

# On Weighted Graph Sparsification by Linear Sketching\*

Yu Chen<sup>†</sup>

Sanjeev Khanna<sup>†</sup>

Huan Li<sup>†</sup>

## Abstract

A seminal work of [Ahn-Guha-McGregor, PODS’12] showed that one can compute a cut sparsifier of an *unweighted* undirected graph by taking a near-linear number of linear measurements on the graph. Subsequent works also studied computing other graph sparsifiers using linear sketching, and obtained near-linear upper bounds for spectral sparsifiers [Kapralov-Lee-Musco-Musco-Sidford, FOCS’14] and first non-trivial upper bounds for spanners [Filtser-Kapralov-Nouri, SODA’21]. All these linear sketching algorithms, however, only work on *unweighted* graphs, and are extended to weighted graphs by weight grouping, a non-linear operation not implementable in, for instance, general turnstile streams.

In this paper, we initiate the study of *weighted* graph sparsification by linear sketching by investigating a natural class of linear sketches that we call *incidence sketches*, in which each measurement is a linear combination of the weights of edges incident on a *single* vertex. This class captures *all* aforementioned linear sketches for unweighted sparsification. It also covers linear sketches implementable in the *simultaneous communication model*, where edges are distributed across  $n$  machines. Our results are:

1. **Weighted cut sparsification:** We give an algorithm that computes a  $(1 + \epsilon)$ -cut sparsifier using  $\tilde{O}(n\epsilon^{-3})$ <sup>1</sup> linear measurements, which is nearly optimal. This also implies a turnstile streaming algorithm with  $\tilde{O}(n\epsilon^{-3})$  space. Our algorithm is achieved by building a so-called “weighted edge sampler” for each vertex.
2. **Weighted spectral sparsification:** We give an algorithm that computes a  $(1 + \epsilon)$ -spectral sparsifier using  $\tilde{O}(n^{6/5}\epsilon^{-4})$  linear measurements. This also implies a turnstile streaming algorithm with  $\tilde{O}(n^{6/5}\epsilon^{-4})$  space. Key to our algorithm is a novel analysis of how the effective resistances change under *vertex sampling*. Complementing our algorithm, we then prove a superlinear lower bound of  $\Omega(n^{21/20-o(1)})$  measurements for computing some  $O(1)$ -spectral sparsifier using incidence sketches.
3. **Weighted spanner computation:** We first show that any  $o(n^2)$  linear measurements can only recover a spanner of stretch that in general depends *linearly* on  $\frac{w_{\max}}{w_{\min}}$ . We thus focus on graphs with  $\frac{w_{\max}}{w_{\min}} = O(1)$  and study the stretch’s dependence on  $n$ . On such graphs, the algorithm in [Filtser-Kapralov-Nouri, SODA’21] can obtain a spanner of stretch  $\tilde{O}(n^{\frac{2}{3}(1-\alpha)})$  using  $\tilde{O}(n^{1+\alpha})$  measurements for any  $\alpha \in [0, 1]$ . We prove that, for incidence sketches, this tradeoff is *optimal up to an  $n^{o(1)}$  factor for all  $\alpha < 1/10$* .

We prove both our lower bounds by analyzing the “effective resistances” in certain *matrix-weighted* graphs, where we develop a number of new tools for reasoning about such graphs – most notably (i) a matrix-weighted analog of the widely used *expander decomposition* of ordinary graphs, and (ii) a proof that a random *vertex-induced subgraph* of a matrix-weighted expander is also an expander. We believe these tools are of independent interest.

---

\*This work was supported in part by NSF awards CCF-1763514, CCF-1934876, and CCF-2008304.

<sup>†</sup>University of Pennsylvania. Email: {chenyu2, sanjeev, huanli}@cis.upenn.edu

<sup>1</sup> $\tilde{O}$  hides polylog( $n, \epsilon^{-1}, \frac{w_{\max}}{w_{\min}}$ ) factors, where  $w_{\max}$  and  $w_{\min}$  are the largest and smallest edge weights.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Our results . . . . .	2
1.2	Roadmap . . . . .	4
<b>2</b>	<b>Overview of weighted cut and spectral sparsification algorithms</b>	<b>5</b>
2.1	Overview of the algorithm for weighted cut sparsification . . . . .	5
2.2	Overview of the algorithm for weighted spectral sparsification . . . . .	6
<b>3</b>	<b>Overview of lower bound for weighted spectral sparsification</b>	<b>8</b>
3.1	The hard distribution . . . . .	9
3.2	A bound on the success probability via effective resistance . . . . .	10
3.3	Warm-up: one-row signed sketches have small TV-distance . . . . .	12
3.4	The general case: proof of Theorem 1.5 . . . . .	14
3.4.1	Techniques for proving Theorems 3.11, 3.13, 3.14 . . . . .	17
<b>4</b>	<b>Preliminaries</b>	<b>19</b>
4.1	Matrices . . . . .	19
4.2	Multivariate Gaussian distributions . . . . .	19
4.3	$\ell_2$ -heavy hitter, $\ell_p$ -sampler and $\ell_2$ estimation . . . . .	20
4.4	Edge strengths and cut sparsifiers . . . . .	21
4.5	Graph matrices, leverage scores, and spectral sparsifiers . . . . .	21
<b>5</b>	<b>A linear sketching algorithm for weighted cut sparsification</b>	<b>23</b>
5.1	A model oblivious algorithm for weighted cut sparsification . . . . .	23
5.2	Implementation by linear sketching . . . . .	24
5.3	Proof of Lemma 5.3 . . . . .	26
<b>6</b>	<b>A linear sketching algorithm for weighted spectral sparsification</b>	<b>28</b>
6.1	A vertex sampling lemma . . . . .	28
6.2	Recovery of heavy edges . . . . .	31
6.3	Main algorithm for weighted spectral sparsification . . . . .	34
6.4	Sparsification of $G^{\text{sq}}$ . . . . .	41
6.4.1	Sparsification of $G^{\text{sq}}$ by heavy edge recovery . . . . .	41
6.4.2	Recovery of heavy edges in $G^{\text{sq}}$ . . . . .	47
<b>7</b>	<b>Preliminaries on matrix-weighted graphs</b>	<b>50</b>
<b>8</b>	<b>Almost regular graphs have only few small eigenvalues</b>	<b>52</b>
<b>9</b>	<b>Almost regular graph decomposition</b>	<b>56</b>
<b>10</b>	<b>Almost regular expander decomposition</b>	<b>64</b>
10.1	Analysis of Algorithm 3 . . . . .	68
10.2	Analysis of Algorithm 4 . . . . .	74

<b>11 Expanders are preserved under vertex sampling</b>	<b>77</b>
11.1 Warm-up: ordinary expanders are preserved under vertex sampling . . . . .	77
11.2 Matrix-weighted expanders are preserved under vertex sampling . . . . .	87
<b>12 A lower bound for weighted spectral sparsification</b>	<b>96</b>
<b>13 A lower bound for weighted spanner computation</b>	<b>99</b>
<b>A Proof of Proposition 1.6</b>	<b>103</b>
<b>B A hard instance for decomposing matrix-weighted graphs into large subgraphs without small nonzero eigenvalues</b>	<b>105</b>
<b>C Missing proofs from Section 3</b>	<b>106</b>
<b>D Missing proofs from Section 7</b>	<b>110</b>
<b>E Missing proofs from Section 10</b>	<b>111</b>
<b>F Missing proofs from Section 11</b>	<b>114</b>

# 1 Introduction

Graph sparsification is a process that reduces the number of edges in a dense graph significantly while preserving certain useful properties. Besides being an interesting problem in its own right, graph sparsification has also been used as a fundamental building block in many modern graph algorithms such as maximum flow and minimum cut algorithms [BK15, She13, KLOS14, Pen16], solvers for graph structured linear systems [ST14, CKM<sup>+</sup>14, JS21], and graph clustering [CSWZ16].

In addition to designing fast algorithms in the classic computational model, a rich body of work has also studied graph sparsification by *linear sketching*. In this setting, we can only access the input graph by taking *linear measurements*, each of which returns a linear combination of the edge weights, and the goal is then to compute a sparsifier of the input graph using as few measurements as possible. To state the previously known results in this setting, let us first recall the definitions of three extensively studied graph sparsifiers that we will study in this work.

**Definition 1.1** (Cut sparsifiers). Given a weighted graph  $G = (V, E, w)$  and a parameter  $\epsilon \in (0, 1)$ , another weighted graph  $H = (V, F, w')$  with  $F \subseteq E$  is called a  $(1 + \epsilon)$ -cut sparsifier of  $G$  if for every cut  $(S, V - S)$ , its weight  $w_G(S, V - S)$  in  $G$  and its weight  $w_H(S, V - S)$  in  $H$  satisfy that  $(1 - \epsilon)w_G(S, V - S) \leq w_H(S, V - S) \leq (1 + \epsilon)w_G(S, V - S)$ .

**Definition 1.2** (Spectral sparsifiers). Given a weighted graph  $G = (V, E, w)$  with Laplacian matrix  $L_G$  and a parameter  $\epsilon \in (0, 1)$ , another weighted graph  $H = (V, F, w')$  with Laplacian matrix  $L_H$  and  $F \subseteq E$  is called a  $(1 + \epsilon)$ -spectral sparsifier of  $G$  if for every vector  $x \in \mathbb{R}^n$  we have  $(1 - \epsilon)x^T L_G x \leq x^T L_H x \leq (1 + \epsilon)x^T L_G x$ .

**Definition 1.3** (Spanners). Given a weighted graph  $G = (V, E, w)$  and a parameter  $t \geq 1$  (called the *stretch*), another weighted graph  $H = (V, F, w')$  with  $F \subseteq E$  is called a  $t$ -spanner of  $G$  if for every vertex pair  $u, v$ , the shortest path length  $d_G(u, v)$ <sup>2</sup> between  $u, v$  in  $G$  and the shortest path length  $d_H(u, v)$  in  $H$  satisfy  $d_G(u, v) \leq d_H(u, v) \leq t \cdot d_G(u, v)$ .

A seminal work by Ahn, Guha, and McGregor [AGM12b] showed that one can compute a  $(1 + \epsilon)$ -cut sparsifier of an unweighted graph using  $\tilde{O}(n\epsilon^{-2})$  linear measurements, which is nearly optimal. Then subsequent works by Kapralov, Lee, Musco, Musco, and Sidford [KLM<sup>+</sup>14] and Kapralov, Mousavifar, Musco, Musco, Nouri, Sidford, and Tardos [KMM<sup>+</sup>20] showed that one can also compute a  $(1 + \epsilon)$ -spectral sparsifier of an unweighted graph using  $\tilde{O}(n\epsilon^{-2})$  linear measurements. Finally, a recent work by Filtser, Kapralov, and Nouri [FKN21] showed that one can compute an  $\tilde{O}(n^{2/3})$ -spanner of an unweighted graph using  $\tilde{O}(n)$  linear measurements. [FKN21] also showed that one can compute an  $\tilde{O}(n^{2/3(1-\alpha)})$ -spanner of an unweighted graph using  $\tilde{O}(n^{1+\alpha})$  linear measurements for any  $\alpha \in [0, 1]$ , and conjectured that this tradeoff might be close to optimal.

In all of these works [AGM12b, KLM<sup>+</sup>14, KMM<sup>+</sup>20, FKN21], the authors also showed that their linear sketching algorithms can also be applied to computing graph sparsifiers of weighted graphs in *dynamic streams*, where the input graph is given by a stream of *insertions and deletions of weighted edges*, and the goal is to compute a sparsifier of the input graph using a small amount of space. In all these works, this is achieved by grouping the edge weights geometrically, and

---

<sup>2</sup>Note that in a weighted graph, the length of a path is defined as the total edge weight along the path.

then applying the linear sketching algorithm for *unweighted graphs* to the subgraph induced by edges in each weight group. Note that this approach crucially requires that if an edge  $e$  is inserted with weight  $w_e$ , then any subsequent deletion of the edge  $e$  again reveals its weight and it must be *identical* to  $w_e$ . This is crucial to ensuring that each edge  $e$  is inserted into or deleted from the same geometric group.

However, the operation of grouping edges by weight is not linear, and as a result, the above approach for extending unweighted sketches to weighted ones is not implementable in the more general *turnstile streams*, where the graph is given as a stream of *arbitrary edge weight updates*. Surprisingly, little seems to be known about graph sparsification in this setting.

## 1.1 Our results

In this paper, we initiate the study of *weighted* graph sparsification by linear sketching. To state our results, we need to introduce some notation first. Let  $w \in \mathbb{R}_{\geq 0}^{n \choose 2}$  denote the weights of the edges of the input graph, where  $w_e = 0$  means that there is no edge in the edge slot  $e$ . A *linear sketch* of  $N = N(n)$  measurements consists of a (random) *sketching matrix*  $\Phi \in \mathbb{R}^{N \times {n \choose 2}}$ , and a (randomized) *recovery algorithm*  $\mathcal{A}$  that takes as input  $\Phi w \in \mathbb{R}^N$  and outputs a sparsifier of the graph with edge weights  $w$ . Note that, by definition, the linear sketch is *non-adaptive*.

We focus on a natural class of linear sketches that we call *incidence sketches*, in which each row of the sketching matrix  $\Phi$  is supported<sup>3</sup> on edges incident on a *single* vertex (which could be different for different rows). This class captures all linear sketches that are implementable in a distributed computing setting, where the edges are stored across  $n$  machines such that machine  $i$  has all edges incident on the  $i^{\text{th}}$  vertex (a.k.a. *simultaneous communication model*). Moreover, it also covers *all* aforementioned linear sketches used in previous works for unweighted cut sparsification [AGM12b], spectral sparsification [KLM<sup>+</sup>17, KMM<sup>+</sup>20], and spanner computation [FKN21].

We now present our results for computing these three kinds of sparsifiers in weighted graphs. When describing our results, we use  $w_{\max}$  and  $w_{\min}$  to denote the largest and the smallest non-zero edge weights, respectively, and always assume  $w_{\max} \geq 1 \geq w_{\min}$ . We also write  $\tilde{O}(\cdot)$  to hide  $\text{polylog}(n, \epsilon^{-1}, \frac{w_{\max}}{w_{\min}})$  factors.

**Weighted cut sparsification.** We design an incidence sketch with a near-linear number of measurements for computing a  $(1 + \epsilon)$ -cut sparsifier of a weighted graph.

**Theorem 1.1** (Algorithm for weighted cut sparsification). *For any  $\epsilon \in (0, 1)$ , there exists an incidence sketch with random sketching matrix  $\Phi_1 \in \mathbb{R}^{N_1 \times {n \choose 2}}$  satisfying  $N_1 \leq \tilde{O}(n\epsilon^{-3})$  and a recovery algorithm  $\mathcal{A}_1$ , such that for any  $w \in \mathbb{R}_{\geq 0}^{n \choose 2}$ ,  $\mathcal{A}_1(\Phi_1 w)$  returns, with probability  $1 - \frac{1}{\text{poly}(n)}$ , a  $(1 + \epsilon)$ -cut sparsifier of the graph with edge weights  $w$ .*

Thus, we achieve a similar performance as the linear sketch for unweighted graphs given in [AGM12b], which uses  $O(n\epsilon^{-2}\text{poly}(\log n))$  measurements. It is well known that even to detect the connectivity of a graph,  $\Omega(n)$  linear measurements are needed. Therefore, our upper bound in Theorem 1.1 is nearly optimal in  $n$ .

---

<sup>3</sup>Recall that the support of a vector is the set of indices at which it is non-zero.

Similar to [Ind06, AGM12b, KLM<sup>+</sup>14, KMM<sup>+</sup>20, FKN21], by using Nisan’s well known pseudorandom number generator [Nis92], we can turn our linear sketching algorithm to a low space streaming algorithm.

**Corollary 1.2** (of Theorem 1.1). *There is a single pass turnstile streaming algorithm with  $\tilde{O}(n\epsilon^{-3})$  space that, at any given point of the stream, recovers a  $(1 + \epsilon)$ -cut sparsifier of the current graph with high probability.*

Note that in the turnstile model, the stream consists of arbitrary edge weight updates.

**Weighted spectral sparsification.** We design an incidence sketch with about  $n^{6/5}$  measurements for computing a  $(1 + \epsilon)$ -spectral sparsifier of a weighted graph.

**Theorem 1.3** (Algorithm for weighted spectral sparsification). *For any  $\epsilon \in (0, 1)$ , there exists an incidence sketch with random sketching matrix  $\Phi_2 \in \mathbb{R}^{N_2 \times \binom{n}{2}}$  satisfying  $N_2 \leq \tilde{O}(n^{6/5}\epsilon^{-4})$  and a recovery algorithm  $\mathcal{A}_2$ , such that for any  $w \in \mathbb{R}_{\geq 0}^{\binom{n}{2}}$ ,  $\mathcal{A}_2(\Phi_2 w)$  returns, with probability  $1 - \frac{1}{\text{poly}(n)}$ , a  $(1 + \epsilon)$ -spectral sparsifier of the graph with edge weights  $w$ .*

Similar to the cut sparsification case, we have the following corollary:

**Corollary 1.4** (of Theorem 1.3). *There is a single pass turnstile streaming algorithm with  $\tilde{O}(n^{6/5}\epsilon^{-4})$  space that, at any given point of the stream, recovers a  $(1 + \epsilon)$ -spectral sparsifier of the current graph with high probability.*

We complement this result by showing that a superlinear number of measurements are indeed necessary for any incidence sketch to recover some  $O(1)$ -spectral sparsifier.

**Theorem 1.5** (Lower bound for weighted spectral sparsification). *There exist constants  $\epsilon, \delta \in (0, 1)$  such that any incidence sketch of  $N$  measurements that computes a  $(1 + \epsilon)$ -spectral sparsifier with probability  $\geq 1 - \delta$  on any  $w$  must satisfy  $N \geq n^{21/20-o(1)}$ .*

Note that this is in sharp contrast to the unweighted case, where a near-linear number of incidence sketch measurements are sufficient for computing an  $O(1)$ -spectral sparsifier [KLM<sup>+</sup>17, KMM<sup>+</sup>20]. Theorem 1.5 also draws a distinction between spectral sparsification and cut sparsification, as for the latter a near-linear number of measurements are enough even in the weighted case (by Theorem 1.1).

**Weighted spanner computation.** We first show that any  $o(n^2)$  linear measurements can only recover a spanner of stretch that in general depends *linearly* on  $\frac{w_{\max}}{w_{\min}}$ . This differs fundamentally from the case of cut or spectral sparsification, where we can recover an  $O(1)$ -sparsifier, whose error is completely independent of  $\frac{w_{\max}}{w_{\min}}$ , using a sublinear-in- $n^2$  number of measurements. Specifically, we prove the following proposition, whose proof appears in Appendix A.

**Proposition 1.6.** *Any linear sketch (not necessarily an incidence sketch) of  $N$  measurements that computes an  $o(\frac{w_{\max}}{w_{\min}})$ -spanner with probability  $\geq .9$  on any  $w$  must satisfy  $N \geq \Omega(n^2)$ .*

This proposition is a consequence of that edge weights are *proportional* to the edge lengths. More specifically, consider a complete graph where we weight a uniformly random edge by  $w_{\min}$  and all other  $\binom{n}{2} - 1$  edges by  $w_{\max}$ . Then, while we can ignore the  $w_{\min}$ -weight edge in an  $O(1)$ -cut or spectral sparsifier, we have to include it in any  $o(\frac{w_{\max}}{w_{\min}})$ -spanner, since otherwise the shortest path length between its two endpoints would have been blown up by at least a factor of  $\frac{2w_{\max}}{w_{\min}}$ . Now note that, as the weight of this edge is smaller than all other edges, in order to find it we have to essentially recover all entries of the edge weight vector  $w$ , which inevitably requires  $\Omega(n^2)$  linear measurements<sup>4</sup>.

In light of the above proposition, we turn our focus to graphs with  $\frac{w_{\max}}{w_{\min}} = O(1)$ , and study the stretch's optimal dependence on  $n$ . On such graphs, the approach in [FKN21] is able to obtain the following tradeoff between the stretch of the spanner and the number of linear measurements needed.

