that $\bigcup_i B''_i = U$. Finally, we add vertices from B''_i to B_i for every i . We obtain the desired partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$.

Now we prove that $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ satisfies the desired requirements. Fix i . We upper bound $\delta_G(P_i, V \setminus (P_i \cup B_i))$. Note that if edge (u, v) goes from P_i to $V \setminus (P_i \cup B_i)$ then u is a vertex of the first type and $|Q_u \cap P'_i| \geq |Q_u|/2$ and either

- v is a vertex of the first type and $Q_v \subseteq P_j \cup B_j$ for some $j \neq i$, or
- v is a vertex of the second type and at least one half of the vertices in Q_v are not in B_i (and none of them are in P_i).

To summarize, in either case at least a half of the vertices in Q_u lie in P'_i and at least half of vertices in Q_v do not lie in $P'_i \cup B'_i$. Thus, at least one quarter of all edges from Q_u to Q_v contribute to $\delta_{G'}(P'_i, V' \setminus (P'_i, B'_i))$, and their total contribution is at least $c_{uv}/4$. We conclude that

$$\delta_G(P_i, V \setminus (P_i \cup B_i)) \leq 4\delta_{G'}(P'_i, V' \setminus (P'_i \cup B'_i)).$$

⁷If $Q_u \subseteq B'_i$, let us assume that u is of the first type.

Now we lower bound $w(P_i)$. Let A be the set of vertices u of the first type such that $Q_u \subseteq P'_i \cup B'_i$. Note that $P_i \subseteq A$ and $P'_i \subseteq \bigcup_{u \in A} Q_u$. Consider $u \in A$. If $u \in P_i$, then $w(P_i \cap \{u\}) = |Q_u| \geq |Q_u \cap P'_i| - |Q_u \cap B'_i|$. If $u \notin P_i$, then $w(P_i \cap \{u\}) = 0 \geq |Q_u \cap P'_i| - |Q_u \cap B'_i|$, since $|Q_u \cap P'_i| < |Q_u|/2 \leq |Q_u \cap B'_i|$. We have,

$$w(P_i) = \sum_{u \in A} w(P_i \cap \{u\}) \geq \sum_{u \in A} |Q_u \cap P'_i| - |Q_u \cap B'_i| \geq |P'_i| - |B'_i| \geq (1 - \varepsilon')|P'_i|.$$

We have,

$$\phi_G(P_i \parallel B_i) = \frac{\delta_G(P_i, V \setminus (P_i \cup B_i))}{w(P_i)} \leq \frac{4}{1 - \varepsilon'} \frac{\delta_{G'}(P'_i, V \setminus (P'_i \cup B'_i))}{|P'_i|} = \frac{O(c(\delta)) \log k}{\varepsilon} \lambda_{k'}(L_G).$$

It remains to show that partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$ is ε -buffered. We already showed that $w(P_i) \geq (1 - \varepsilon')w(P'_i)$. Now we upper bound $w(B_i)$. First, $w(B''_i) \leq \varepsilon|P'_i|/2 \leq \varepsilon w(P_i)/2(1 - \varepsilon') \leq 5\varepsilon w(P_i)/9$.

Then, $u \in B_i \setminus B''_i$ if and only if $|Q_u \cap B'_i| \geq |Q_u|/2 = w_u/2$. Therefore,

$$w(B_i \setminus B''_i) \leq 2 \sum_{u \in B_i \setminus B''_i} |Q_u \cap B'_i| \leq 2|B'_i| \leq 2\varepsilon'|P'_i| \leq \frac{2\varepsilon'}{1 - \varepsilon'} w(P_i) \leq \frac{2\varepsilon}{9} w(P_i).$$

We conclude that $w(B_i) = w(B_i \setminus B''_i) + w(B''_i) = \frac{7\varepsilon}{9} w(P_i)$, as required. This completes the proof for the case when all the vertex weights are integers. By linearity, inequality (E.4) also holds when all the weights are rational numbers, and by continuity, it follows that inequality (E.4) holds when weights are arbitrary positive real numbers. $\square$

F Lower Bound on $h_G^{k,\varepsilon}$

In this section, we prove Theorem 1.4, which we now restate as follows.