**Theorem 1.7** (Algorithm for weighted spanner computation [FKN21]). *For any constant  $\alpha \in [0, 1]$ , there exists an incidence sketch with random sketching matrix  $\Phi_3 \in \mathbb{R}^{N_3 \times \binom{n}{2}}$  satisfying  $N_3 \leq \tilde{O}(n^{1+\alpha})$  and a recovery algorithm  $\mathcal{A}_3$ , such that for any  $w \in \mathbb{R}_{\geq 0}^{\binom{n}{2}}$  with  $\frac{w_{\max}}{w_{\min}} \leq O(1)$ ,  $\mathcal{A}_3(\Phi_3 w)$  returns, with probability  $1 - \frac{1}{\text{poly}(n)}$ , an  $\tilde{O}(n^{\frac{2}{3}(1-\alpha)})$ -spanner of the graph with edge weights  $w$ .*

[FKN21] also conjectured that on unweighted graphs, to obtain a spanner of stretch  $O(n^{\frac{2}{3}-\epsilon})$  for any constant  $\epsilon > 0$ , a superlinear number of measurements are needed for any linear sketch (in other words, the tradeoff is optimal at  $\alpha = 0$ ). We make progress on this question by showing that this is indeed true for a natural class of linear sketches (i.e. incidence sketches) on “almost” unweighted graphs (i.e. those with  $\frac{w_{\max}}{w_{\min}} = O(1)$ ). In fact, we show that in such a setting, the tradeoff obtained in the above theorem is *optimal for all  $\alpha < 1/10$* .

**Theorem 1.8** (Lower bound for weighted spanner computation). *For any constant  $\alpha \in (0, 1/10)$ , there exist constants  $C \geq 1, \delta \in (0, 1)$  such that any incidence sketch of  $N$  measurements that computes an  $o(n^{\frac{2}{3}(1-\alpha)})$ -spanner with probability  $\geq 1 - \delta$  on any  $w$  with  $\frac{w_{\max}}{w_{\min}} \leq C$  must satisfy  $N \geq n^{1+\alpha-o(1)}$ .*

## 1.2 Roadmap

The rest of this paper is structured as follows.

We start by giving an overview of the techniques used in proving our main results. Specifically, Section 2 gives an overview of our algorithms for weighted cut and spectral sparsification. Then in Section 3, we give an overview of our lower bound for weighted spectral sparsification. Note that we do *not* give a separate overview of our lower bound for weighted spanner computation, because the ideas are similar to the ones described in Section 3.

Section 4 contains some preliminaries that we will rely on throughout. Then Sections 5 and 6 present in detail our algorithms and their analysis for weighted cut and spectral sparsification, respectively.

---

<sup>4</sup>In our actual proof of the proposition, we have to add some random Gaussian noise to each edge's weight for the lower bound to carry out.

The remaining Sections 7–13 are devoted to proving our lower bounds for weighted spectral sparsification and spanner computation, both of which rely on a number of new tools for analyzing certain *matrix-weighted* graphs. Section 7 first defines such matrix-weighted graphs and sets up some notation. Then Sections 8, 9, 10, 11 present the new tools we develop for analyzing them. Finally, Sections 12, 13 prove our lower bounds for weighted spectral sparsification and spanner computation, respectively.

## 2 Overview of weighted cut and spectral sparsification algorithms

### 2.1 Overview of the algorithm for weighted cut sparsification

**Recap of the unweighted cut sparsification algorithm in [AGM12b].** At a high-level, the approach taken is to reduce cut sparsification to (repeatedly) recovering a spanning forest of a subgraph of the input graph, obtained by sampling edges *uniformly* at some rate  $p \in (0, 1)$  known beforehand. This task is then further reduced to the task of sampling an edge connecting  $S$  to  $\bar{S}$  for an arbitrary subset  $S$  of vertices, as this can be used to create a spanning forest by growing connected components.

Now to implement this latter task, in the sketching phase, we apply an  $\ell_0$ -sampler sketch to the incidence vector of each vertex  $u$  (i.e. each column of the edge-vertex incidence matrix) in the *sub-sampled graph*. Then in the recovery phase, in order to recover an edge going out of a vertex set  $S$ , we add up the sketches of the vertices inside  $S$ . By linearity, this summed sketch is taken over the sum of the incidence vectors of vertices inside  $S$ , and the latter contains exactly the edges going out of  $S$ , since the edges inside cancel out. As a result, we can recover an edge going out of  $S$ , and create a spanning forest of the sub-sampled graph. Note that this approach crucially utilizes the fact that the edges are sampled uniformly at a rate that is known *beforehand*. This means that we can sample all  $\binom{n}{2}$  edge slots beforehand, and apply the linear sketch only to the sampled edge slots.

**Our approach for weighted cut sparsification.** We also reduce the task to recovering a spanning forest in a sub-sampled graph. However, the latter graph is now obtained by sampling edges *non-uniformly*. Specifically, we need to recover a spanning forest in a subgraph obtained by sampling each edge  $e$  with probability  $\min\{w_{ep}, 1\}$  for some parameter  $p \in (0, 1)$  that is known beforehand. Therefore, in order to apply the idea as in the unweighted case, we will now need to design a variant of  $\ell_0$ -sampler that, given a vector  $x \in \mathbb{R}_{\geq 0}^N$  and a parameter  $p \in (0, 1)$ , recovers a nonzero entry of  $x$  after each entry  $i \in [N]$  is sampled with probability  $\min\{x_ip, 1\}$ . We call such a sampler “weighted edge sampler”.

Note that the edge weights are *not* known to us beforehand, so we cannot sample the edge slots with our desired probabilities as in the unweighted graph case. We instead build such a weighted edge sampler using a *rejection sampling* process, in which we sample edges uniformly at  $\tilde{O}(1)$  geometric rates, but use  $\ell_1$ -samplers to try recovering edges at each rate, and only output a recovered edge  $e$  if the sampling rate  $\approx w_{ep}$ . We then show that with high probability, we can efficiently find a desired edge.

Roughly, our analysis involves proving that there exists a geometric rate  $q$  such that, after uniformly sampling edges at rate  $q$ , the total weight of edges  $e$  satisfying  $w_{ep} \approx q$  accounts for a

large portion of that of all sampled edges. As a result, by using a few independent  $\ell_1$ -samplers, we can find one such edge with high probability.

## 2.2 Overview of the algorithm for weighted spectral sparsification

As in previous linear sketches for unweighted graphs [KLM<sup>+</sup>14, KMM<sup>+</sup>20], the key task is to recover edges with  $\tilde{\Omega}(1)$  effective resistances (or in weighted case,  $\tilde{\Omega}(1)$ -leverage scores), which we refer to as *heavy edges*. The high-level idea used in previous works is to (i) compute, for each vertex pair  $s, t$ , a set of vertex potentials  $x_{s,t} \in \mathbb{R}^n$  induced by an electrical flow from  $s$  to  $t$ , and then (ii) apply an  $\ell_2$ -heavy hitter to  $B_G x_{s,t} \in \mathbb{R}^{\binom{n}{2}}$  to try recovering the edge  $(s, t)$ , where  $B_G$  is the edge-vertex incidence matrix of  $G$  (see Section 4.5). They achieve (i) by simulating an iterative refinement process in [LMP13]. To achieve (ii), they make a key observation that

$$\|B_G x_{s,t}\|_2^2 = x_{s,t}^T B_G^T B_G x_{s,t} = x_{s,t}^T L_G x_{s,t}$$

is the energy of  $x_{s,t}$ , and the entry of  $B_G x_{s,t}$  indexed by edge  $(s, t)$  is  $(B_G x_{s,t})_{(s,t)} = b_{s,t}^T x = x_s - x_t$ . Therefore by the energy minimization characterization of effective resistances (Fact 4.17), whenever the effective resistance between  $s, t$  is  $b_{s,t}^T L_G^\dagger b_{s,t} \geq \tilde{\Omega}(1)$ , we have

$$(B_G x_{s,t})_{(s,t)}^2 \geq \tilde{\Omega}(1) \|B_G x_{s,t}\|_2^2,$$

and hence the entry  $(s, t)$  is an  $\ell_2$ -heavy hitter.

However, when the graph is weighted, we are only allowed to access the graph through linear measurements on its weight vector  $w_G$ . As a result, we can only apply  $\ell_2$ -heavy hitters to  $W_G B_G x_{s,t}$ , whose squared  $\ell_2$ -norm is  $x_{s,t}^T B_G^T W_G^2 B_G x_{s,t}$ . Now notice that  $B_G^T W_G^2 B_G$  is the Laplacian matrix of a “squared” graph (call it  $G^{\text{sq}}$ ), which has the same edges as  $G$ , but whose edges are weighted by  $w_e^2$  as opposed to  $w_e$ . Therefore, we will be recovering edges that are heavy in  $G^{\text{sq}}$  instead of in  $G$  if we apply the same approach as in previous works. Unfortunately, a heavy edge in  $G$  is not in general heavy in  $G^{\text{sq}}$ , since the energy on the edges with very large weights will blow up when we square the edge weights (i.e.  $w_e^2(x_u - x_v)^2 \gg w_e(x_u - x_v)^2$ ), and hence make the total energy grow unboundedly.

To see an intuitive example, suppose  $G$  is a “block cycle graph” on  $n$  vertices whose edges are generated as follows (see also Figure 1):

1. Partition the vertices into  $n^{4/5}$  blocks  $S_0, \dots, S_{n^{4/5}-1}$ , each with  $n^{1/5}$  vertices.
2. For each  $0 \leq i < n^{4/5}$ , add on  $S_i$  a complete graph of  $n^{1/5}$  vertices with edge weights  $n^{2/5}$ , i.e.  $n^{2/5} K_{n^{1/5}}$ .
3. For each  $0 \leq i < n^{4/5}$ , add on  $(S_i, S_{i+1})$  a complete bipartite graph of  $2n^{1/5}$  vertices with edge weights  $n^{2/5}$  and bipartition  $(S_i, S_{i+1})^5$ , i.e.  $n^{2/5} K_{n^{1/5}, n^{1/5}}$ .
4. Finally, add a “crossing edge”  $e^*$  of weight 1 between a randomly chosen vertex pair  $s, t$ .

We note that, in this construction, the crossing edge  $e^*$  spans  $\Omega(n^{4/5})$  consecutive blocks. Note that, typically, the crossing edge  $e^*$  spans  $\Omega(n^{4/5})$  consecutive blocks, and therefore has effective resistance (and also leverage score)  $\Omega(1)$ .

---

<sup>5</sup>Note that we consider  $i + 1$  as 0 when  $i = n^{4/5} - 1$ .

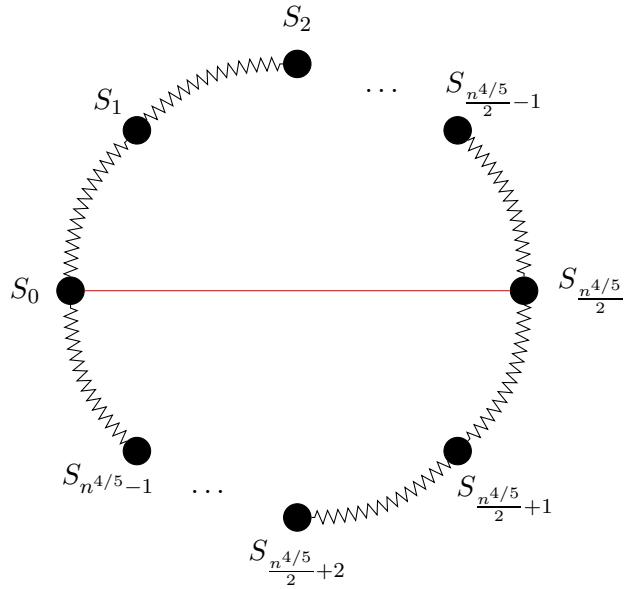


Figure 1: A block cycle graph on  $n$  vertices. Each  $S_i$  represents a block of  $n^{1/5}$  vertices connected by a clique and each zigzag represents the edges of a complete bipartite graph between adjacent blocks. The red edge represents the crossing edge. All edges along the cycle have weights  $n^{2/5}$ , and the crossing edge has weight 1.

**Proposition 2.1.** *If  $s \in S_i, t \in S_j$  such that  $\min \{|i - j|, n^{4/5} - |i - j|\} \geq \Omega(n^{4/5})$ , then the effective resistance of  $e^*$  satisfies  $r_{e^*} \geq \Omega(1)$ .*

*Proof.* Let  $s, t$  be the endpoints of the crossing edge  $e^*$ . By the energy minimization characterization of effective resistances (Fact 4.17), it suffices to show that there is a set of vertex potentials whose normalized energy with respect to  $s, t$  is  $O(1)$ . Specifically, consider the set of potentials  $x \in \mathbb{R}^n$  such that  $x_u = \frac{i}{n^{4/5}}$  for all  $u \in S_i$ . Then we have  $x_s - x_t = \Theta(1)$ , and the total energy is  $\sum_{e=(u,v)} w_e (x_u - x_v)^2 = n^{6/5} \cdot n^{2/5} (\Theta(n^{-4/5}))^2 + \Theta(1) = \Theta(1)$ , as desired.  $\square$

However, in the squared graph  $G^{\text{sq}}$ , all edge weights along the cycle are blown up by a factor of  $n^{2/5}$ , and thus  $e^*$  only has leverage score  $O(n^{-2/5})$  in  $G^{\text{sq}}$ . To recover in a vector  $x$  every entry with  $\ell_2$ -contribution  $\geq n^{-2/5} \|x\|_2^2$ , one will need an  $\Omega(n^{-2/5})$  factor blowup in the number of linear measurements, resulting in a total of  $n^{7/5}$  measurements needed to recover  $e^*$ .

We can in fact improve the number of linear measurements needed for recovering  $e^*$  to  $\tilde{O}(n^{6/5})$  using the *vertex sampling* trick, an idea first used in [FKN21] for sketching spanners. Namely, consider sampling a vertex set  $C \subset V$  by including each vertex with probability  $n^{-1/5}/100$ , and looking at the vertex-induced subgraph  $G^{\text{sq}}[C]$ . Then one can show that, conditioned on  $e^* \in G^{\text{sq}}[C]$ , with constant probability, the two endpoints of  $e^*$  will be disconnected in  $G^{\text{sq}}[C] \setminus e^*$ . As a result, the leverage score of  $e^*$  becomes 1 in  $G^{\text{sq}}[C]$ , and we can recover  $e^*$  by recovering heavy edges in  $G^{\text{sq}}[C]$ , which, as will show, can be done using  $\tilde{O}(|C|) \approx \tilde{O}(n^{4/5})$  measurements. Since  $e^* \in G[C]$  with probability  $\approx \frac{1}{n^{2/5}}$ , repeating this sampling process independently for  $\tilde{O}(n^{2/5})$  times allows us to recover  $e^*$  in at least one vertex induced subgraph. This results in a linear sketch of  $\tilde{O}(n^{6/5})$  measurements.

What if we slightly increase each block's size to  $n^{1/5+\delta}$  and decrease the edge weights along the cycle to  $n^{2/5-3\delta}$ ? While one can still verify that the crossing edge  $e^*$  has leverage score  $\Omega(1)$ , applying the same vertex sampling process as above will not disconnect the endpoints of  $e^*$  with  $\Omega(1)$  probability. However, one can alternatively show that, with constant probability, the number of edges along the cycle reduces by a factor of  $n^{2/5}$ . Since now the energy of each edge only blows up by a factor smaller than  $n^{2/5}$  in  $G^{\text{sq}}$ , this will also make the leverage score of  $e^*$  become  $\Omega(1)$  in  $G^{\text{sq}}[C]$ , and thus we can apply the same linear sketch of  $\tilde{O}(n^{6/5})$  measurements.

The above warm-up seems to suggest that the sampling rate of  $\approx n^{-1/5}$  is a sweet spot for recovering heavy edges in any graphs with the block cycle structure. Indeed, we prove a key vertex sampling lemma showing that in *any* weighted graph  $G$ , a heavy edge  $e$  in  $G$  is also likely heavy in a vertex-induced subgraph of  $G^{\text{sq}}$  obtained by sampling vertices at rate  $\approx n^{-1/5}$ . This is proved by carefully analyzing the structures of the edges of different weights after vertex sampling, and then explicitly constructing a set of vertex potentials with small total energy in the induced subgraph. Finally, by integrating this lemma into an iterative refinement process in [LMP13] (as the authors did in [KLM<sup>+</sup>14, KMM<sup>+</sup>20]) and a spectral sparsification algorithm in [Kou14], we are able to recover a spectral sparsifier of  $G$  using  $\tilde{O}(n^{6/5})$  linear measurements. We note that the latter step of using heavy edge recovery to build a spectral sparsifier is also more involved than in the unweighted case and requires a few extra techniques; we refer the reader to the overview at the beginning of Section 6.3 and the discussion therein for more details.

We note that this method of recovering heavy edges by vertex sampling is inspired by the one used in [KLM<sup>+</sup>14] for spanners. However, for spectral sparsification, the correctness of such a method follows from fairly different reasoning, and the proof is arguably more involved.

### 3 Overview of lower bound for weighted spectral sparsification

In this section we give an overview of our lower bound for weighted spectral sparsification (Theorem 1.5). We prove our lower bound on a family of hard instances that turn out to have the exact same structure as the one in Figure 1, which we used to illustrate the difficulty of recovering spectral sparsifiers in weighted graphs. Specifically, our hard instances are weighted “block cycle graphs” plus an extra *crossing edge* that is included with probability 1/2. In a block cycle graph, the vertices are partitioned into blocks that are arranged in a cyclic manner. Each block is a complete graph, and the vertices of adjacent blocks are connected by a complete bipartite graph. Here we draw all edge weights from Gaussian distributions and permute the vertices uniformly at random. On such graphs, for a suitable choice of edge weights, computing a spectral sparsifier essentially boils down to detecting the presence/absence of the crossing edge. We then show that for the latter task, the success probability of any incidence sketch can be bounded by the “effective resistance” of the crossing edge in a *matrix-weighted* graph, where the matrix weights are in turn determined by the sketching matrix.

We note that this family of hard instances has a similar structure to the ones conjectured in [FKN21]. However, instead of using Bernoulli distributions on the edges as suggested in [FKN21], we use Gaussian distributions, which makes it easier to build the connection to effective resistances.

In order to show that the effective resistance is small for any incidence sketch with a limited number of measurements, we develop a number of new tools for analyzing such matrix-weighted graphs. Most importantly, we present (i) a matrix-weighted analog of the widely used expander

decomposition of ordinary graphs [GR99, KVV04, ST04], and (ii) a proof that a random vertex-induced subgraph of a matrix-weighted expander is also an expander with high probability. We highly recommend reading Section 3.3 to get intuition on why these two techniques are useful, where we use the ordinary graph version of (i) and (ii) to prove a lower bound for a simple class of sketches.

The rest of this section is structured as follows. In Section 3.1 we describe the distribution from which we generate our hard instances. In Section 3.2 we explain how we bound the success probability of an incidence sketch by the effective resistance in a matrix-weighted graph. In Section 3.3 we prove, as a warm-up, a lower bound for a simple class of sketches, where we only need to analyze the effective resistances in *ordinary graphs*. Finally in Section 3.4 we outline our proof for arbitrary incidence sketches, which requires analyzing matrix-weighted graphs. We note that some proofs in this section are deferred to Appendix C.

### 3.1 The hard distribution

We first state how we generate the input weighted graph  $G = (V, E, w)$ . Let  $n$  be the number of vertices and define  $s \stackrel{\text{def}}{=} n^{1/5}$  and  $\ell \stackrel{\text{def}}{=} n^{4/5}$ . We choose a random permutation  $\pi : 1..n \rightarrow 1..n$  and construct a block cycle graph as follows. The  $i$ -th block (where  $0 \leq i < \ell$ ) consists of vertices  $\pi(si+1), \dots, \pi(si+s)$ . For simplicity we denote the  $a$ -th vertex in the  $i$ -th block (i.e.  $\pi(si+a)$ ) as  $u_{i,a}$ . The block index  $i$  will always be modulo  $\ell$  implicitly. We then add a complete graph to each block, and a complete bipartite graph between each pair of adjacent blocks. Namely, for each  $0 \leq i < \ell$ , we add a graph  $G_i$  with edges connecting  $u_{i,a}, u_{i,b}$  for all  $a < b \in \{1, \dots, s\}$ , and add another bipartite graph  $G_{i,i+1}$  with edges connecting  $u_{i,a}, u_{i+1,b}$  for all  $a, b \in \{1, \dots, s\}$ . Finally, with probability  $1/2$ , we add an edge between vertices  $\pi(1)$  and  $\pi(n/2+1)$  (assume  $n$  is even). We refer to this edge as the *crossing edge* with respect to  $\pi$  and any other edge in  $G$  as a *non-crossing edge* with respect to  $\pi$ . We will omit “with respect to  $\pi$ ” when the underlying permutation  $\pi$  is clear.

We next describe how the edge weights are determined. The weights of all non-crossing edges are drawn independently from  $\mathcal{N}(8n^{2/5}, n^{4/5} \log^{-1} n)$  (the Gaussian distribution with mean  $8n^{2/5}$  and variance  $n^{4/5} \log^{-1} n$ ). The weight of the crossing edge is drawn from the standard Gaussian  $\mathcal{N}(0, 1)$ . If the crossing edge has negative weight, we say the input is *invalid*, and accept any sketch as a valid sketch. Our goal will be to detect the presence/absence of the crossing edge with high probability.

In the following, we will call the conditional distribution on the presence of the crossing edge the *Yes distribution*, and call the conditional distribution on the absence of the crossing edge the *No distribution*. We then show that with high probability, the effective resistance of the crossing edge is large, and therefore any linear sketch for computing spectral sparsifiers must distinguish between the two distributions with good probability.

**Proposition 3.1.** *With probability at least  $1 - 1/n$ , all non-crossing edges have weights in the range  $[4n^{2/5}, 12n^{2/5}]$ , and as a result the effective resistance between vertices  $\pi(1)$  and  $\pi(n/2+1)$  is at least  $1/48$  in a No instance.*

**Proposition 3.2.** *Any linear sketch that can compute a 1.0001-spectral sparsifier with probability 0.9 can distinguish between the Yes and No distributions with probability 0.6.*

The first proposition follows from an application of the Chernoff bound (Theorem 4.7). The proof of the second proposition is deferred to Appendix C.

In the following, we will assume, for ease of our analysis, that the sketch will be given the permutation  $\pi$  *after* computing the linear sketch. That is, the recovery algorithm  $\mathcal{A}$  takes as input both  $\Phi w$  and  $\pi$ . We will show that even with this extra piece of information, any incidence sketch with  $n^{21/20-\epsilon}$  measurements for constant  $\epsilon > 0$  cannot distinguish between the Yes and No distributions with high probability.

### 3.2 A bound on the success probability via effective resistance

We first show that for our lower bound instance, any incidence sketch can be reduced to a more restricted class of linear sketches by only increasing the number of measurements by an  $O(\log n)$  factor. Specifically, let us fix an arbitrary orientation of the edges, and consider sketches taken over the weighted signed edge-vertex incidence matrix  $B^w \in \mathbb{R}^{\binom{n}{2} \times n}$ , where the latter is given by

$$B_{eu}^w = \begin{cases} w_e & e \in E \text{ and } u \text{ is } e\text{'s head} \\ -w_e & e \in E \text{ and } u \text{ is } e\text{'s tail} \\ 0 & \text{otherwise.} \end{cases}$$

That is, the algorithm must choose a (random) sketching matrix  $\Phi \in \mathbb{R}^{k \times \binom{n}{2}}$  with the  $e^{\text{th}}$  column  $\phi_e \in \mathbb{R}^k$  corresponding to the edge slot  $e$ . The sketch obtained is then  $\Phi B^w \in \mathbb{R}^{k \times n}$ . Notice that the total number of measurements in  $\Phi B^w$  is  $kn$ , as each vertex applies the sketching matrix  $\Phi$  to its incident edges. Let us call this class of sketch *signed sketches*. By Yao's minimax principle [Yao77], to prove a lower bound for distinguishing the Yes and No distributions, it suffices to focus on deterministic sketches. The proof of the proposition below appears in Appendix C.

**Proposition 3.3** (Reduction to signed sketches). *Consider any incidence sketch of  $N$  measurements with a deterministic sketching matrix  $\Phi \in \mathbb{R}^{N \times \binom{n}{2}}$  and a recovery algorithm  $\mathcal{A}$  that, given  $\Phi w$  and  $\pi$ , distinguishes between the Yes and No distributions with probability at least 0.6. Then there exists a signed sketch with a sketching matrix  $\Phi' \in \mathbb{R}^{k \times \binom{n}{2}}$ , where  $k = O(1) \cdot \max\{1, \frac{N \log n}{n}\}$ , and a recovery algorithm  $\mathcal{A}'$  that, given  $\Phi' B^w$  and  $\pi$ , distinguishes between the Yes and No distributions with probability at least 0.55.*

Let us now fix a sketching matrix  $\Phi \in \mathbb{R}^{k \times \binom{n}{2}}$  and aim to obtain an upper bound on the success probability of any signed sketch using  $\Phi$ . For notational convenience, let us write  $(\Phi B^w)_{\text{yes}}$  to denote  $\Phi B^w$  conditioned on the presence of the crossing edge and  $(\Phi B^w)_{\text{no}}$  to denote  $\Phi B^w$  conditioned on the absence of the crossing edge. We will also write  $(\Phi B^w)_{\pi, \text{yes}}$  or  $(\Phi B^w)_{\pi, \text{no}}$  to denote an extra conditioning on the permutation being  $\pi$  in addition to the presence/absence of the crossing edge. Then to bound the success probability of any signed sketch using  $\Phi$ , it suffices to show that the total variation distance (TV-distance) between  $(\Phi B^w)_{\text{yes}}$  and  $(\Phi B^w)_{\text{no}}$  is small.

To state our upper bound on the TV-distance, we need to first introduce some notation. For

an edge  $(u, v)$ , define  $b_{uv} \in \mathbb{R}^{nk}$  by writing it as a block vector (with block size  $k$ ) as follows:

$$b_{uv} = \begin{pmatrix} 0 \\ \vdots \\ \phi_{uv} \\ 0 \\ \vdots \\ -\phi_{uv} \\ 0 \\ \vdots \end{pmatrix} \begin{array}{l} u^{\text{th}} \text{ block} \\ \\ v^{\text{th}} \text{ block} \end{array} \in \mathbb{R}^{nk}, \quad (1)$$

where  $\phi_{uv} \in \mathbb{R}^k$  is the column of  $\Phi$  corresponding to the edge slot  $(u, v)$ . For a permutation  $\pi$ , we then define  $L_\pi = \sum_{\text{non-crossing } (u, v)} n^{4/5} \log^{-1} n b_{uv} b_{uv}^T$ . The following proposition is essentially a consequence of Theorem 4.6 [DMR18], which bounds the TV-distance between multivariate Gaussians with the same mean. We give its proof in Appendix C. Note that we use  $\dagger$  to denote taking the Moore-Penrose pseudoinverse of a matrix.

**Proposition 3.4.** *For any permutation  $\pi$  such that  $b_{\pi(1)\pi(n/2+1)}$  is in the range<sup>6</sup> of  $L_\pi$ ,*

$$d_{\text{TV}}((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}}) \leq O(1) \cdot \min \left\{ 1, b_{\pi(1)\pi(n/2+1)} L_\pi^\dagger b_{\pi(1)\pi(n/2+1)} \right\}.$$

Our plan is then to show that  $b_{\pi(1)\pi(n/2+1)} L_\pi^\dagger b_{\pi(1)\pi(n/2+1)}$  is small on average for every choice of a signed sketch with  $k = n^{1/20-\epsilon}$  for constant  $\epsilon > 0$ .

Note that if  $k = 1$  and each  $\phi_{uv} \in \{0, 1\}$ , then  $L_\pi$  is exactly the Laplacian matrix of the graph (call it  $\mathcal{H}_\pi$ ) that is formed by the non-crossing edges  $(u, v)$  such that  $\phi_{uv} = 1$ , where each edge is weighted by  $n^{4/5} \log^{-1} n$ . Therefore,  $b_{\pi(1)\pi(n/2+1)} L_\pi^\dagger b_{\pi(1)\pi(n/2+1)}$  is the effective resistance between  $\pi(1)$  and  $\pi(n/2 + 1)$  in  $\mathcal{H}_\pi$ , if  $\phi_{\pi(1)\pi(n/2+1)} \neq 0$  (otherwise  $b_{\pi(1)\pi(n/2+1)}$  is the zero vector).

In fact, to get a quick intuition as to why we should expect the effective resistance between  $\pi(1)$  and  $\pi(n/2 + 1)$  to be small, let us assume  $\phi_{uv} = 1$  for all edge slots  $(u, v)$ . Then, for any permutation  $\pi$ ,  $\mathcal{H}_\pi$  is the graph formed by *all* non-crossing edges, each weighted by  $n^{4/5} \log^{-1} n$ . Note that these weights are about  $n^{2/5}$  times larger than the weights  $\Theta(n^{2/5})$  in the actual input graph (Proposition 3.1). As a result, the effective resistance between  $\pi(1)$  and  $\pi(n/2 + 1)$  is about  $n^{2/5}$  times smaller than the effective resistance between them in the input graph (the former roughly equals  $n^{-2/5}$ ).

When  $k > 1$ , we can view  $L$  as the Laplacian of a *matrix-weighted* graph (again, call it  $\mathcal{H}_\pi$ ) formed by the non-crossing edges, where each edge  $(u, v)$  has a  $k \times k$  matrix weight  $n^{4/5} \log^{-1} n \phi_{uv} \phi_{uv}^T$ . Now  $b_{\pi(1)\pi(n/2+1)} L_\pi^\dagger b_{\pi(1)\pi(n/2+1)}$  can be seen as the (generalized) effective resistance between  $\pi(1)$  and  $\pi(n/2 + 1)$  in  $\mathcal{H}_\pi$ .

---

<sup>6</sup>Recall that the range of a symmetric matrix is the linear span of its columns.

### 3.3 Warm-up: one-row signed sketches have small TV-distance

As a warm-up, we show that for any signed sketch, in the case that  $k = 1$  and the sketching matrix  $\Phi$  has 0/1 entries, we have, for any constant  $\epsilon > 0$ ,

$$\mathbb{E}_\pi [d_{\text{TV}} ((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}})] \leq \frac{1}{n^{1/5-O(\epsilon)}}. \quad (2)$$

By Proposition 3.4, we know that  $d_{\text{TV}} ((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}})$  can be bounded by the effective resistance between  $\pi(1), \pi(n/2 + 1)$  in  $\mathcal{H}_\pi$  if  $\phi_{\pi(1)\pi(n/2+1)} = 1$ , and is zero otherwise. Here  $\mathcal{H}_\pi$  is formed by the non-crossing edges whose  $\phi_{uv} = 1$ , where each edge  $(u, v)$  has scalar weight  $n^{4/5} \log^{-1} n$ . We can focus on the  $\Phi$ 's whose number of nonzero entries is at least  $n^{9/5+\epsilon}$ , since otherwise

$$\Pr_\pi [\phi_{\pi(1)\pi(n/2+1)} = 1] = \frac{\text{nnz}(\Phi)}{\binom{n}{2}} \leq \frac{1}{n^{1/5-O(\epsilon)}},$$

and we would already have our desired result (2).

Our proof of (2) will rely on decomposing  $\mathcal{H}_\pi$  into expanders with large minimum degree. Since  $\mathcal{H}_\pi$ 's edges all have the same weight  $n^{4/5} \log^{-1} n$ , it is more convenient to work with the *unweighted version of  $\mathcal{H}_\pi$* , which we denote by  $H_\pi$ . We now briefly review the definition of unweighted expanders, as well as state a known expander decomposition lemma that we will utilize.

**Definition 3.1** (Expander). An unweighted graph  $H = (V, E)$  is a  $\zeta$ -expander for some  $\zeta \in [0, 1]$  if its conductance is at least  $\zeta$ , namely, for every nonempty  $S \subset V$ , we have

$$|E(S, V - S)| \geq \zeta \cdot \min \{ \text{vol}(S), \text{vol}(V - S) \},$$

where  $\text{vol}(S)$  is the total degree of vertices in  $S$ .

Note that in the lemma below, we slightly abuse the notion of “regular graphs”. Specifically, we will say a graph is regular if its minimum vertex degree  $d_{\min}$  is not much smaller than the average degree  $d$ .

**Lemma 3.5** (Almost regular expander decomposition, see e.g. [KKTY21]). *Given an unweighted graph  $H = (V, E)$  with average degree  $d \geq 16$ , there exists a subgraph  $I = (U, F)$  where  $U \subseteq V$  and  $F \subseteq E$  such that  $I$  is a  $\frac{1}{16 \log n}$ -expander with minimum degree  $d_{\min} \geq \frac{d}{16}$ .*

We will also need the following lemma, which shows that a random vertex-induced subgraph of an expander with large minimum degree is almost certainly an expander. We give the proof of this lemma in Section 11.1. To the best of our knowledge, even this result was not known before.

**Lemma 3.6** (Expanders are preserved under vertex sampling). *There exists a  $\theta = \theta(n) = n^{o(1)}$  with the following property. Consider an unweighted  $\frac{1}{16 \log n}$ -expander  $H = (V, E)$  with minimum degree  $d_{\min} \geq 4 \cdot 10^6 \cdot \theta(n)$ . For any  $s \geq \frac{4 \cdot 10^6}{d_{\min}} \cdot \theta(n) \cdot n$ , let  $C \subseteq V$  be a uniformly random vertex subset of size  $s$ . Then with probability at least  $1 - 1/n^7$ , the vertex-induced subgraph  $H[C]$  is a  $\frac{1}{n^{o(1)}}$ -expander with minimum degree at least  $\frac{s}{2n} \cdot d_{\min}$ .*

*Proof of (2) using Lemmas 3.5,3.6.* As argued above we can assume w.l.o.g. that  $\text{nnz}(\Phi) \geq n^{9/5+\epsilon}$ . We want to obtain, for each edge slot  $e$  satisfying  $\phi_e = 1$ , conditioned on  $e$  being the crossing edge w.r.t.  $\pi$ , an upper bound (call it  $u_e$ ) on the typical effective resistance between the endpoints of  $e$  in the graph  $\mathcal{H}_\pi$ . In other words, conditioned on  $e$  being the crossing edge,  $u_e$  should be an upper bound on the effective resistance between the endpoints of  $e$  in  $\mathcal{H}_\pi$  with high probability over  $\pi$ . Then the total variation distance between  $(\Phi B^w)_{\text{yes}}$  and  $(\Phi B^w)_{\text{no}}$  can be bounded by

$$\mathbb{E}_\pi [d_{\text{TV}} ((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}})] \leq O(1) \cdot \frac{1}{\binom{n}{2}} \sum_{e:\phi_e=1} u_e. \quad (3)$$

To obtain the  $u_e$ 's, let us define the unweighted graph  $H_\phi = (V, E_\phi)$  where  $E_\phi$  contains all edges  $e$  whose  $\phi_e = 1$  (including the ones not present in the input graph, i.e.  $|E_\phi| = \text{nnz}(\Phi)$ ). Now consider the following process, where we repeatedly delete an expander subgraph from  $H_\phi$  and obtain  $u_e$ 's for the edges in the expander.

1. While  $|E_\phi| \geq 10^9 n^{9/5+\epsilon}$ :

- (a) Find a subgraph  $I = (U, F)$  of  $H_\phi = (V, E_\phi)$  that is a  $\frac{1}{16 \log n}$ -expander with minimum degree  $d_{\min} \geq \frac{|E_\phi|}{8n}$  (existence is guaranteed by Lemma 3.5).
- (b) For each edge  $f \in F$ , let  $u_f \leftarrow \left( \frac{10^9 n^{9/5+\epsilon}}{|E_\phi|} \right)^2$ .
- (c) Delete the edges in  $F$  from  $H_\phi$  by letting  $E_\phi \leftarrow E_\phi \setminus F$ .

2. Let  $u_f \leftarrow 1$  for all  $f$  in the remaining  $E_\phi$ .

To show that  $u_e$ 's are valid upper bounds, let us consider a fixed iteration of the while loop. For  $i = 0, \dots, n^{4/5} - 1$ , let  $U_i$  denote the vertices in  $I$  that are in the  $i^{\text{th}}$  block of the input block cycle graph:

$$U_i \stackrel{\text{def}}{=} U \cap \left\{ \pi(n^{1/5}i + 1), \dots, \pi(n^{1/5}i + n^{1/5}) \right\}.$$

Then by Chernoff bounds, with probability at least  $1 - 1/n^5$  over the random choice of  $\pi$ , we have  $|U_i| \geq \frac{|U|}{2n^{4/5}} \geq \frac{4 \cdot 10^6}{d_{\min}} \cdot |U|^{1+\epsilon}$ . Then by invoking Lemma 3.6, with probability at least  $1 - 1/n^4$  over  $\pi$ , all vertex-induced subgraphs  $I[U_i \cup U_{i+1}]$  are  $\frac{1}{n^{o(1)}}$ -expanders with minimum degree at least  $\frac{|E_\phi|}{16n^{9/5}}$ . Using this fact, we obtain the following claim, whose proof appears in Appendix C.

**Claim 3.7.** For each edge  $f \in F$ , conditioned on  $f$  being the crossing edge, with probability at least  $1 - 1/n^2$  over  $\pi$ , the effective resistance between the endpoints of  $f$  in  $\mathcal{H}_\pi$  is at most  $u_f$ .

Now let us divide the above process for obtaining  $u_e$ 's into  $O(\log n)$  phases, where in phase  $i \in \{1, \dots, O(\log n)\}$ , we have  $|E_\phi| \in ((\binom{n}{2})/2^i, (\binom{n}{2})/2^{i-1}]$ . Then we have

$$\begin{aligned} \sum_{e:\phi_e=1} u_e &= \sum_{e:\phi_e=1} \left( \frac{10^9 n^{9/5+\epsilon}}{|E_\phi|} \right)^2 \leq n^{O(\epsilon)} \cdot \sum_{i=1}^{O(\log n)} \frac{\binom{n}{2}}{2^i} \cdot \left( \frac{2^i}{n^{1/5}} \right)^2 \\ &\leq n^{8/5+O(\epsilon)} \sum_1^{O(\log n)} 2^i \leq n^{9/5+O(\epsilon)} \end{aligned}$$

where in the first line we have used  $n^{O(\epsilon)}$  to hide the constant factors, and the last inequality holds since in the last phase we have  $2^i \leq n^{1/5}$ . Plugging this into (3) finishes the proof.  $\square$

### 3.4 The general case: proof of Theorem 1.5

Note that even though for  $k = 1$ , the TV-distance is  $\tilde{O}(n^{-1/5})$ , this does not imply that  $k$  must be large for the TV-distance to become  $\Omega(1)$ .

By Proposition 3.3, in order to prove Theorem 1.5, it suffices to prove the following:

**Theorem 3.8.** *For any fixed sketching matrix  $\Phi \in \mathbb{R}^{k \times \binom{n}{2}}$  where  $k \leq n^{1/20-\epsilon}$  for some constant  $\epsilon > 0$ , we have*

$$\mathbb{E}_\pi [d_{\text{TV}}((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}})] \leq o(1).$$

By Proposition 3.4, our goal is to bound the “effective resistance”  $b_{\pi(1)\pi(n/2+1)} L_\pi^\dagger b_{\pi(1)\pi(n/2+1)}$  between vertices  $\pi(1), \pi(n/2+1)$  in the matrix-weighted graph  $\mathcal{H}_\pi$  consisting of the non-crossing edges, where edge  $(u, v)$  has matrix weight  $n^{4/5} \log^{-1} n \phi_{uv} \phi_{uv}^T \in \mathbb{R}^{k \times k}$ . We will do so by (significantly) generalizing our previous approach based on expander decomposition for ordinary graphs in Section 3.3. Our approach for the  $k = 1$  case essentially consists of two steps: (i) decomposing the graph  $H_\phi$  into large expander subgraphs and (ii) proving that a random vertex induced subgraph of an expander is still an expander.