THEOREM F.1. *Consider a d -regular graph $G = (V, E)$ and its ε -buffered partition $(P_1, \dots, P_k \parallel B_1, \dots, B_k)$. Then for every $i \in [k]$,*

$$\lambda_k \leq 2\phi_G(P_1, \dots, P_k \parallel B_1, \dots, B_k) + \varepsilon.$$

Thus,

$$\lambda_k \leq 2h_G^{k,\varepsilon} + \varepsilon.$$

Proof. By the Courant-Fischer min-max theorem,

$$\lambda_k = \min_H \max_{z \in H: z \neq 0} \frac{z^T L_G z}{\|z\|^2},$$

where the minimum is over k -dimensional subspaces H of $\mathbb{R}^n$. Let b_i be the indicator vector of P_i : $b_i(u) = 1$ if $u \in P_i$ and $b_i(u) = 0$, otherwise. Let H be the linear span of $b_1, \dots, b_k$ and $z = \sum_{i=1}^k \alpha_i b_i$. Then,

$$\lambda_k \leq \max_{(\alpha_1, \dots, \alpha_k) \neq 0} \frac{z^T L_G z}{\|z\|^2}$$

First note that vectors b_i have disjoint supports and thus are mutually orthogonal. Therefore, $\|z\|^2 =$

$\sum_{i=1}^k \alpha_i^2 \|h_i\|^2 = \sum_{i=1}^k \alpha_i^2 |P_i|$. Now we upper bound $z^T L_G z$. We will use that $|\delta(P_i, B_i)| \leq d|B_i| \leq \varepsilon d|P_i|$.

$$\begin{aligned}
 dz^T L_G z &\stackrel{\text{by (1.1)}}{=} \sum_{\substack{i,j \in [k] \\ i < j}} (\alpha_i - \alpha_j)^2 \cdot |\delta(P_i, P_j)| + \sum_{i \in [k]} (\alpha_i - 0)^2 \cdot \left| \delta\left(P_i, \bigcup_j B_j\right) \right| \\
 &\leq \sum_{\substack{i,j \in [k] \\ i < j}} (2\alpha_i^2 + 2\alpha_j^2) \cdot |\delta(P_i, P_j)| + \sum_{i \in [k]} \left(\alpha_i^2 \cdot \left| \delta\left(P_i, \bigcup_{j:j \neq i} B_j \setminus B_i\right) \right| + \alpha_i^2 \cdot |\delta(P_i, B_i)| \right) \\
 &= 2 \sum_{\substack{i,j \in [k] \\ i \neq j}} \alpha_i^2 \cdot |\delta(P_i, P_j)| + \sum_{i \in [k]} \alpha_i^2 \cdot \left(\left| \delta\left(P_i, \bigcup_{j:j \neq i} B_j \setminus B_i\right) \right| + |\delta(P_i, B_i)| \right) \\
 &\leq \sum_{i \in [k]} \alpha_i^2 \cdot \left(2 \left| \delta\left(P_i, \bigcup_{j:j \neq i} P_j \cup \left(\bigcup_{j:j \neq i} B_j \setminus B_i\right) \right) \right| + |\delta(P_i, B_i)| \right) \\
 &\leq \sum_{i \in [k]} \alpha_i^2 \cdot \left(2|\delta(P_i, V \setminus (P_i \cup B_i))| + \varepsilon d|P_i| \right).
 \end{aligned}$$

Therefore,

$$\frac{z^T L_G z}{\|z\|^2} \leq \frac{1}{d} \max_{i \in [k]} \frac{2|\delta(P_i, V \setminus (P_i \cup B_i))| + \varepsilon d|P_i|}{|P_i|} = \max_{i \in [k]} \frac{2|\delta(P_i, V \setminus (P_i \cup B_i))|}{d|P_i|} + \varepsilon.$$

□

G Gaussian Distribution

In this section, we present several useful estimates on the Gaussian distribution. Let $X \sim \mathcal{N}(0, 1)$ be a one-dimensional Gaussian random variable. Denote the probability that $X \geq t$ by $\bar{\Phi}(t)$:

$$\bar{\Phi}(t) = \Pr\{X \geq t\}.$$

The first lemma gives an accurate estimate on $\bar{\Phi}(t)$ for large t .