First note that there does not appear to be a combinatorial analog of conductance in matrix-weighted graphs, which suggests that we should define expanders in an algebraic way. Let us first recall the algebraic characterization of expanders for ordinary, unweighted graphs. The definition is based on eigenvalues of the *normalized Laplacian* of the graph, which is given by  $N = D^{-1/2} L D^{-1/2}$ , where  $D$  is a diagonal matrix with  $D_{uu}$  equal to the degree  $d_u$  of  $u$ .

**Definition 3.2** (Algebraic definition of ordinary, unweighted expanders). An unweighted graph  $H$  is a  $\zeta$ -expander for some  $\zeta \in [0, 1]$  if the smallest nonzero eigenvalue of its normalized Laplacian matrix  $N$  is at least  $\zeta$ .

By Cheeger’s inequality [AM85], for  $\zeta \geq \tilde{\Omega}(1)$ , this definition translates to that the graph  $H$  is a union of vertex-disjoint combinatorial expanders, each with conductance  $\tilde{\Omega}(1)$ . To come up with an analogous definition for matrix-weighted graphs, let us first define their associated matrices formally.

**Matrices associated with matrix-weighted graphs.** We consider a  $k \times k$  matrix-weighted graph  $H = (V, E)$  with  $|V| = n$ . For each edge  $(u, v) \in E$ , there is a vector  $\phi_{uv} \in \mathbb{R}^k$ , indicating that  $(u, v)$  is weighted by the  $k \times k$  rank-1 matrix  $\phi_{uv} \phi_{uv}^T$ .

**Definition 3.3** (Degree matrices). For a vertex  $u$ , its generalized degree is given by

$$D_u = \sum_{u \sim v} \phi_{uv} \phi_{uv}^T \in \mathbb{R}^{k \times k}.$$

We then define the  $nk \times nk$  degree matrix  $D$  as a block diagonal matrix (with block size  $k \times k$ ), with the  $u^{\text{th}}$  block on the diagonal being  $D_{uu} = D_u \in \mathbb{R}^{k \times k}$ .

**Definition 3.4** (Laplacian matrices). The Laplacian matrix is given by  $L = \sum_{(u,v) \in E} b_{uv} b_{uv}^T$ , where  $b_{uv}$ 's are defined in (1).

We will call  $b_{uv}$  the *incidence vector* of edge  $(u, v)$ . Note that the Laplacian matrix here differs from the connection Laplacian matrix [KLP<sup>+</sup>16] which is also defined to be a block matrix. In particular, the Laplacian matrix of a matrix-weighted graph is not necessarily block diagonally dominant (bDD) (Definition 1.1 of [KLP<sup>+</sup>16]).

**Definition 3.5** (Normalized Laplacian matrices). The normalized Laplacian matrix is given by  $N \stackrel{\text{def}}{=} D^{\dagger/2} L D^{\dagger/2}$ . Equivalently, we have  $N = \sum_{(u,v) \in E} D^{\dagger/2} b_{uv} b_{uv}^T D^{\dagger/2}$ .

We will call  $D^{\dagger/2} b_{uv}$  the *normalized incidence vector* of edge  $(u, v)$ .

The following proposition says that similar to scalar-weighted graphs, the eigenvalues of the normalized Laplacian of a matrix-weighted graph are also between  $[0, 2]$ . The proof this proposition appears in Appendix D (in “Proof of Proposition 7.7”).

**Proposition 3.9.** *The eigenvalues of  $N$  are between  $[0, 2]$ .*

Now, a first attempt might be to define matrix-weighted expanders to also be graphs whose normalized Laplacians' nonzero eigenvalues are large, and then try to decompose any matrix-weighted graph into large expander subgraphs. However, we show that the latter goal may not be achievable in general, by presenting in Appendix B a hard instance, for which any large subgraph has a small nonzero eigenvalue.

**Our approach.** In light of the hard instance, we loosen the requirement of being an expander by allowing small eigenvalues, but requiring instead that each edge, compared to the average, does not have too large “contribution” to the small eigenvectors. Formally, we want that every edge's normalized incidence vector has small (weighted) projection onto the bottom eigenspace. We will also need an analog of “almost regularity”, which for ordinary, unweighted graphs says that the minimum degree is large. We give the formal definition of an almost regular matrix-weighted expander below.

**Definition 3.6** (Almost regular matrix-weighted expanders). For a  $k \times k$  matrix-weighted graph  $H$ , let  $\lambda_1 \leq \dots \leq \lambda_{nk}$  be the eigenvalues of its normalized Laplacian  $N$ , and let  $f_1, \dots, f_{nk} \in \mathbb{R}^{nk}$  be a set of corresponding orthonormal eigenvectors. We say  $H$  is a  $(\gamma, \zeta, \psi)$ -almost regular expander if

1. ( $\gamma$ -almost regularity) For every vertex  $u$  and every incident edge  $(u, v) \in E$ , we have

$$\phi_{uv}^T D_u^\dagger \phi_{uv} \leq \frac{\gamma \cdot k}{n}. \quad (4)$$

2.  $((\zeta, \psi)$ -expander) For every edge  $(u, v) \in E$  we have

$$\left( D^{\dagger/2} b_{uv} \right)^T \left( \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} f_i f_i^T \right) D^{\dagger/2} b_{uv} \leq \frac{\psi \cdot k^2}{n^2}. \quad (5)$$

The LHS of (4) is the so-called *leverage score* of  $\phi_{uv}$  w.r.t.  $D_u$  (Definition 4.3). It is known that the sum of leverage scores equals the rank of the matrix:

**Proposition 3.10.** *For any fixed vertex  $u$ ,  $\sum_{(u,v) \in E} \phi_{uv}^T D_u^\dagger \phi_{uv} = \text{rank}(D_u)$ .*

Since  $D_u$  is a  $k \times k$  matrix, we have  $\text{rank}(D_u) \leq k$ . Therefore, in the case that  $u$  has  $\Omega(n)$  incident edges, (4) is essentially saying that no incident edge's leverage score exceeds the average by too much.

To get intuition for condition (5), we need the following two results. The first theorem is proved in Section 8. The second proposition is proved in Appendix C.

**Theorem 3.11.** *Let  $H$  be a  $k \times k$ -matrix weighted graph that is  $\gamma$ -almost regular (in the sense of (4)). Then for any  $\zeta \in (0, 1)$ , the number of eigenvalues of its normalized Laplacian that are between  $(0, \zeta]$  is at most  $\frac{\gamma k^2}{(1-\zeta)^2}$ .*

**Proposition 3.12.** *Let  $\ell$  be the number of  $\lambda_i$ 's that are between  $(0, \zeta]$ . Then*

$$\sum_{(u,v) \in E} \left( D^{\dagger/2} b_{uv} \right)^T \left( \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} f_i f_i^T \right) D^{\dagger/2} b_{uv} = \ell. \quad (6)$$

Therefore, in the case that  $|E| = \Omega(n^2)$ , (5) is essentially saying that the LHS for every edge  $(u, v)$  does not exceed the average by too much.

We then show that every dense enough matrix-weighted graph can indeed be made into an expander by downscaling a small number of edges. To this end, let us define, for a *scaling*  $s : E \rightarrow [0, 1]$ , the rescaled graph  $H^s$ , which is obtained from  $H$  by rescaling each edge  $(u, v)$ 's weight to  $s_{uv}^2 \phi_{uv} \phi_{uv}^T$ . The proof of the following theorem appears in Sections 9 and 10.

**Theorem 3.13.** *There is an algorithm that, given any  $k \times k$  matrix-weighted graph  $H = (V, E)$  with  $|E| \geq \Omega(n^2)$ , outputs a scaling  $s : E \rightarrow [0, 1]$  such that*

1. *The rescaled graph  $H^s$  is a  $(\gamma, \zeta, \psi)$ -almost regular expander for*

$$\gamma = 8 \log n, \quad \zeta = \frac{1}{\log n}, \quad \psi = 16k^2 \log^3 n.$$

2. *The number of edges  $(u, v) \in E$  with  $s_{uv} < 1$  is  $o(n^2)$ .*

We next show that almost-regular expanders are preserved under vertex sampling. However, we will now use a different notion of ‘‘preservation’’. To state our specific result, let us define some additional notations. For a vertex subset  $C \subseteq V$ , we write  $L_{G[C]}$  to denote the Laplacian of the vertex-induced subgraph  $G[C]$ . We also let  $D_{CC}$  be the submatrix of  $D$  (the degree matrix of the original graph  $H$ ) with rows and columns restricted to vertices in  $C$ , and let  $(f_i)_C$  denote, for an eigenvector  $f_i$ , the vector  $f_i$  with indices restricted to  $C$ . We then have the following theorem, whose proof appears in Section 11.2.

**Theorem 3.14.** *There exists a  $\theta = \theta(n) \leq n^{o(1)}$  with the following property. Let  $H = (V, E)$  be a  $k \times k$  matrix-weighted,  $(\gamma, \zeta, \psi)$ -almost regular expander where  $\zeta \leq 1/\log n$ . For an  $s \geq 2 \cdot 10^6 \gamma \psi \zeta^{-1} k^2 \theta(n)$ , let  $C \subseteq V$  be a uniformly random vertex subset of size  $s$ . Then with probability at least  $1 - 1/n^5$ , we have that*

1. The null space of  $D_{CC}^{\dagger/2} L_{G[C]} D_{CC}^{\dagger/2}$  is exactly the linear span of  $\{(f_i)_C : \lambda_i = 0\}$ .

2. For all vectors  $x \in \mathbb{R}^{|C|k}$  such that  $x^T (f_i)_C = 0, \forall i : \lambda_i = 0$ ,

$$x^T \left( D_{CC}^{\dagger/2} L_{G[C]} D_{CC}^{\dagger/2} \right)^\dagger x \leq n^{o(1)} \cdot x^T \left( \frac{n^2}{s^2} \sum_{i:\lambda_i \in (0,\zeta]} \frac{1}{\lambda_i} (f_i)_C (f_i)_C^T + \frac{n}{s} \cdot \frac{1}{\zeta} I \right) x. \quad (7)$$

We argue that (7) is roughly saying that the pseudoinverse of the subgraph  $G[C]$  can be bounded by the pseudoinverse of the original graph that is (i) restricted to indices in  $C$  and (ii) rescaled in a certain way. For technical reasons, on the LHS of (7) we normalize the Laplacian of the vertex-induced subgraph using the degree matrix of the *original graph*  $H$ . As for the RHS, we can see it as a rescaled version of the pseudoinverse of  $N$  restricted to  $C$ , by noting that

$$\left( N^\dagger \right)_{CC} = \sum_{\lambda_i > 0} \frac{1}{\lambda_i} (f_i)_C (f_i)_C^T.$$

Thus, on the RHS of (7) we blow up the small eigenvalues quadratically in  $1/(\text{sampling rate})$ , but blow up the large eigenvalues linearly in  $1/(\text{sampling rate})$ .

With these tools, we are finally able to prove Theorem 1.5. We present the proof in Section 12.

### 3.4.1 Techniques for proving Theorems 3.11, 3.13, 3.14

We now explain, at a very high level, the techniques that we use to prove these three key theorems, as well as their connections to previous works. More details can be found in the subsequent sections.

**Proof of Theorem 3.11.** We consider the “spectral embedding” induced by the bottom eigenvectors of the normalized Laplacian, which maps each vertex to a rectangular matrix. Such a spectral embedding may be seen as the matrix-weighted counterpart of the ones for scalar-weighted graphs, which map each vertex to a *vector*. The latter embeddings were previously used to prove higher-order Cheeger inequalities [LRTV12, LGT14]. We show that, for matrix-weighted graphs that are almost regular (in the sense of (4)), the spectral embedding has vertex-wise bounded spectral norm (Lemma 8.2), and as a result the number of bottom eigenvectors must be small (Theorem 8.1).

**Proof of Theorem 3.13.** Our proof consists of two steps: (i) decomposing the graph into an almost regular graph (Theorem 9.1), and (ii) decomposing the graph into an almost regular expander (Theorem 10.1). In achieving (ii), we actually invoke (i) repeatedly to maintain the almost regularity of the graph.

As noted above, the almost regularity condition (4) is essentially saying that no incident edge has leverage score too large comparing to the average. A similar task to (i) has in fact been investigated by a previous work [CLM<sup>+</sup>15], where the authors showed that given a set of vectors, one can, by downscaling a small number of them, make every vector have small leverage score comparing to the average. This result is achieved by an algorithm that iteratively downscales

vectors with large leverage scores, while analyzing how each vector's leverage score changes in the process. While it is possible to directly invoke the result from [CLM<sup>+</sup>15] to get a large almost regular graph, its guarantee does not suffice for our purpose of smoothly incorporating (i) into (ii). In particular, since we will repeatedly invoke (i) in (ii), we need, in addition to that the number of rescaled edges is small, an extra bound on the number of completely deleted edges (i.e. those rescaled to 0) that is proportional to the rank change of the degree matrix  $D$ . As a result, we design a more involved algorithm for obtaining the scaling as well as carry out a more careful analysis of the algorithm.

Achieving (ii) turns out to be much more challenging. Although the LHS of (5) may be seen as a leverage score, there is the intrinsic difficulty that whenever the edge weights change, so do the eigenvalues and eigenvectors of the normalized Laplacian, as well as the degree matrix itself (hence also the normalized incidence vectors). Thus it is not clear how the LHS of (5) will change. As a result, when trying to obtain a desired scaling, we have to use some global measure of progress. This is in contrast to (i), where we can track the leverage score change of each edge *locally*. We resolve this issue by considering, as a potential function, the determinant of the normalized Laplacian restricted to the bottom eigenspace. In other words, our potential function is the product of the eigenvalues of  $N$  that are between  $(0, \zeta]$ <sup>7</sup>. We show that, by a delicate global analysis of such a potential function, we are able to make the graph an expander by only downscaling a small number of edges.

**Proof of Theorem 3.14.** Our proof is motivated by the *approximate Gaussian elimination* of the Laplacian matrices of scalar-weighted graphs, which was previously used as an algorithmic tool for solving graph structured linear systems [KLP<sup>+</sup>16, KS16] and building data structures for dynamically maintaining effective resistances [DKP<sup>+</sup>17, LZ18, LPYZ20]. Our approach also relies on analyzing matrix-valued martingales which played a key role in [KS16], which have played key roles in constructing vertex/subspace sparsifiers [KS16, LS18, FGL<sup>+</sup>21].

Let us first briefly review the Gaussian elimination of the Laplacian matrix of a scalar-weighted graph. Roughly speaking, by eliminating the row and column of  $L$  corresponding to a vertex  $u$ , we can obtain another Laplacian matrix  $L'$  supported on  $V \setminus \{u\}$  whose pseudoinverse equals the pseudoinverse of the original  $L$  restricted to  $V \setminus \{u\}$  (i.e.  $(L')^\dagger = (L^\dagger)_{V \setminus \{u\} V \setminus \{u\}}$ ). Given a vertex subset  $C \subseteq V$ , one can also eliminate the vertices outside of  $C$  one by one and get a Laplacian matrix  $L''$  supported on  $C$  with the same property that  $(L'')^\dagger = (L^\dagger)_{CC}$ . The matrix  $L''$  is referred to as the *Schur complement* of  $L$  onto  $C$ . However, the graphs associated with  $L'$  and  $L''$  could be dense, which are inefficient for algorithm design. Therefore [KS16] showed that one can perform an *approximate* Gaussian elimination by, upon each elimination, implicitly sub-sampling the edges in  $L'$ . They then showed that we eventually get a good approximation to  $L''$  by analyzing a matrix-valued martingale induced by this process.

We now explain how to apply this idea to prove Theorem 3.14. Since we are considering an induced subgraph  $G[C]$  where  $C$  is a uniformly random subset of size  $s$ , we can also view the process for choosing  $C$  as deleting a sequence of  $n - s$  vertices from  $V$  uniformly at random. Our goal is to compare the pseudoinverse of  $G[C]$  with that of the original graph, therefore it suffices to compare it with the Schur complement of  $L$  onto  $C$ . We will in fact do such a comparison upon the elimination of every vertex. That is, if we let  $C_i$  be the set of remaining vertices at

---

<sup>7</sup>Due to technical reasons, the actual potential function slightly differs from the one stated here.

the  $i^{th}$  step, then we want to compare the Laplacian of  $G[C_i]$  with the Schur complement of  $L$  onto  $C_i$ . At a high level, we do so by setting up a matrix-valued martingale, and show that it has good concentration when  $G$  is a matrix-weighted expander (in the sense of Definition 3.6).

## 4 Preliminaries

### 4.1 Matrices

**Definition 4.1** (Pseudoinverse). Let  $A$  be an  $n \times n$  symmetric matrix. Let  $\lambda_1, \lambda_2, \dots, \lambda_n$  be the eigenvalues of  $A$  and let  $v_1, v_2, \dots, v_n$  be orthonormal (i.e. unit and orthogonal) eigenvectors of  $A$ , then by the spectral theorem  $A = \sum_{i=1}^n \lambda_i v_i v_i^T$ . The pseudoinverse of  $A$  is then defined as

$$A^\dagger \stackrel{\text{def}}{=} \sum_{\lambda_i \neq 0} \frac{1}{\lambda_i} v_i v_i^T. \quad (8)$$

**Definition 4.2** (Matrix partial ordering). For two  $n \times n$  symmetric matrices  $A, B$ , we write  $A \preceq B$  if for any vector  $x \in \mathbb{R}^n$  we have  $x^T A x \leq x^T B x$  (i.e.  $B - A$  is positive semi-definite).

**Fact 4.1** (Properties of the matrix partial ordering).

1. If  $A \preceq B$  and  $C \preceq D$ , then  $A + C \preceq B + D$ .
2. If  $A \preceq B$  where  $A, B \in \mathbb{R}^{n \times n}$ , then for any  $W$  with  $n$  rows,  $W^T A W \preceq W^T B W$ .
3. For two positive semidefinite matrices  $A, B$  that have the same null space,  $A \preceq B$  implies that  $B^\dagger \preceq A^\dagger$ .

**Theorem 4.2** (Matrix chernoff bound [Tro12]). Let  $X_1, \dots, X_m \in \mathbb{R}^{n \times n}$  be independent random positive semidefinite matrices such that  $\lambda_{\max}(X_i) \leq R, \forall i$  and  $\mathbb{E}[\sum_{i=1}^m X_i] = I_{n \times n}$  where  $I_{n \times n}$  is the  $n \times n$  identity matrix. Then with probability at least  $1 - 2n \exp\left\{-\frac{\epsilon^2}{3R}\right\}$

$$(1 - \epsilon)I \preceq \sum_{i=1}^m X_i \preceq (1 + \epsilon)I. \quad (9)$$

**Definition 4.3** (Leverage scores). Let  $a_1, \dots, a_m \in \mathbb{R}^n$  and  $A = \sum_{i=1}^m a_i a_i^T \in \mathbb{R}^{n \times n}$ . The leverage score of  $a_i$  w.r.t.  $A$  is defined to be  $\tau_i(A) = a_i^T A^\dagger a_i$ .

**Fact 4.3.** Let  $a_1, \dots, a_m \in \mathbb{R}^n$  and  $A = \sum_{i=1}^m a_i a_i^T \in \mathbb{R}^{n \times n}$ . Also let  $B = (a_1, \dots, a_m) \in \mathbb{R}^{n \times m}$ , so we have  $A = BB^T$ . Let  $b \in \mathbb{R}^n$  be a vector in the span of  $a_1, \dots, a_m$ . Then we have

$$\min_{Bx=b} \|x\|_2^2 = b^T (BB^T)^\dagger b.$$

### 4.2 Multivariate Gaussian distributions

**Definition 4.4.** Let  $\mu \in \mathbb{R}^d$  be a vector and  $\Sigma \in \mathbb{R}^{d \times d}$  be a matrix. We say a random vector  $x \in \mathbb{R}^d$  follows a multivariate Gaussian distribution with mean  $\mu$  and covariance matrix  $\Sigma$ , denoted by  $\mathcal{N}(\mu, \Sigma)$ , if

1. Each  $x_i$  distributes as  $\mathcal{N}(\mu_i, \Sigma_{ii})$ , a univariate Gaussian with mean  $\mu_i$  and variance  $\Sigma_{ii}$ .
2.  $\mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)] = \Sigma_{ij}$  for all pairs  $i, j$ . Or equivalently,

$$\mathbb{E}[(x - \mu)(x - \mu)^T] = \Sigma. \quad (10)$$

**Fact 4.4.** The covariance matrix  $\Sigma$  is symmetric and positive semi-definite.

**Fact 4.5.** Let  $x_1, x_2 \in \mathbb{R}^d$  be independent random vectors such that  $x_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$  and  $x_2 \sim \mathcal{N}(\mu_2, \Sigma_2)$ . Then  $x_1 + x_2 \sim \mathcal{N}(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$ .

**Theorem 4.6** ( $\ell_1$ -distance between multivariate Gaussians with the same mean [DMR18]). *Let  $\mu \in \mathbb{R}^d$  and let  $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d \times d}$  be positive semidefinite such that*

1.  $\Sigma_1$  and  $\Sigma_2$  have the same null space.
2.  $\Sigma_2 = \Sigma_1 + ww^T$  for some  $w \in \mathbb{R}^d$ .

Then we have

$$d_{TV}(\mathcal{N}(\mu, \Sigma_1), \mathcal{N}(\mu, \Sigma_2)) = \Theta\left(\min\left\{1, w^T \Sigma_1^\dagger w\right\}\right).$$

**Theorem 4.7** (Chernoff bound for univariate Gaussians, Theorem 9.3 of [MU17]). *Let  $X$  be a univariate Gaussian with mean  $\mu$  and variance  $\sigma^2 > 0$ :  $X \sim \mathcal{N}(\mu, \sigma^2)$ . Then for any  $a > 0$*

$$\Pr[|X - \mu| \geq a\sigma] \leq 2e^{-a^2/2}. \quad (11)$$

### 4.3 $\ell_2$ -heavy hitter, $\ell_p$ -sampler and $\ell_2$ estimation

**Proposition 4.8** ( $\ell_2$ -heavy hitters [KLM<sup>+</sup>14]). *For any  $\eta > 0$ , there is a decoding algorithm  $D$  and a distribution on matrices  $A \in \mathbb{R}^{O(\eta^{-2}\text{polylog}(N)) \times N}$  such that, for any  $x \in \mathbb{R}^N$ , given  $Ax$ , the algorithm  $D$  returns a vector  $\tilde{x}$  such that  $\tilde{x}$  has  $O(\eta^{-2}\text{polylog}(N))$  non-zeros and satisfies*

$$\|x - \tilde{x}\|_\infty \leq \eta \|x\|_2$$

with high probability over the choice of  $A$ . The sketch  $Ax$  can be maintained and decoded in  $O(\eta^{-2}\text{polylog}(N))$  space.

Given a vector  $x$  of size  $N$  and a number  $\delta > 0$ , an  $\ell_p$  sampler is an algorithm that output an index  $i$  with probability

$$p_i \in (1 \pm N^{-c}) \frac{|x_i|^p}{\|x\|_p^p} \pm O(N^{-c})$$

where  $c$  is arbitrary constant. The algorithm may also output `Fail` with probability at most  $\delta$ .

For any  $0 \leq p \leq 2$ , there is a polylog size linear sketch for  $\ell_p$ -sampler.

**Proposition 4.9** ([JST11, CF14]). *For any constant  $c$  and  $0 < \delta < 1$ , there is a linear sketch for  $\ell_0$ -sampling with measurement size  $O(\log^2 n \log 1/\delta)$ .*

**Proposition 4.10** ([JW21]). *For any  $0 < p < 2$ , any  $\epsilon, \delta_1, \delta_2 > 0$  and any constant  $c > 0$ , there is a linear sketch with measurement size  $O(\log^2 N (\log \log N)^2 \log(1/\delta_1) + \epsilon^{-p} \log N \log^2(1/\delta_2) \log(1/\delta_1))$  such that given an  $N$ -dimensional vector  $x$ , with probability  $(1 - \delta_1)$ , it can recover an index  $i$ , such that the probability of outputting  $i$  is*

$$p_i \in (1 \pm N^{-c}) \frac{|x_i|^p}{\|x\|_p^p} \pm O(N^{-c})$$

Moreover, if the sketch does output an index  $i$ , then it also recovers a value  $x'_i$  such that  $|x_i| \leq x'_i \leq (1 + \epsilon) |x_i|$ , with probability  $1 - \delta_2$ .

**Proposition 4.11** ([KLM<sup>+</sup>14]). *For any  $\epsilon > 0$  and any constant  $c > 0$ , there is a linear sketch of size  $O(\epsilon^{-2} \text{polylog}(N))$  such that given an  $N$ -dimensional vector  $x$ , with probability  $N^{-c}$ , we can recover a vector  $x'$  such that  $\|x - x'\|_\infty \leq \epsilon \|x\|_2$ .*

**Proposition 4.12** ([JL84]). *For any  $0 < \delta < 1$  and any constant  $c > 0$ , there is a linear sketch of size  $O(\delta^{-2} \log N)$  such that given an  $N$ -dimensional vector  $x$ , with probability  $N^{-c}$ , we can recover a vector  $x'$  such that  $\|x'\|_2 \in (1 \pm \delta) \|x\|_2$ .*

#### 4.4 Edge strengths and cut sparsifiers

Given a graph  $G$ , a  $k$ -strongly connected component is a maximal vertex induced subgraph whose minimum cut size is at least  $k$ . Thus, all  $k$ -strongly connected components form a partition of the entire vertex set. The following fact gives an equivalent way of obtaining the  $k$ -strongly connected components.

**Fact 4.13.** Given any graph and integer  $k$ , consider a process where we iteratively remove (the edges across) an arbitrary cut of size strictly smaller than  $k$ , until there is no such cut left. Then the connected components in the resulting graph are the  $k$ -strongly connected components of the original graph.

The *strength* of an edge  $e$  in the graph, denoted  $k_e$ , is the maximum value of  $k$  such that a  $k$ -strong component of  $G$  contains both endpoints of  $e$ . The weighted sum of the inverse of the strength of every edge in a graph is at most  $n - 1$ .

**Claim 4.14** ([BK15]). For any weighted graph  $G = (V, F, w)$  on  $n$  vertices,  $\sum_{f \in F} \frac{w_f}{k_f} \leq n - 1$ .

Given a graph  $G$ , an  $(1 \pm \epsilon)$ -cut sparsifier  $G'$  is a subgraph of  $G$  such that any cut in  $G$  is preserved in  $G'$  to within a factor of  $(1 \pm \epsilon)$ . The seminal result [BK15] shows that, for any graph  $G$ , if we construct a graph  $G'$  as follows: we include each edge  $e$  in  $G'$  with probability  $p_e \geq \Omega(\frac{w_e \log n}{\epsilon^2 k_e})$ , and give weight  $\frac{w_e}{p_e}$  if it gets chosen, then  $G'$  is a  $(1 \pm \epsilon)$  cut sparsifier of  $G$  with high probability.

#### 4.5 Graph matrices, leverage scores, and spectral sparsifiers

Fix an arbitrary orientation of all possible  $\binom{n}{2}$  edge slots of the graph. Let  $B \in \mathbb{R}^{\binom{n}{2} \times n}$  be the *edge-vertex incidence matrix* of an undirected, unweighted complete graph over  $n$  vertices. That is, for every edge  $e = (u, v)$  oriented from  $u \rightarrow v$ , there is a row  $b_e \in \mathbb{R}^n$  in  $B$  corresponding to

$e$  such that the column  $u$  has value 1, the column  $v$  has value  $-1$ , and all other columns have value 0. We also write  $b_e$  as  $b_{u,v}$ . For a graph  $G$ , we write  $B_G \in \mathbb{R}^{\binom{n}{2} \times n}$  to denote the matrix obtained from  $B$  by zeroing out rows corresponding to absent edges in  $G$ .

Given any weighted graph  $G$ , let  $W_G \in \mathbb{R}^{\binom{n}{2} \times \binom{n}{2}}$  be the diagonal matrix whose diagonal entries are the weights of the edges corresponding to them, i.e.  $(W_G)_{ee} = w_e$ . If an edge  $e$  is not present in the graph, then  $(W_G)_{ee} = 0$ . The *Laplacian matrix* of  $G$  is given by  $L_G = B_G^T W_G B_G = (W_G^{1/2} B_G)^T (W_G^{1/2} B_G)$ . Notice that for unweighted graphs, we have  $L_G = B_G^T B_G$ .

For any edge  $e$ , its *leverage score*  $\tau_e$  is given by  $\tau_e = (\sqrt{w_e} b_e)^T L_G^\dagger (\sqrt{w_e} b_e) = w_e b_e^T L_G^\dagger b_e$ , where  $L_G^\dagger$  is the Moore-Penrose pseudoinverse of the Laplacian matrix  $L_G$ .

**Fact 4.15.** The sum of the leverage scores of all edges  $\sum_e \tau_e = \text{rank}(W_G^{1/2} B_G) \leq n - 1$ .

In general, for *any* matrix  $C \in \mathbb{R}^{m \times n}$  whose  $i^{\text{th}}$  row is denoted by  $c_i \in \mathbb{R}^n$ , we define the leverage score of the  $i^{\text{th}}$  row by  $\tau_i = c_i^T (C^T C)^\dagger c_i$ , and we once again have  $\sum_{i=1}^m \tau_i = \text{rank}(C)$ .

If we view the graph  $G$  as an electrical network where each edge  $e$  has resistance  $1/w_e$ , then the *effective resistance* between the two vertices  $s, t$  is given by  $r_{s,t} = b_{s,t}^T L_G^\dagger b_{s,t}$ . For an edge  $e = (s, t)$ , we also define its effective resistance as  $r_e = r_{s,t}$ . Thus, the leverage score  $\tau_e = w_e r_e$ .

If we inject  $f$  units of electrical flow into a vertex  $s$  and extract 1 unit from a vertex  $t$ , then  $f L_G^\dagger b_{s,t} \in \mathbb{R}^n$  is referred to as *the set of vertex potentials induced by the electrical flow*. The relation between vertex potentials of the electrical flow is characterized by Ohm's Law.

**Fact 4.16** (Ohm's Law). Let  $x = f L_G^\dagger b_{s,t} \in \mathbb{R}^n$  be the set of vertex potentials when we send  $f$  units of electrical flow from  $s$  to  $t$ . Then we have  $x_s - x_t = f(b_{s,t} L_G^\dagger b_{s,t})$ . Moreover, for any edge  $e = (u, v)$  with  $x_u \geq x_v$ , the flow on this edge is in the direction of  $u \rightarrow v$  and has amount exactly  $w_e(x_u - x_v)$ .

It is also known that the vertex potentials induced by an electrical flow minimizes the total energy. Specifically, for an arbitrary set of vertex potentials  $x \in \mathbb{R}^n$ , we define its *energy* to be  $x^T L_G x = \sum_{e=(u,v) \in G} w_e (x_u - x_v)^2$ , and define its *normalized energy with respect to vertices  $s, t$*  to be  $\frac{x^T L_G x}{(x^T b_{s,t})^2}$  (i.e. the energy divided by  $(x_s - x_t)^2$ ). Then we have:

**Fact 4.17.** The vertex potentials induced by an electrical flow from  $s$  to  $t$  minimizes the normalized energy with respect to  $s, t$ . That is, for any  $f > 0$ , we have

$$f L_G^\dagger b_{s,t} \in \arg \min_{x \in \mathbb{R}^n} \frac{x^T L_G x}{(x^T b_{s,t})^2},$$

and thus, by plugging in  $x = L_G^\dagger b_{s,t}$ , the smallest normalized energy w.r.t.  $s, t$  is

$$\min_{x \in \mathbb{R}^n} \frac{x^T L_G x}{(x^T b_{s,t})^2} = \frac{b_{s,t}^T L_G^\dagger L_G L_G^\dagger b_{s,t}}{(b_{s,t}^T L_G^\dagger b_{s,t})^2} = \frac{b_{s,t}^T L_G^\dagger b_{s,t}}{(b_{s,t}^T L_G^\dagger b_{s,t})^2} = \frac{1}{b_{s,t}^T L_G^\dagger b_{s,t}},$$

exactly 1 over the effective resistance between  $s, t$ .

Given a graph  $G$  with Laplacian matrix  $L_G$ , a  $(1 \pm \epsilon)$ -spectral sparsifier  $G'$  is a graph with Laplacian matrix  $L_{G'}$  such that for any  $x \in \mathbb{R}^n$ ,  $x^T L_{G'} x \in (1 \pm \epsilon)x^T L_G x$ . In other words,  $\frac{1}{1+\epsilon} L_G \preceq L_{G'} \preceq (1+\epsilon)L_G$  (recall Definition 4.2). We also use  $L_G \approx_{1+\epsilon} L_{G'}$  to denote the

same relation between  $L_G$  and  $L_{G'}$ . For two scalars  $a, b \geq 0$ , we also write  $a \approx_{1+\epsilon} b$  to denote  $\frac{a}{1+\epsilon} \leq b \leq (1+\epsilon)a$ .

If we sample each edge of a graph with probability proportional to its leverage score or larger, and reweight it accordingly, then with high probability we get a spectral sparsifier. In fact, this sampling process gives a good spectral approximation for any  $C \in \mathbb{R}^{m \times n}$

**Theorem 4.18** ([SS11, Tro12]). *Let  $\epsilon > 0$ . Given a matrix  $C \in \mathbb{R}^{m \times n}$ , let  $p_1, \dots, p_m \in \mathbb{R}$  be such that  $1 \geq p_i \geq \min\{1, 100\tau_i\epsilon^{-2} \log n\}$  for all  $i \in [m]$ . Let  $\tilde{W} \in \mathbb{R}^{m \times m}$  be a diagonal matrix such that  $\tilde{W}_{ii} = 1/p_i$  with probability  $p_i$  and  $\tilde{W}_{ii} = 0$  otherwise. Then with high probability,*

$$C^T \tilde{W} C \approx_\epsilon C^T C.$$

Recall that when  $C = W_G^{1/2} B_G$ , we have  $L_G = C^T C$ . Thus if we sample each edge  $e$  with probability  $p_e = \min\{1, 100\tau_e\epsilon^{-2} \log n\}$  and reweight it to  $w_e/p_e$  if sampled, we get, by the claim above and that  $\sum_e \tau_e \leq n - 1$ , a  $(1 + \epsilon)$ -spectral sparsifier of  $O(n\epsilon^{-2} \log n)$  edges. We could also do an oversampling, where we sample each edge  $e$  with some probability  $p_e \geq \min\{1, 100\tau_e\epsilon^{-2} \log n\}$  and also reweight it to  $w_e/p_e$  if sampled, and then we get a  $(1 + \epsilon)$ -spectral sparsifier of  $O(\sum_e p_e)$  edges.

## 5 A linear sketching algorithm for weighted cut sparsification

In this section, we present a linear sketch with  $\tilde{O}(n\epsilon^{-3})$  measurements that computes a  $(1 + \epsilon)$ -cut sparsifier of a weighted graph. In particular, our linear sketch will be an incidence sketch. Our algorithm is obtained by generalizing the cut sparsifier algorithm of [RSW18], which is an  $O(\log n)$  round cut query algorithm that works only for unweighted graphs. We first describe the weighted generalization of the algorithm of [RSW18] in a model-oblivious manner, and then show how to implement it by linear sketching.

### 5.1 A model oblivious algorithm for weighted cut sparsification

We present a model-oblivious algorithm `WEIGHTEDCUTSPARSIFY` for weighted cut sparsification in Figure 2, which generalizes the idea presented in [RSW18] for unweighted graphs. Here, we assume the edge weights of the input graph are between  $[1, U]$  for some  $U \geq 1$  that is known to us. We characterize the performance of the algorithm in the lemma below.

**Lemma 5.1.** *Fix  $\alpha = 800, \beta = 400, \gamma = 100$  at Line 2 of the algorithm `WEIGHTEDCUTSPARSIFY`. Let  $\epsilon$  be an arbitrary number in  $(0, 1)$ . Then, the algorithm `WEIGHTEDCUTSPARSIFY` outputs, with high probability, a  $(1 + \epsilon)$ -cut sparsifier of the input graph  $G$  with  $O(n\epsilon^{-2} \log n)$  edges.*

To prove the proposition, we will need the following lemma.

**Lemma 5.2.** *In each iteration  $\ell$ , with high probability:*

1. *The edges in  $\tilde{G}$  that survive the contractions have strength at most  $2^{\ell-1}$ .*
2. *The edges in  $\tilde{G}$  that are within the components  $C_1, \dots, C_r$  have strength  $[2^{\ell-4}, 2^\ell]$ .*

*Proof.* We show that at the end of each iteration  $\ell$ , the edges within  $C_1, \dots, C_r$  have strength  $\geq 2^{\ell-4}$ , and the edges between them have strength at most  $2^{\ell-1}$ . Then by noting that all edges in the graph have strength at most  $nU$ , the lemma follows by an induction on  $\ell$ .

Consider iteration  $\ell$  of the for loop. Let  $V_1, \dots, V_t$  be the partition of  $\tilde{G}$  into maximal  $2^{\ell-4}$ -strongly connected components. Then by Fact 4.13 there exists a way to arrive at these components by starting from the entire graph  $\tilde{G}$  and iteratively removing a cut with size  $< 2^{\ell-4}$ . By applying a Chernoff bound and a union bound over the sequence of (at most  $n - 1$ ) cuts that we remove in this process, we have that after sampling, each of these cuts has size at most  $50 \log n$ . As a result, the partition  $C_1, \dots, C_r$  of  $G_\ell$  into maximal  $100 \log n$ -strongly connected components is a refinement of  $V_1, \dots, V_t$ . This implies that all edges within  $C_1, \dots, C_r$  have strength  $\geq 2^{\ell-4}$ .

Now consider the partition  $V_1, \dots, V_s$  of  $\tilde{G}$  into maximal  $2^{\ell-1}$ -strongly connected components. After sampling at Line 3a, these components still have min cut  $\geq 100 \log n$  with high probability. Thus, the partition  $C_1, \dots, C_r$  of  $G_\ell$  into maximal  $100 \log n$ -strongly connected components is a coarsening of  $V_1, \dots, V_s$ . This implies that all edges going across different  $C_i$ 's have strength at most  $2^{\ell-1}$ . This finishes the proof of the lemma.  $\square$

*Proof of Lemma 5.1.* By Lemma 5.2, whenever we sample an edge  $e$  of the graph, we sample it with probability  $p_e \geq \min\{1, 50w_e k_e^{-1} \epsilon^{-2} \log n\}$  and re-weight it to  $w_e/p$ . Moreover, in the last iteration  $\ell = 0$ , all edges are within the components  $C_1, \dots, C_r$  (since each edge's strength is at least 1). Therefore ultimately all edges get sampled in our algorithm. As a result, we get with high probability a  $(1 + \epsilon)$ -cut sparsifier. On the other hand, the probability  $p_e$  with which we sample  $e$  also satisfies  $p_e \leq 800w_e k_e^{-1} \epsilon^{-2} \log n$ , and thus we get a sparsifier with  $O(n\epsilon^{-2} \log n)$  edges.  $\square$

## 5.2 Implementation by linear sketching

Now we show how to implement the algorithm by linear sketching. The implementation is motivated by the techniques first used in [AGM12a]. Note that it suffices to implement the two edge sampling processes at Lines 3a, 3(b)iii and the contraction operations at Line 3(b)iv.