LEMMA G.1. (see [CMM06, Lemma A.1]) For every $t > 0$,

$$\frac{t}{\sqrt{2\pi}(t^2 + 1)} e^{-\frac{t^2}{2}} < \bar{\Phi}(t) < \frac{1}{\sqrt{2\pi}t} e^{-\frac{t^2}{2}} \text{ and } \bar{\Phi}(t) = \Theta\left(\frac{e^{-\frac{t^2}{2}}}{t+1}\right).$$

LEMMA G.2. (see [CMM06, Lemma A.1, part 2]) For any $\rho \geq 1$ and $t \geq 0$, there exists a constant C such that

$$\bar{\Phi}(\rho t) \leq \frac{1}{t} (C t \bar{\Phi}(t))^{\rho^2}.$$

LEMMA G.3. Let X and Y be jointly $\mathcal{N}(0, 1)$ -Gaussian random variables. Denote $\delta^2 = 1/2 \mathbf{Var}[X - Y]$. Choose $m > 3$, threshold $t > 1$ such that $\bar{\Phi}(t) = 1/m$, and a parameter $\varepsilon \in [0, t]$. Then

$$\Pr\{X \geq t \text{ and } Y \leq t - \varepsilon\} \leq O(\delta^2 \varepsilon^{-1} \sqrt{\log m/m}).$$

Proof. Note that (1) the covariance of X and Y is $\mathbf{E}[XY] = 1 - \mathbf{Var}[X - Y]/2 = 1 - \delta^2$, and (2) by Lemma G.1, $t = \Theta(\sqrt{\log m})$. Denote $p = \Pr\{X \geq t \text{ and } Y \leq t - \varepsilon\}$. Note that if $\delta^2 \varepsilon^{-1} t \geq 1/32$, then the lemma trivially holds,

$$p = \Pr\{X \geq t \text{ and } Y \leq t - \varepsilon\} \leq \Pr\{X \geq t\} = \frac{1}{m} \leq O\left(\frac{\delta^2 \varepsilon^{-1} \sqrt{\log m}}{m}\right),$$

as required. So we assume below that $\varepsilon > 32\delta^2 t$. Let $\alpha = \mathbf{E}[XY] = 1 - \delta^2$. Consider Gaussian random variable $Z = \alpha X - Y$. Note that Z has mean 0 and variance $\mathbf{E}[Z^2] = \alpha^2 + 1 - 2\alpha^2 = 2\delta^2 - \delta^4$. Further, the covariance of X and Z is 0: $\mathbf{E}[XZ] = \alpha \mathbf{E}[X^2] - \mathbf{E}[XY] = 0$. In particular, for every $\tau \geq 0$,

$$(G.5) \quad \Pr\{Z \geq \tau\} = \bar{\Phi}(\tau/\sqrt{2\delta^2 - \delta^4}) \leq \bar{\Phi}\left(\frac{\tau}{\sqrt{2\delta}}\right) \stackrel{\text{by Lemma G.1}}{\leq} O\left(e^{-\left(\frac{\tau}{\sqrt{2\delta}}\right)^2/2}\right).$$

Therefore, X and Z are independent. We have,

$$p = \Pr\{X \geq t \text{ and } Y \leq t - \varepsilon\} = \Pr\{X \geq t \text{ and } \alpha X - Z \leq t - \varepsilon\} = \frac{1}{m} \Pr\{Z \geq \varepsilon + \alpha X - t \mid X \geq t\}$$

Define random variable $\Delta = X - t$. Then

$$\varepsilon + \alpha X - t = \varepsilon + (1 - \delta^2)(t + \Delta) - t \geq \varepsilon/2 + (1 - \delta^2)\Delta \geq \frac{\varepsilon + \Delta}{2},$$

where we used that $\varepsilon/2 - \delta^2 t \geq 0$ and $\delta^2 \leq \varepsilon/(2t) \leq t/(2t) = 1/2$. We have,

$$p \leq \frac{1}{m} \Pr\{Z \geq (\varepsilon + \Delta)/2 \mid \Delta \geq 0\} \stackrel{\text{by (G.5)}}{\leq} \frac{\mathbf{E}[e^{-(\frac{\varepsilon+\Delta}{2})^2/(4\delta^2)} \mid \Delta \geq 0]}{m}.$$

Let us upper bound the probability density function $f_\Delta(x)$ of Δ conditioned on the event $\Delta \geq 0$.

$$\begin{aligned} f_\Delta(x) &= \frac{e^{-(x+t)^2/2}}{\sqrt{2\pi}} \Big/ \Pr\{\Delta \geq 0\} = (t+1) \cdot \frac{e^{-t^2/2}}{\sqrt{2\pi}(t+1)} \cdot e^{-x^2/2-tx} \cdot m \\ &\leq O(t \cdot \bar{\Phi}(t) \cdot e^{-x^2/2-tx} \cdot m) = O(te^{-x^2/2-tx}) = O(te^{-tx}). \end{aligned}$$