We note that the implementation of both sampling processes can be seen as the following task: we first independently generate a uniformly random real number  $R_e \in [0, 1]$  for each edge slot  $e$ , and then recover all edges satisfying  $R_e < w_e p$  for some given  $p$  (which in iteration  $\ell$  equals  $\beta 2^{-\ell} \log n$  for the first process and  $\alpha \epsilon^{-2} 2^{-\ell} \log n$  for the second process). Here we can generate the  $R_e$ 's offline, but have to recover the sampled edges using linear sketching. We achieve the latter by repeatedly finding a spanning forest formed by the sampled edges. We will show that the sampled edges can be found by performing  $\tilde{O}(1)$  iterations of spanning forest recovery.

We shall first show how to recover a spanning forest formed by the sampled edges via linear sketching. To this end, we need a linear sketching subroutine that we call *weighted edge sampler*, with the following guarantee.

**Lemma 5.3.** *Let  $R_1, \dots, R_N$  be  $N$  numbers independently and uniformly at random generated from  $[0, 1]$ , let  $c > 0$  be an arbitrary constant, and let  $p \in (0, 1)$  be a parameter. There exists a linear sketch of  $\text{polylog}(N, 1/p)$  measurements with the following guarantee. For any vector*

$$H \leftarrow \text{WEIGHTEDCUTSPARSIFY}(G, \epsilon)$$

1. Initially, let the sparsifier  $H$  be an empty graph, and let  $\tilde{G} \leftarrow G$ .
2. Fix some sufficiently large constants  $\alpha > \beta > \gamma > 0$ .
3. For  $\ell \in \{\log(nU), \log(nU) - 1, \dots, 1, 0\}$ :
  - (a) Obtain  $G_\ell$  from  $\tilde{G}$  by keeping each edge w.p.  $p_e := \min\{\beta \cdot w_e \cdot 2^{-\ell} \log n, 1\}$  and with weight  $(\beta \cdot w_e \cdot 2^{-\ell} \log n)/p_e$ .
  - (b) In each connected component of  $G_\ell$ :
    - i. While there exists a cut of weight  $\leq \gamma \cdot \log n$ , remove the edges in that cut and recurse on both sides; repeat until there is no such cut.
    - ii. Let  $C_1, \dots, C_r$  be the connected components induced by the remaining edges.
    - iii. For each edge  $e$  in  $\tilde{G}$  with endpoints in the same  $C_i$ , add it to  $H$  with probability  $p_e := \min\{\alpha \cdot w_e \cdot \epsilon^{-2} 2^{-\ell} \log n, 1\}$  and with weight  $w_e/p_e$ .
    - iv. Update  $\tilde{G}$  by contracting each of  $C_1, \dots, C_r$ .

Figure 2: Model oblivious algorithm for weighted cut sparsification.

$w \in \mathbb{R}^N$ , if there exists an entry  $e$  such that  $R_e \leq w_e p$  and  $w_e > N^{-c}$ , then with high probability, the sketch recovers an index  $e'$  such that  $R_{e'} \leq (1 + \epsilon)w_{e'} p$  along with a  $(1 + \epsilon)$ -approximate estimate of  $w_{e'}$ .

To find a spanning forest of the sampled edges (those with  $R_e < w_e p$ ), we apply the weighted edge sampler sketch to the incidence vectors of the vertices in  $G$ . Specifically, we fix an arbitrary orientation of each of the  $\binom{n}{2}$  potential edges. Then for a vertex  $u \in G$ , we consider its incidence vector  $b_u \in \mathbb{R}^{\binom{n}{2}}$  given by

$$(b_u)_e = \begin{cases} w_e & e \in G, u \text{ is } e\text{'s head} \\ -w_e & e \in G, v \text{ is } e\text{'s tail} \\ 0 & e \notin G \text{ or } e \text{ does not touch } u. \end{cases} \quad (12)$$

Let  $t = \text{polylog}(n)$ . For any  $i \in [t]$ , let  $A_i$  be an independently generated weighted edge sampler sketching matrix. For each  $i \in [t]$  and vertex  $u \in G$ , we compute the sketch  $A_i b_u$ . Thus we make  $\tilde{O}(n)$  measurements in total. Now using the sketches we have taken, we recover a spanning forest of sampled edges via a  $\tilde{O}(1)$ -round process as follows.

In the first round, for each vertex  $u$ , we find an arbitrary outgoing edge using the weighted edge sampler sketch  $A_1 b_u$ . We then find all connected components induced by these edges, and add up the sketches  $A_2 b_u$  of vertices within the same component. Note that the edges within the same component cancel out in the summation, so the resulting sketches are in fact taken over the outgoing edges of each component. As a result, in the next round we are able to find an outgoing edge of each component. We then proceed similarly in the  $i^{\text{th}}$  round using sketches

$A_i b_u$ 's. Since in each round, the number of components is at least reduced by a factor of 2, we can find a spanning forest in  $O(\log n)$  rounds of this process.

In order to iteratively find  $\tilde{O}(1)$  edge-disjoint spanning forests, each time we find one, we “delete” the found edges from the other linear sketches, and restart the  $O(\log n)$ -round process above. Note that however, since we do not have the exact weights of the edges (Lemma 5.3 only gives approximation of them), we do not delete the found edges completely, but rather decrease each of their weights by an  $\Omega(1)$  factor.

Finally, to implement the contraction operations at Line 3(b)iv, we once again add the sketches of the vertices within each contracted component, just as we did in finding spanning forests. Then starting from the next iteration, the sketches work for the contracted graph  $\tilde{G}$ .

We conclude this subsection by proving that the sampled edges can be recovered by  $\tilde{O}(1)$  edge-disjoint spanning forests.

**Lemma 5.4.** *The edges in  $G_\ell$  can be found by  $O(\log^2 n)$  edge-disjoint spanning forests.*

*Proof.* We first show that the edges in  $G_\ell$  all have low strength.

**Claim 5.5.** Every edge in the graph  $G_\ell$  has strength  $O(\log n)$ .

*Proof.* By Lemma 5.2, we know that at the beginning of iteration  $\ell$ , all edges in  $\tilde{G}$  have strength at most  $2^\ell$ . This means that there is a way of removing all edges in the graph by iteratively removing a cut of size  $\leq 2^\ell$ . Then after sampling at Line 3a, these cuts all have size  $O(\log n)$ . Thus it follows that all edges in  $G_\ell$  have strength  $O(\log n)$ .  $\square$

Let  $C > 0$  be a constant such that all edges in  $G_\ell$  have strength  $\leq C \log n$ . Then this claim implies that  $G_\ell$  is *uniformly sparse*, in the sense that for any vertex induced subgraph  $G[S]$  where  $S \subseteq V$ , the number of edges in  $G[S]$  is  $(|S| - 1)C \log n$ . Indeed, each edge in  $G[S]$  has strength only smaller than in  $G$ , and thus all edges in  $G[S]$  can be removed by iteratively (for at most  $|S| - 1$  times) removing a cut of size  $C \log n$ . This in particular means that at any point, a spanning forest contains an  $1/(C \log n)$  fraction of the total remaining edges. Thus  $O(\log^2 n)$  edge-disjoint spanning forests recover all edges in  $G_\ell$ .  $\square$

### 5.3 Proof of Lemma 5.3

Roughly, we will simulate the non-uniform sampling process, where we want to sample each  $e$  with probability  $w_{ep}$ , by sampling the elements uniformly, but at different geometric rates. We will then essentially implement a *rejection sampling* process. Specifically, when subsampling all elements at some uniform rate  $q$ , we use  $\ell_1$ -samplers to recover a few elements that are sampled. We will then check if any one of the sampled elements  $e$  satisfies  $w_{ep} \approx q$ . If so, we will output this element; otherwise, we go the next sampling rate and repeat this step. We will show that with high probability, we will successfully recover a desired element.

*Proof of Lemma 5.3.* We will analyze the linear sketch given in Figure 3. The basic idea of the algorithm is as follows: for each  $0 \leq j \leq \gamma$ , we maintain  $\text{polylog}(N, \frac{1}{p})$  number of  $\gamma_1$ -sketches given by Proposition 4.10 with failure probability  $\delta_1, \delta_2 = N^{-100}$  that work for  $w^j$  where  $w^j$  is the vector generated by sampling each entry of  $w$  with probability  $2^{-j}$  (i.e.  $q$  in the overview above). For each sampled element  $e$ , we check if  $w_{ep}$  is indeed larger than  $2^{-j}$ , if so,  $e$  indeed gets sampled and we can output  $e$ , otherwise  $e$  might not get sampled, and we discard  $e$ . We

$e \leftarrow \text{WEIGHTEDEDGEAMPLER}(w \in \mathbb{R}^N, p, R_1, \dots, R_N)$

- Let  $t = 10^4 c\gamma \log N$  where  $\gamma$  is the minimum integer such that  $2^{-\gamma} < pN^{-c}$ .
- For each  $1 \leq i \leq t$  and each  $0 \leq j \leq \gamma$ , independently generate an  $\gamma_1$ -sampling matrix  $A_i^j$ .
- For each  $0 \leq j \leq \gamma$ , check if  $R_e > 2^{-j}$ . If so, for each  $1 \leq i \leq t$ , replace the column of  $A_i^j$  that corresponds to element  $e$  by the 0 vector.
- Compute  $A_i^j w$  for each pair of  $i$  and  $j$ , and recover a sampled element  $e_i^j$  and a weight  $w'_{e_i^j}$ , which is a  $(1 + \epsilon)$ -factor approximation to  $|w_{e_i^j}|$ .
- For each recovered pair  $e_i^j, w'_{e_i^j}$ , check if  $R_{e_i^j} \leq w'_{e_i^j} p$ . If there exists such an element  $e_i^j$ , output an arbitrary one of them, otherwise output **NoElement**.

Figure 3: Weighted edge sampler.

prove that whenever there are elements that get sampled, we will find one of them with high probability.

The analysis is conditioned on the event that none of the  $\gamma_1$ -samplers fails, which is a high probability event. Since the algorithm only uses  $\gamma_1$  samplers, we can without loss of generality assume that each element has positive weight. Note that the algorithm uses  $t\gamma$  number of  $\gamma_1$  sampling matrices in total, so the total number of measurements of this sketch is  $\text{polylog}(N, U, \frac{1}{p})$ . Moreover, by Proposition 4.10, if the algorithm outputs an element  $e$ , we have  $R_e \leq w'_e \leq (1 + \epsilon)w_e p$ . It is sufficient to prove that if there exists  $e$  such that  $R_e \leq w_e p$ , with high probability, we will not output **No Element**.

For any  $1 \leq k \leq \gamma$ , let  $S_k$  be the set of elements  $e$  such that  $2^{-k} \leq w_e p < 2^{-k+1}$ , and let  $S_0$  be the set of elements  $e$  such that  $w_e p \geq 1$ . For any  $0 \leq j \leq \gamma$ , let  $S_k^j$  be the set of elements  $e \in S_k$  such that  $R_e \leq 2^{-j}$ . For any  $j$  and  $k$ , let  $n_k = |S_k|$ ,  $n_k^j = |S_k^j|$  and  $W_k^j$  be the total weight of elements in  $S_k^j$ . The following claim follows from Chernoff bound.

**Claim 5.6.** With high probability, for any  $0 \leq j, k \leq \gamma$ , if  $n_k > 1000 \cdot 2^j \log N$ , then  $|2^j n_k^j - n_k| < n_k/2$ , otherwise  $n_k^j < 1500 \log N$ .

*Proof.* For each element  $e \in S_k$ , the probability that  $e \in S_k^j = 2^{-j}$ , so the size of  $n_k^j$  is the sum of  $|n_k|$  independent random 0/1 variables each with expectation  $2^{-j}$ . The expected size of  $n_k^j$  is  $2^{-j} n_k$ . If  $|n_k| > 1000 \cdot 2^j \log N$ , by Chernoff bound, the probability that  $|2^j n_k^j - n_k| > n_k/2$  is at most  $< 2e^{\frac{n_k}{2^{j+4}}} < N^{-50}$ .

If  $|n_k| \leq 1000 \cdot 2^j \log N$ , the probability that  $|n_k^j| \geq 1500 \log N$  is at most the probability of the case when  $n_k = 1000 \cdot 2^j \log N$ . So the probability is less than  $N^{-50}$ .  $\square$

Let  $k^*$  be the index that maximizes  $\frac{n_{k^*}}{2^{k^*}}$ . If  $n_{k^*} > 1000 \cdot 2^{k^*} \log N$ , by Claim 5.6,  $n_{k^*}^{k^*} > \frac{n_{k^*}}{2^{k^*+1}}$ .

Since each element in  $S_{k^*}$  has weight at least  $2^{-k^*}/p$ ,  $W_{k^*}^{k^*} > \frac{n_{k^*}}{2^{2k^*+1}p}$ . On the other hand, for any  $k > k^*$ , we have  $n_k^{k^*} < \max\{1500 \log N, \frac{3n_k}{2^{k^*+1}}\}$  by Claim 5.6. Since each element in  $S_k$  has weight at most  $2^{-k+1}/p$ , we have  $W_k^{k^*} < \max\{\frac{3000 \log N}{2^k p}, \frac{3n_k}{2^{k+k^*+1}}\}$ . As  $n_{k^*} > 1000 \cdot 2^{k^*} \log N$ ,  $\frac{3000 \log N}{2^k p} < \frac{3n_{k^*}}{2^{k+k^*} p} < 3W_{k^*}^{k^*}$ . Also, by definition of  $k^*$ ,  $\frac{n_{k^*}}{2^{k^*}} \geq \frac{n_k}{2^k}$ , which means  $\frac{3n_k}{2^{k+k^*+1}} \leq \frac{3n_{k^*}}{2^{2k^*+1}} < 3W_{k^*}^{k^*}$ . Thus, we have  $W_k^{k^*} < 3W_{k^*}^{k^*}$ . Note that for any  $1 \leq i \leq t$ ,  $e_i^{k^*}$  is obtained by an  $\gamma_1$  sampler from  $\cup_{k=0}^{\gamma} S_k^{k^*}$ , so with probability at least  $\frac{1}{3\gamma+3}$ ,  $e_i^{k^*}$  is an element in  $S_k$  such that  $k \leq k^*$ . In this case,  $R_e \leq 2^{-k^*} \leq w_e p$ , and so the algorithm will output an element that gets sampled. Since  $t = 10^4 \gamma \log N$ , there exists such an  $i$  with high probability.

If  $n_{k^*} \leq 1000 \cdot 2^{k^*} \log N$ , then for any  $k$ ,  $\frac{n_k}{2^k} \leq 1000 \log N$ . Let  $e$  be a maximum weight element such that  $R_e < w_e p$ . Suppose  $e \in S_k$ , then we have  $R_e < w_e p < 2^{-k+1}$ . Let  $k'$  be the largest index such that  $e \in S_k^{k'}$ , we have  $k' \geq k-1$ . For any  $e'$  such that  $e' \in S_{k''}$  with  $k'' < k$ , by definition of  $e$  and  $k$ ,  $R_{e'} > w_{e'} p \geq 2^{-k''} \geq 2^{-k+1} \geq 2^{k'}$ , which means  $e' \notin S_{k''}^{k'}$ . So  $S_{k''}^{k'} = \emptyset$  for any  $k'' < k$ . On the other hand, for any  $k'' \geq k$ ,  $n_{k''}^{k'} < \max\{1500 \log N, \frac{3n_{k''}}{2^{k'+1}}\}$  by Claim 5.6. Since any element in  $S_{k''}^{k'}$  has weight at most  $2^{-k''+1}/p$ ,  $W_{k''}^{k'} < \frac{3000 \log N}{2^{k''} p}, \frac{3n_{k''}}{2^{k''+k'+1}}$ . Since  $w_e \geq \frac{1}{2^k p}$  and  $k'' \geq k$ ,  $\frac{3000 \log N}{2^{k''} p} \leq 3000 \log N w_e$ . Moreover, since  $\frac{n_{k''}}{2^{k''}} \leq 1000 \log N$ ,  $\frac{3n_{k''}}{2^{k''+k'+1}} \leq \frac{3000 \log N}{2^{k'+1} p} \leq 3000 \log N w_e$ . So  $W_{k''}^{k'} < 3000 \log N w_e$  for any  $k'' \geq k$ , which means for any  $1 \leq i \leq t$ ,  $e_i^{k'} = e$  with probability at least  $\frac{1}{3000(\gamma+1) \log N}$ . Since  $t = 10^4 \gamma \log N$ , with high probability, there exists one  $i$  such that  $e_i^{k'} = e$  and the algorithm will output an element that gets sampled.

So in both cases, the algorithm succeeds with high probability.  $\square$

## 6 A linear sketching algorithm for weighted spectral sparsification

In this section, we present a linear sketch with  $\tilde{O}(n^{6/5} \epsilon^{-4})$  measurements that computes a  $(1 + \epsilon)$ -spectral sparsifier of a weighted graph. Our linear sketch will be an incidence sketch.

This section is structured as follows. First in Section 6.1 we prove a key vertex sampling lemma, which says that a heavy edge in  $G$  is likely heavy in a random vertex-induced subgraph of  $G^{\text{sq}}$ . Then in light of this lemma, in Section 6.2 we present our linear sketch for recovering heavy edges in  $G$ , where we also assume black-box access to two other linear sketches for sparsifying and recovering heavy edges in  $G^{\text{sq}}$ , respectively. Next in Section 6.3, we present our main linear sketching algorithm for weighted spectral sparsification using the heavy edge recovery sketch in Section 6.2. Finally in Section 6.4 we present the linear sketches for sparsifying and recovering heavy edges in  $G^{\text{sq}}$  that we invoke in Section 6.2.

### 6.1 A vertex sampling lemma

In this subsection we prove a key vertex sampling lemma. This lemma will enable us to recover heavy edges in  $G$  by subsampling vertices at rate  $\approx n^{-1/5}$  and then recovering edges in the vertex-induced subgraph of  $G^{\text{sq}}$ .

**Lemma 6.1** (Vertex sampling lemma). *Let  $\eta \in (0, 1)$ . Given a weighted graph  $G$ , let  $C$  be a vertex set obtained by including each vertex in  $G$  with probability  $\frac{\eta}{100n^{1/5}}$  independently. For any edge  $e$  in  $G$  with leverage score  $w_e b_e^T L_G^\dagger b_e \geq \eta$ , conditioned on  $e \in G[C]$ , with probability at least  $1$ , its leverage score in  $G^{\text{sq}}[C]$  satisfies  $w_e^2 b_e^T L_{G^{\text{sq}}[C]}^\dagger b_e \geq 1/1000$ .*

Roughly, our proof of the lemma proceeds as follows: (i) group vertices according to their potentials induced by an electrical flow between the endpoints of  $e$  in  $G$ ; (ii) analyze the structure of the edges in the vertex-induced subgraph based on their weights and the potential difference between their endpoints; (iii) explicitly construct a set of vertex potentials in  $G^{\text{sq}}$  that certifies the heaviness of the edge  $e$ .

*Proof of Lemma 6.1.* Let  $e = (s, t)$  and without loss of generality, assume  $w_e = 1$  (since we could always scale all edge weights simultaneously without changing any leverage scores). Since the leverage score  $\tau_e \geq \eta$ , we also have that the effective resistance  $b_e^T L_G^\dagger b_e \geq \eta$ . We use the electrical network view of the graph  $G$ , and let  $x = \frac{L_G^\dagger b_e}{b_e^T L_G^\dagger b_e} \in \mathbb{R}^n$  be the set of vertex potentials induced by an electrical flow from  $s$  to  $t$  of  $\frac{1}{b_e^T L_G^\dagger b_e} \leq \frac{1}{\eta}$  units. We also assume without loss of generality  $x_t = 0$  (since we could always shift all vertex potentials by  $x_t$  otherwise). Since  $b_e^T L_G^\dagger b_e$  is the effective resistance between  $s$  and  $t$ , we have  $x_s = 1$  by Ohm's law. Moreover, by Fact 4.17, the normalized energy of  $x$  w.r.t.  $s, t$  satisfies that

$$x^T L_G x = \frac{1}{b_e^T L_G^\dagger b_e} \leq 1/\eta.$$

We now partition the vertices other than  $s$  and  $t$  into  $n^{4/5}$  groups based on their potentials. Specifically, the  $i^{\text{th}}$  group  $S_i$  contains all vertices satisfying  $(i - 1) \cdot n^{-4/5} \leq x_u \leq i \cdot n^{-4/5}$ , where we break ties arbitrarily. For an edge  $f = (u, v)$ , we say  $f$  passes through  $S_i$  if  $x_u \leq (i - 1) \cdot n^{-4/5} < i \cdot n^{-4/5} \leq x_v$  or  $x_v \leq (i - 1) \cdot n^{-4/5} < i \cdot n^{-4/5} \leq x_u$ .

**Claim 6.2.** For any  $i \in [n^{4/5}]$ , the total weight of edges that pass through  $S_i$  is at most  $n^{4/5}/\eta$ .

*Proof.* Consider an edge  $f = (u, v)$  that passes through  $S_i$ , and assume without loss of generality  $x_u \geq x_v$ . By Ohm's law, the flow on edge  $f$  is in the direction  $u \rightarrow v$  and has amount  $w_f(x_u - x_v) \geq w_f n^{-4/5}$ . Since the total amount of flow across the cut

$$(\{t\} \cup S_1 \cup \dots \cup S_i, S_{i+1} \cup \dots \cup S_{n^{4/5}} \cup \{s\})$$

is at most  $1/\eta$ , we have

$$\sum_{f: f \text{ passes through } S_i} w_f n^{-4/5} \leq 1/\eta,$$

which means that the total weight of such edges is at most  $n^{4/5}/\eta$ .  $\square$

We now consider what happens when we look at a vertex-induced subgraph  $G[C]$  where  $C$  is obtained by including each vertex (other than  $s, t$ ) with probability  $\frac{\eta}{100n^{1/5}}$ , and then also including  $s, t$ . We say an edge  $f = (u, v)$  is *intermediate* if  $\{u, v\} \cap \{s, t\} = \emptyset$ . We say a group

$S_i$  is *good* if (i) none of the vertices in  $S_i$  gets sampled in  $C$ , and (ii) all intermediate edges that pass through  $S_i$  and have both endpoints in  $C$  have weight at most  $n^{2/5}$ . We say  $S_i$  is *bad* otherwise.

**Claim 6.3.** With probability at least  $2/3$ , the number of good  $S_i$ 's with  $i \in (\frac{1}{4}n^{4/5}, \frac{3}{4}n^{4/5}]$  is at least  $n^{4/5}/20$ .

*Proof.* First, by Markov's inequality, at least  $.8$  fraction of the  $S_i$ 's with  $i \in (\frac{1}{4}n^{4/5}, \frac{3}{4}n^{4/5}]$  have size at most  $10n^{1/5}$ . For any fixed  $S_i$  with  $|S_i| \leq 10n^{1/5}$ , we have

$$\mathbb{E}[|S_i \cap C|] \leq \frac{|S_i|\eta}{100n^{1/5}} \leq \frac{\eta}{10} \leq \frac{1}{10}.$$

Therefore, once again by Markov's inequality, the probability that (i) happens for any fixed  $S_i$  with  $|S_i| \leq 10n^{1/5}$  is at least  $.9$ .

On the other hand, by Claim 6.2, the total number of edges with weight  $> n^{2/5}$  that pass through  $S_i$  is at most  $n^{2/5}/\eta$ . The probability of any intermediate edge belonging to  $G[S]$  is  $\left(\frac{\eta}{100n^{1/5}}\right)^2 = \frac{\eta^2}{10000n^{2/5}}$ . These combined give us that the expected number of intermediate edges with weight  $> n^{2/5}$  that pass through  $S_i$  and have both endpoints in  $C$  is at most  $1/10000$ . Now an application of Markov's inequality gives that (ii) happens for  $S_i$  with probability at least  $1 - 1/10000$ .

Therefore, by a union bound, each  $S_i$  with  $|S_i| \leq 10n^{1/5}$  is good with probability at least  $.89$ . Thus the expected number of bad  $S_i$ 's with  $|S_i| \leq 10n^{1/5}$  is at most  $.11n^{4/5}$ . By Markov's inequality, the number of bad  $S_i$ 's with  $|S_i| \leq 10n^{1/5}$  is at most  $.33n^{4/5}$  with probability  $\geq 2/3$ . Thus, with probability at least  $2/3$ , the number of good  $S_i$ 's with  $i \in (\frac{1}{4}n^{4/5}, \frac{3}{4}n^{4/5}]$  is at least  $.4n^{4/5} - .33n^{4/5} \geq .05n^{4/5}$ , as desired.  $\square$

We will now construct a set of vertex potentials (call it  $y \in \mathbb{R}^{|C|}$ ) in  $G[C]$  from  $x$ . We will show that the energy of the new set of potentials is small in  $G^{\text{sq}}[C]$  (i.e. even with edge weights squared), but the potential difference between  $s, t$  is still large, which result in a small normalized energy w.r.t.  $s, t$ , and thus certify the “heaviness” of edge  $(s, t)$  in  $G^{\text{sq}}[C]$ .

Specifically, we obtain  $y$  by “collapsing” the vertex potentials within each bad  $S_i$ , so that the intermediate edges that do not pass through any good  $S_i$ 's will have both endpoints getting the same potential. To take care of the edges incident on  $s$  or  $t$  that span less than  $1/4$  fraction of the groups, we will also collapse the vertex potentials in the range  $[0, 1/4]$  and  $[3/4, 1]$ . Precisely,  $y_u$  is given as follows for each  $u \in C$ :

1. If  $x_u \geq 3/4$ , then set  $y_u \leftarrow 1$ .
2. Otherwise, suppose  $x_u \in S_i$  for some  $i \leq \frac{3}{4}n^{4/5}$ . Count the number of good  $S_j$ 's with  $j \in (\max\{i, \frac{1}{4}n^{4/5}\}, \frac{3}{4}n^{4/5}]$  and let  $k$  denote that number. Then set  $y_u \leftarrow 1 - kn^{-4/5}$ .

**Claim 6.4.**  $y$  satisfies the following properties:

1. For any  $u, v$ ,  $|y_u - y_v| \leq |x_u - x_v|$ .
2. With probability at least  $2/3$ ,  $y_s - y_t \geq 1/20$ .

3. With probability at least .8,  $y^T L_{G^{\text{sq}}[C]} y \leq 2$ .

*Proof.* Note that for any  $i$ ,  $C \cap S_i \neq \emptyset$  implies that  $S_i$  is bad. Therefore, by our construction, all vertices within the same bad  $S_i$  will end up having the same potentials in  $y$ . As a result, for any  $u, v$ ,  $|y_u - y_v|$  equals  $n^{-4/5}$  times the number of good groups  $S_i$  between  $x_u$  and  $x_v$  such that  $i \in (\frac{1}{4}n^{4/5}, \frac{3}{4}n^{4/5}]$ , which implies 1.

By Claim 6.3, with probability 2/3, the number of good  $S_i$ 's with  $i \in (\frac{1}{4}n^{4/5}, \frac{3}{4}n^{4/5}]$  is at least  $n^{4/5}/20$ . Thus we have 2.

We then prove 3. First note that for edges that do not pass through any good  $S_i$ , both their endpoints have the same potential in  $y$ . So the total energy contributed by these edges is zero. Now the remaining edges can be divided into two types: (A) edges that are incident on  $s$  or  $t$ ; (B) intermediate edges that pass through some good  $S_i$ . By the definition of good  $S_i$ 's, edges of type (B) have weight at most  $n^{2/5}$  each.

For edges of type (A), they are of the form  $(s, u)$  or  $(v, t)$ . If  $x_u \geq 3/4$  or  $x_v \leq 1/4$ , then once again both endpoints of the edge have the same potential in  $y$ , and the energy contribution from such edges is zero. We thus focus on the edges of type (A) such that  $x_u < 3/4$  or  $x_v > 1/4$ , and refer to those edges as type (A'). For any type (A') edge  $f = (a, b)$ , we have  $|x_a - x_b| \geq 1/4$ , and thus by Ohm's law the amount of flow on  $f$  is at least  $w_f/4$ . Since the total amount of flow going out of  $s$  or going into  $t$  is upper bounded by  $1/\eta$ , the total weight of type (A') edges is at most  $8/\eta$ . This means that the number of type (A') edges  $f$  with  $w_f > 1$  is at most  $8/\eta$ . Thus the expected number of such edges in  $G[C]$  is at most  $(8/\eta) \frac{\eta}{100n^{1/5}} \leq \frac{1}{10}$ . By Markov's inequality, none of such edges is in  $G[C]$  with probability at least .9 (call this event  $\mathcal{E}_1$ ). As for any type (A') edge  $f = (a, b)$  with  $w_f \leq 1$ , we have  $w_f^2(y_a - y_b)^2 \leq w_f(x_a - x_b)^2$ . This combined with the fact that each  $f$  belongs to  $G[C]$  with probability  $\frac{\eta}{100n^{1/5}}$ , the expected contribution of type (A') edges with weight at most 1 to  $y^T L_{G^{\text{sq}}[C]} y$  is at most  $\frac{\eta}{100n^{1/5}} \cdot x^T L_G x \leq \frac{1}{100n^{1/5}} \leq \frac{1}{100}$ . Thus this contribution does not exceed 1 with probability .99 (call this event  $\mathcal{E}_2$ ).

Finally we consider type (B) edges. Since their weights are at most  $n^{2/5}$  each, squaring the edge weights blows up the energy on them by at most a factor of  $n^{2/5}$ . On the other hand, since these are intermediate edges, the probability that any such edge belongs to  $G[C]$  is  $\left(\frac{\eta}{100n^{1/5}}\right)^2 = \frac{\eta^2}{10000n^{2/5}}$ . Therefore, the expected total contribution of type (B) edges to  $y^T L_{G^{\text{sq}}[C]} y$  is at most  $\frac{\eta^2}{10000n^{2/5}} \cdot n^{2/5} x^T L_G x \leq \frac{\eta}{10000} \leq \frac{1}{10000}$ . Thus this contribution does not exceed 1 with probability  $1 - 1/10000$  (call this event  $\mathcal{E}_3$ ).

By a union bound,  $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$  simultaneously happen with probability at least .8, in which case we have  $y^T L_{G^{\text{sq}}[C]} y \leq 2$ .  $\square$

By Claim 6.4 and a union bound over the events in the claim, we have with probability at least .1 that the normalized energy of  $y$  with respect to  $s, t$  in  $G^{\text{sq}}[C]$  is  $\frac{y^T L_{G^{\text{sq}}[C]} y}{(y_s - y_t)^2} \leq 800$ .  $\square$

## 6.2 Recovery of heavy edges

Armed with the vertex sampling lemma, we are now ready to design a linear sketch to recover all heavy edges in  $G$ . In doing so, we will also need to invoke two other linear sketches for sparsifying and recovering heavy edges in  $G^{\text{sq}}$ , respectively. We summarize their performance in the two lemmas below, and prove them later in Section 6.4. We note that the sketch designed in Lemma 6.6 is basically a direct application of  $\ell_2$ -heavy hitters. The sketch designed in Lemma 6.5

is essentially a reduction from sparsification to heavy edge recovery, which will be very similar to our main linear sketching algorithm for weighted spectral sparsification in the next subsection (Section 6.3).

**Lemma 6.5.** *For any parameter  $\epsilon_2 > 0$  and any integer  $n$ , there exists a linear sketch with sketching matrix  $S^{\text{sq}} \in \mathbb{R}^{n\epsilon_2^{-4}\text{polylog}(n,\epsilon_2^{-1},\frac{w_{\max}}{w_{\min}}) \times \binom{n}{2}}$  and recovery algorithm **SQUARERECOVERY** such that, given an input graph  $G$  of  $n$  vertices with weight vector  $w_G$ , **SQUARERECOVERY**( $S^{\text{sq}}w_G$ ) returns a  $(1 + \epsilon_2)$ -spectral sparsifier of  $G^{\text{sq}}$  with high probability.*

**Lemma 6.6.** *For any parameters  $\eta_3, \epsilon_3 \in (0, 1)$  and any integer  $n$ , there exists a linear sketch with sketching matrix  $S^{\text{sqh}} \in \mathbb{R}^{n\eta_3^{-1}\epsilon_3^{-2}\text{polylog}(n) \times \binom{n}{2}}$  and recovery algorithm **SQUAREHEAVYRECOVERY** such that, for an input graph  $G$  of  $n$  vertices with weight vector  $w_G$  and another graph  $\tilde{G}$ , **SQUAREHEAVYRECOVERY**( $S^{\text{sqh}}w_G, \tilde{G}$ ) recovers a set of edges  $F$  in  $G$  along with estimates of their weights  $\tilde{w}_f$ 's such that with high probability*

1.  $F$  contains all edges  $e$  satisfying

$$\frac{((w_G)_e b_e^T L_{\tilde{G}}^\dagger b_e)^2}{b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \geq \eta_3.$$

2. All edges  $f \in F$  satisfy  $\frac{1}{1+\epsilon_3}(w_G)_f \leq \tilde{w}_f \leq (1 + \epsilon_3)(w_G)_f$ .

We remark that to understand the first guarantee of the above lemma, one should think of  $\tilde{G}$  as a good spectral sparsifier of  $G^{\text{sq}}$ , so that the numerator  $((w_G)_e b_e^T L_{\tilde{G}}^\dagger b_e)^2 \approx ((w_G)_e b_e^T L_{G^{\text{sq}}}^\dagger b_e)^2$  and the denominator  $b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e \approx b_e^T L_{G^{\text{sq}}}^\dagger b_e$ , and thus the LHS  $\approx (w_G)_e^2 b_e^T L_{G^{\text{sq}}}^\dagger b_e$ , the leverage score of  $e$  in  $G^{\text{sq}}$ . That is, this guarantee is essentially saying that all heavy edges in  $G^{\text{sq}}$  will be recovered.

We now describe the linear sketch for recovering heavy edges in  $G$  in Figure 4, and characterize its performance in the lemma below.

**Lemma 6.7.** *For any parameters  $\eta_1 \in (n^{-4/5} \log n, 1)$ ,  $\epsilon_1 \in (0, 1)$  and integer  $n$ , there exists a linear sketch with sketching matrix  $S^{\text{hv}} \in \mathbb{R}^{n^{6/5}\eta_1^{-1}\epsilon_1^{-2}\text{polylog}(n,\frac{w_{\max}}{w_{\min}}) \times \binom{n}{2}}$  and recovery algorithm **HEAVYEDGERECOVERY** such that, for an input graph  $G$  of  $n$  vertices with weight vector  $w_G$ , **HEAVYEDGERECOVERY**( $S^{\text{hv}}w_G$ ) recovers a set  $F$  of edges in  $G$  along with estimates of their weights  $\tilde{w}_f$ 's such that with high probability*

1. All edges  $e$  whose leverage score in  $G$  satisfy  $(w_G)_e b_e^T L_G^\dagger b_e \geq \eta_1$  belong to  $F$ .
2. All edges  $f \in F$  satisfy  $\frac{1}{1+\epsilon_1}(w_G)_f \leq \tilde{w}_f \leq (1 + \epsilon_1)(w_G)_f$ .

We now prove Lemma 6.7 using Lemmas 6.1, 6.5, 6.6.

*Proof of Lemma 6.7 using Lemmas 6.1, 6.5, 6.6.* **Number of linear measurements.** We observe that with high probability,  $|V_i| \leq 100\eta_1 n^{4/5}$  for all  $i \in [t]$ , and thus each  $S_i^{\text{sq}}w_G[V_i] \in$

HEAVYEDGESKETCH( $G, \eta_1, \epsilon_1$ )

1. Let  $t = \lceil 10000\eta_1^{-2}n^{2/5} \log n \rceil$  and let  $V_1, \dots, V_t$  be vertex subsets, each obtained by including each vertex in  $G$  with probability  $\frac{\eta_1}{100n^{1/5}}$  independently.  
(Subsample the vertices sufficiently many times to cover every edge)
2. For each  $i \in [t]$ , let  $S_i^{\text{sq}} \in \mathbb{R}^{|V_i|\epsilon_2^{-4}\text{polylog}(|V_i|, \epsilon_2^{-1}, \frac{w_{\max}}{w_{\min}}) \times \binom{|V_i|}{2}}$  be a sketching matrix with error  $\epsilon_2 = .01$  (Lemma 6.5), and let  $S_i^{\text{sqh}} \in \mathbb{R}^{|V_i|\eta_3^{-1}\epsilon_3^{-2}\text{polylog}(|V_i|) \times \binom{|V_i|}{2}}$  with  $\eta_3 = 1/2000$  and  $\epsilon_3 = \epsilon_1/100$  (Lemma 6.6).
3. Concatenate the following sketches as  $S^{\text{hv}} w_G$ :
  - (a)  $S_1^{\text{sq}} w_{G[V_1]}, \dots, S_t^{\text{sq}} w_{G[V_t]}$ , where  $w_{G[V_i]} \in \mathbb{R}^{\binom{|V_i|}{2}}$  is the weight vector of the vertex induced subgraph  $G[V_i]$ .  
(Create a sparsification sketch for each vertex-induced subgraph)
  - (b)  $S_1^{\text{sqh}} w_{G[V_1]}, \dots, S_t^{\text{sqh}} w_{G[V_t]}$ .  
(Create a heavy edge sketch for each vertex-induced subgraph)

HEAVYEDGERECOVERY( $S^{\text{hv}} w_G$ )

1. For each  $i \in [t]$ :
  - (a) Use Lemma 6.5 to recover from  $S_i^{\text{sq}} w_{G[V_i]}$  a sparsifier  $\tilde{G}_i$  of  $G^{\text{sq}}[V_i]$ .  
( $\tilde{G}_i$  is a 1.01-spectral sparsifier of  $G^{\text{sq}}[V_i]$ )
  - (b) Use the recovery algorithm from Lemma 6.6 to recover from  $S_i^{\text{sqh}} w_{G[V_i]}$  and  $\tilde{G}_i$  a set  $F$  of edges along with their estimated weights  $\tilde{w}_f$ 's, and mark all edges in  $F$  as heavy.  
(Feed the sparsifier  $\tilde{G}_i$  to the recovery algorithm to get all heavy edges)
2. Return all edges marked heavy along with the estimates of their weights (if for some edge there are multiple estimates of its weight, pick an arbitrary one).

Figure 4: Linear sketch for recovering heavy edges in  $G$ .