We conclude that

$$\begin{aligned} pm &\leq O(1) \int_0^\infty e^{-\frac{(x+t)^2}{16\delta^2}} (te^{-tx}) dx = O(te^{4t^2\delta^2+\varepsilon t}) \int_0^\infty e^{-\frac{(x+8t\delta^2+\varepsilon)^2}{16\delta^2}} dx \\ &\stackrel{\text{let } \tilde{x} = \frac{x+8t\delta^2+\varepsilon}{2\sqrt{2}\delta}}{=} O\left(\delta te^{4t^2\delta^2+\varepsilon t}\right) \int_{8t\delta^2+\varepsilon/(2\sqrt{2}\delta)}^\infty e^{-\frac{\tilde{x}^2}{2}} d\tilde{x} \leq O\left(\delta te^{2\varepsilon t} \bar{\Phi}\left(\frac{\varepsilon}{2\sqrt{2}\delta}\right)\right) \\ &\stackrel{\text{by Lemma G.1}}{\leq} O\left(\frac{\delta^2}{\varepsilon} te^{2\varepsilon t - \varepsilon^2/(16\delta^2)}\right) = O\left(\frac{\delta^2 t}{\varepsilon}\right) \end{aligned}$$

here we three times used that $\varepsilon > 32\delta^2 t$. We conclude that $p = O(\delta^2 \varepsilon^{-1} t/m) = O(\delta^2 \varepsilon^{-1} \sqrt{\log m}/m)$, as required. $\square$

H Supplementary Figures

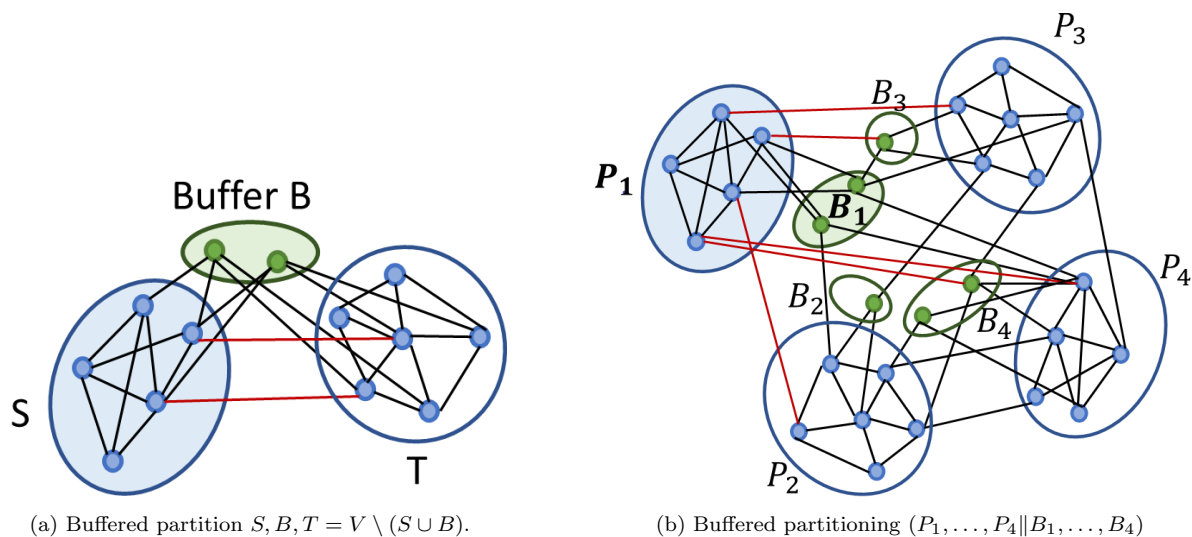


Figure 5: *Left:* The figure on the left shows a partition of the vertex set V into three pieces S, B and $T = V \setminus (S \cup B)$. Here B denotes the buffer, and cost of this cut is $\delta(S, T)$, as denoted by the edges marked in red. The edges marked in grey denote the edges between S and the buffer B .

Right: The illustrative figure shows a $k = 4$ partition P_1, P_2, P_3, P_4 with buffers B_1, B_2, B_3, B_4 . The red edges indicate the edges $\delta(P_1, V \setminus (P_1 \cup B_1))$ that contribute to the cut corresponding to P_1 . All parts $P_1, \dots, P_4$ and $B_1, \dots, B_4$ are disjoint.