$\mathbb{R}^{\eta_1 n^{4/5} \text{polylog}(n, \frac{w_{\max}}{w_{\min}})}$  and each  $S_i^{\text{sqh}} w_G[V_i] \in \mathbb{R}^{\eta_1 n^{4/5} \epsilon_1^{-2} \text{polylog}(n)}$ . Therefore, the total number of linear measurements is bounded by

$$t \cdot \eta_1 n^{4/5} \epsilon_1^{-2} \text{polylog}(n, \frac{w_{\max}}{w_{\min}}) \leq n^{6/5} \eta_1^{-1} \epsilon_1^{-2} \text{polylog}(n, \frac{w_{\max}}{w_{\min}}).$$

**Guarantee 1.** Consider fixing any edge  $e$  with  $w_e b_e^T L_G^\dagger b_e \geq \eta_1$ . By Lemma 6.1, for each  $i \in [t]$ , with probability  $\frac{\eta_1^2}{1000n^{2/5}}$ , we have  $e \in G[V_i]$  and  $w_e^2 b_e^T L_{G^{\text{sq}}[V_i]}^\dagger b_e \geq 1/1000$  (call this event  $\mathcal{E}_i$ ). Therefore at least one of  $\mathcal{E}_1, \dots, \mathcal{E}_t$  happens with probability  $\geq 1 - (1 - \frac{\eta_1^2}{1000n^{2/5}})^t \geq 1 - 1/n^{-10}$ . Whenever an  $\mathcal{E}_i$  happens, using the fact that  $\tilde{G}_i$  is a 1.01-spectral sparsifier of  $G^{\text{sq}}[V_i]$  (by Lemma 6.5), we have

$$(w_G)_e b_e^T L_{\tilde{G}_i}^\dagger b_e \geq \frac{1}{1.01} (w_G)_e b_e^T L_{G^{\text{sq}}[V_i]}^\dagger b_e$$

and

$$b_e^T L_{\tilde{G}_i}^\dagger L_{G^{\text{sq}}[V_i]} L_{\tilde{G}_i}^\dagger b_e \leq 1.01 b_e^T L_{\tilde{G}_i}^\dagger L_{\tilde{G}_i} L_{\tilde{G}_i}^\dagger b_e = b_e^T L_{\tilde{G}_i}^\dagger b_e \leq 1.01^2 b_e^T L_{G^{\text{sq}}[V_i]}^\dagger b_e,$$

and thus

$$\frac{((w_G)_e b_e^T L_{\tilde{G}_i}^\dagger b_e)^2}{b_e^T L_{\tilde{G}_i}^\dagger L_{G^{\text{sq}}[V_i]} L_{\tilde{G}_i}^\dagger b_e} \geq 1.01^{-3} (w_G)_e^2 b_e^T L_{G^{\text{sq}}[V_i]}^\dagger b_e \geq 1.01^{-3}/1000 \geq 1/2000.$$

By Lemmas 6.6, with high probability,  $e$  is among the recovered edges. Therefore by a union bound over all such edges  $e$ , we have the desired result.

**Guarantee 2.** This follows directly from Lemma 6.6.  $\square$

### 6.3 Main algorithm for weighted spectral sparsification

We now show how to use the heavy edge recovery sketch in the previous section to obtain a spectral sparsifier of  $G$ .

We first briefly summarize the main ideas. The first idea is to use the iterative refinement process in [LMP13] as in the previous works on unweighted graphs [KLM<sup>+</sup>14, KMM<sup>+</sup>20]. That is, we consider, for some large  $\alpha = \text{poly}(w_{\max}, n)$ ,  $t = \text{polylog}(\epsilon^{-1}, \frac{w_{\max}}{w_{\min}}, n)$  and a constant  $\beta \in (0, 1)$ , the sequence of graphs

$$G + \alpha K_n, G + \beta \alpha K_n, G + \beta^2 \alpha K_n, \dots, G + \beta^t \alpha K_n,$$

which have the properties that

1.  $\alpha K_n$  is an  $O(1)$ -spectral sparsifier of  $G + \alpha K_n$ .
2.  $G + \beta^k \alpha K_n$  is an  $O(1)$ -spectral sparsifier of  $G + \beta^{k+1} \alpha K_n$  for all  $k \geq 0$ .
3.  $G + \beta^t \alpha K_n$  is a  $(1 + \epsilon)$ -spectral sparsifier of  $G$ .

The idea is then to iteratively obtain a sparsifier of each of these graphs, where we use the sparsifier of  $G + \beta^k \alpha K_n$  to guide the sparsification of  $G + \beta^{k+1} \alpha K_n$  (in particular, we use the sparsifier of the former to estimate the effective resistances and leverage scores in the latter). Thus it boils down to how to sparsify  $G^{(k+1)} := G + \beta^{k+1} \alpha K_n$  using heavy edge recovery.

**Remark 6.1.** We remark that, when given access to a heavy edge recovery sketch, the sparsification of  $G^{(k+1)}$  is relatively easy to achieve in the unweighted case, for the following reason. Consider, in an unweighted graph, an edge  $e = (s, t)$  with effective resistance (thus also leverage score)  $r_e$ , and let  $x_{s,t} \in \mathbb{R}^n$  be the set of vertex potentials induced by an electrical flow from  $s$  to  $t$  and assume w.l.o.g.  $x_s - x_t = 1$ . By Fact 4.17, we have  $x^T L_G x = 1/r_e$ . Now notice that we can also assume  $x_u \in [0, 1]$  for all  $u$ , since letting  $x_u \leftarrow 1$  for all  $x_u > 1$  and  $x_v \leftarrow 0$  for all  $x_v < 0$  can only decrease the total energy.

This means that the energy  $(x_s - x_t)^2 = 1$  on edge  $e = (s, t)$  is the largest among all edges, and thus if we sample all edges uniformly at rate  $\approx r_e$ , the total energy will be  $\tilde{O}(1)$  with high probability by Chernoff bounds. This implies that in the latter subsampled graph,  $e$  (if sampled) becomes heavy with high probability. Since  $r_e$  is exactly (up to an  $O(\log n)$  factor) the probability with which we want to sample  $e$ , we can apply the heavy edge sketch to subgraphs of  $G$  obtained by sampling edges at geometrically decreasing rates, and then try to recover each edge  $e$  from the subgraph with sampling rate  $\approx r_e$ .

However, for weighted graphs, the energy on some other edges of very large weights can be unboundedly big. Thus concentration bounds no longer give us high success probability of recovering an edge  $e$  when sampling all edges uniformly at rate  $\approx \tau_e = w_e r_e$ , even though the energy reduces significantly in expectation.

To sparsify  $G^{(k+1)}$ , we will utilize the spectral sparsification framework in [Kou14], which is itself model oblivious. The framework works as follows:

1. Fix some constant  $p \in (0, 1)$ .
2. While the number of edges in the graph is  $> n\epsilon^{-2}\text{polylog}(n)$ :
  - (a) Find all edges whose leverage score  $\geq \eta := \epsilon^2/\text{polylog}(n)$ , and call these edges  $F$ .
  - (b) Sample each edge not in  $F$  with probability  $p$ , and multiply its weight by  $1/p$  if sampled.

Notice that since the leverage scores of all edges sum up to at most  $n - 1$ , the total number of edges in  $F$  is at most  $n\epsilon^{-2}\text{polylog}(n)$ . Thus, in each while loop iteration, the number of edges decreases by a constant factor, and as a result there can be at most  $O(\log n)$  iterations. Then using Theorem 4.18, we have that the final graph is a  $(1 + \epsilon)$ -spectral sparsifier of  $G^{(k+1)}$ . Notice that here, the first step in the while loop is exactly the recovery of heavy edges.

We now describe the difficulty that arises in implementing the above process using linear sketching *non-adaptively*, and our way around it. First, let  $E_0 \supseteq E_1 \supseteq \dots \supseteq E_{O(\log n)}$  be such that  $E_0 = \binom{V}{2}$  and  $E_{i+1}$  is obtained by subsampling each edge slot in  $E_i$  with probability  $p$  (the constant fixed at the first step of the above process). We apply the heavy edge recovery sketch to each  $G^{(k+1)}[E_i]$ . We then implement each iteration of the while loop in the above process.

At first, we recover all heavy edges in  $G^{(k+1)}$  using the sketch of  $G^{(k+1)}[E_0] = G^{(k+1)}$ , and call these edges  $F_0$ . We would like to sample each edge in  $G^{(k+1)}[\binom{V}{2} \setminus F_0]$  with probability

$p$ , and multiply its weight by  $1/p$  if sampled. Then in the next iteration, we would want to recover heavy edges in the latter subsampled graph. That is, we would like to have a sketch of the graph  $(1/p)G^{(k+1)}[E_1 \setminus F_0] + G^{(k+1)}[F_0]$ . However, we only have a sketch of  $G^{(k+1)}[E_1]$ . By linearity, we can multiply it by  $1/p$  and add to it  $G^{(k+1)}[F_0 \setminus E_1]$  and get a sketch of  $(1/p)G^{(k+1)}[E_1] + G^{(k+1)}[F_0 \setminus E_1]$ . Nonetheless, this sketch is still not taken on our desired graph  $(1/p)G^{(k+1)}[E_1 \setminus F_0] + G^{(k+1)}[F_0]$ , since the weights of the edges in  $E_1 \cap F_0$  in the former graph are larger than in the latter by a factor of  $1/p$ . We say these edges are *overweighted* by  $1/p$ , and call the former graph *overweighted* graph.

One might hope to further subtract from the sketch  $(1/p - 1)G^{(k+1)}[E_1 \cap F_0]$  to bring down the weights of the overweighted edges by a factor of  $(1/p)$ . However, notice that we do *not* have the exact weights of the edges in  $F_0$  from our heavy edge recovery sketch. Rather, we only have some estimates of their weights. Moreover, while in the second iteration we are only looking to subtract edges that are overweighted by  $1/p$ , in subsequent iterations, we might need to subtract edges that are overweighted by  $\text{poly}(n)$ , which means that our weight estimates for such edges must have inverse polynomial accuracy for the subtraction to work.

Our way around this issue is to repeatedly *re-estimate* the weights of the overweighted edges. Specifically, we show that the edges that are overweighted the most must be heavy in the overweighted graph. Thus we can apply the heavy edge recovery sketch to the overweighted graph, get estimates of the weights of these edges, and bring their weights down by a factor of  $(1/p)$ . We then repeatedly apply this step  $O(\log n)$  times (where we re-estimate the weights each time) until there are no overweighted edges. Since we only bring down the edge weights by a constant factor each time and always re-estimate the weights once changed, we will never have too large an error.

We now present in Figure 5 our main algorithm for weighted spectral sparsification, which invokes the heavy edge recovery sketch in Section 6.2. Specifically, for each graph  $G^{(k)}$  in the iterative refinement process, we apply independent heavy edge recovery sketches to subgraphs of  $G^{(k)}$  obtained by sampling edges at geometrically decreasing rates. Then in the recovery step, we first simulate the iterative refinement process using an outer for loop of  $k$ , and then implement the framework from [Kou14] in an inner for loop of  $i$ . Inside each iteration of the inner for loop, we start with the sketch of an overweighted graph  $Z$ , and then gradually bring down the weights of the overweighted edges by repeatedly recovering heavy edges in the current (overweighted) graph and subtracting a constant fraction of their weights. Finally when there are no overweighted edges left, we recover the heavy edges in the resulting graph and then go to the next sampling rate.

The performance of our main linear sketching algorithm for weighted graph sparsification is characterized in Theorem 6.8.

**Theorem 6.8.** *For any parameter  $\epsilon > 0$  and any integer  $n$ , there exists a linear sketch with sketching matrix  $S^{\text{ss}} \in \mathbb{R}^{n^{6/5}\epsilon^{-4}\text{polylog}(n,\epsilon^{-1},\frac{w_{\max}}{w_{\min}}) \times \binom{n}{2}}$  and recovery algorithm SPECTRALRECOVERY such that, given an input graph  $G$  of  $n$  vertices with weight vector  $w_G$ , SPECTRALRECOVERY( $S^{\text{ss}}w_G$ ) returns a  $(1 + \epsilon)$ -spectral sparsifier of  $G$  with high probability.*

Before proving the theorem, we shall first give some useful intermediate lemmas. The following proposition directly follows from the definition of spectral sparsifiers.

**Proposition 6.9.** *For any two graphs  $G_1, G_2$  whose weight vectors satisfy that  $(w_{G_1})_e \approx_{1+\epsilon} (w_{G_2})_e$  for all  $e \in \binom{V}{2}$ ,  $G_1$  is a  $(1 + \epsilon)$ -spectral sparsifier of  $G_2$ .*

SPECTRALSKETCH( $G, \epsilon$ )

1. Let  $t = \lceil 10000 \log(10^6 \epsilon^{-1} \frac{w_{\max}}{w_{\min}} \cdot n^{10}) \rceil$ , and then for each  $k = 0, 1, 2, \dots, t$ :
  - (a) Let  $E_0^{(k)} \supseteq E_1^{(k)} \supseteq \dots \supseteq E_t^{(k)}$  be edge subsets where  $E_0^{(k)} = \binom{V}{2}$  and  $E_i^{(k)}$  is obtained by sub-sampling each edge slot in  $E_{i-1}^{(k)}$  with probability  $(1 + 1/1000)^{-1}$ .
  - (b) For each pair  $0 \leq i, j \leq t$ , use Lemma 6.7 to generate a sketching matrix  $(S^{\text{hv}})_{i,j}^{(k)} \in \mathbb{R}^{O(n^{6/5} \eta_1^{-1} \epsilon_1^{-2} \text{polylog}(n, \frac{w_{\max}}{w_{\min}})) \times \binom{n}{2}}$  with  $\eta_1 = \epsilon^2 / (10^{12} t^2 \log n)$ ,  $\epsilon_1 = \epsilon / (10^6 t)$ .
2. For each  $0 \leq k, i, j \leq t$ , let  $G^{(k)} \leftarrow G + (1 + 1/10^4)^{-k} 10^6 w_{\max} n^5 K_n$ , and compute the sketch  $(S^{\text{hv}})_{i,j}^{(k)} w_{G^{(k)}[E_i^{(k)}]}$ , where  $w_{G^{(k)}[E_i^{(k)}]} \in \mathbb{R}^{\binom{n}{2}}$  is the weight vector of  $G^{(k)}[E_i^{(k)}]$ .  
(Take sufficiently many independent heavy edge sketches on each subsampled graph)
3. Concatenate these sketches as  $S^{\text{ss}} w_G$ .

$H^{(t)} = \text{SPECTRALRECOVERY}(S^{\text{ss}} w_G)$

1. Initially, let  $H^{(0)} \leftarrow 10^6 w_{\max} n^5 K_n$ . ( $H^{(0)}$  is a 1.001-spectral sparsifier of  $G^{(0)}$ )
2. For  $k = 1, 2, \dots, t$ :
  - (a) Let  $H^{(k)} \leftarrow \emptyset$ . ( $H^{(k)}$  will be a  $(1 + \epsilon/1000)$ -spectral sparsifier of  $G^{(k)}$ )
  - (b) Set  $c_e \leftarrow 0$  for all  $e \in \binom{V}{2}$ . ( $c_e$  will be s.t.  $e$  is added to  $H^{(k)}$  when  $i = c_e$ )
  - (c) For  $i = 0, 1, \dots, t$ :
    - i. Let  $Z \leftarrow (1 + 1/1000)^i G^{(k)}[E_i^{(k)}] + H^{(k)}[\binom{V}{2} \setminus E_i^{(k)}]$ .  
( $Z$  records the graph on which our linear sketches are currently taken)
    - ii. Compute sketches  $s_j := (S^{\text{hv}})_{i,j}^{(k)} w_Z$ ,  $j \in [0, t]$ , where  $w_Z$  is  $Z$ 's weight vector.
    - iii. For each  $f \in H^{(k)} \cap E_i^{(k)}$ , let  $\delta_f \leftarrow i - c_f$ , and let  $\delta_f \leftarrow 0$  for all other edges.  
(( $w_Z$ )<sub>f</sub> needs to be brought down by a factor of  $(1 + 1/1000)^{\delta_f}$ )
    - iv. Let  $j \leftarrow 0$ . Then while  $\exists f : \delta_f > 0$ , do the following:
      - Use Lemma 6.7 to recover from  $s_j$  a set  $F$  of edges and then let  $j \leftarrow j + 1$ .
      - For each  $f \in F$  such that  $\delta_f > 0$ , let  $\tilde{w}_f$  be the estimate of its weight:
        - $Z \leftarrow Z - (1 - (1 + 1/1000)^{-1}) \tilde{w}_f f$ .
        - $s_{j'} \leftarrow s_{j'} - (1 - (1 + 1/1000)^{-1}) (S^{\text{hv}})_{i,j'}^{(k)} (\tilde{w}_f \chi_f)$  for all  $j' \in [0, t]$ .
        - $\delta_f \leftarrow \delta_f - 1$ .(Bring down ( $w_Z$ )<sub>f</sub> by  $(1 + 1/1000)$  and update all sketches accordingly)
    - v. Use Lemma 6.7 to recover from  $s_j$  a set  $F^*$  of edges.
    - vi. For each edge  $f \in F^*$  with estimated weight  $\tilde{w}_f$  such that  $\tilde{w}_f b_f^T L_{H^{(k-1)}}^\dagger b_f \geq 8\eta_1$  and  $f$  is not already in  $H^{(k)}$ , add  $f$  to  $H^{(k)}$  with weight  $\tilde{w}_f$ , and let  $c_f \leftarrow i$ .  
(Add recovered heavy edges to  $H^{(k)}$ , then go to the next sampling rate)

Figure 5: Linear sketch for weighted spectral sparsification.

This proposition then immediately implies the following two lemmas.

**Lemma 6.10.**  $10^6 w_{\max} n^5 K_n$  is a 1.001-spectral sparsifier of  $G^{(0)}$ .

**Lemma 6.11.** For all  $k \geq 1$ ,  $G^{(k-1)}$  is a 1.001-spectral sparsifier of  $G^{(k)}$ .

**Lemma 6.12.**  $G^{(t)}$  is a  $(1 + \epsilon/2)$ -spectral sparsifier of  $G$ .

*Proof.* By definition

$$\begin{aligned} L_{G^{(t)}} &= L_G + L_{(1+10^{-4})^{-t} 10^6 w_{\max} n^5 K_n} \\ &\preceq L_G + .1\epsilon w_{\min} n^{-5} L_{K_n}, \end{aligned}$$

where the last line follows from our choice of  $t$ . Thus, the largest eigenvalue of the second term is bounded by  $.1\epsilon w_{\min} n^{-4}$ . By standard lower bounds on the second smallest eigenvalue, the second smallest eigenvalue of  $L_G$  is at least  $w_{\min}/n^2$ . Therefore we have

$$-.1\epsilon L_G \preceq L_{G^{(t)}} - L_G \preceq .1\epsilon L_G,$$

which implies that  $G^{(t)}$  is a  $(1 + \epsilon/2)$ -spectral sparsifier of  $G$ .  $\square$

Fix an iteration of the outer for loop of  $k$ . Then for an iteration of the inner for loop of  $i$ , let  $H_i^{(k)}$  be the  $H^{(k)}$  at the *beginning* of the iteration, and let  $F_i^{(k)}$  be the edges in  $H_i^{(k)}$ . Define graph

$$J_i^{(k)} := (1 + 1/1000)^i G^{(k)}[E_i^{(k)} \setminus F_i^{(k)}] + \sum_{\ell=0}^{i-1} (1 + 1/1000)^\ell G^{(k)}[F_{\ell+1}^{(k)} \setminus F_\ell^{(k)}].$$

Notice that  $F_0^{(k)} = \emptyset$  and  $J_0^{(k)} = G^{(k)}$ . Also by the way we are assigning values to  $c_f$  in the algorithm, we have, at the beginning of the for loop (of  $i$ ) iteration,  $f \in F_{c_f+1}^{(k)} \setminus F_{c_f}^{(k)}$  for all  $f \in F_i^{(k)}$ . Thus we also have

$$J_i^{(k)} := (1 + 1/1000)^i G^{(k)}[E_i^{(k)} \setminus F_i^{(k)}] + \sum_{f \in F_i^{(k)}} (1 + 1/1000)^{c_f} (w_{G^{(k)}})_f f. \quad (13)$$

**Lemma 6.13.** Suppose  $H^{(k-1)}$  is a 1.001-spectral sparsifier of  $G^{(k-1)}$ . Then with high probability, for all  $0 \leq i < t$ ,

1. After the while loop inside the  $i^{\text{th}}$  iteration terminates, for all  $f$ ,

$$\frac{1}{1 + \epsilon/10000} (w_{J_i^{(k)}})_f \leq (w_Z)_f \leq (1 + \epsilon/10000) (w_{J_i^{(k)}})_f.$$

2. For all  $f \in F_{i+1}^{(k)} \setminus F_i^{(k)}$ ,

$$\frac{1}{1 + \epsilon/(10^6 t)} (w_{J_{i+1}^{(k)}})_f \leq (w_{H_{i+1}^{(k)}})_f \leq (1 + \epsilon/(10^6 t)) (w_{J_{i+1}^{(k)}})_f.$$

3. All edges in  $F_{i+1}^{(k)}$  have leverage scores in  $J_i^{(k)}$  at least  $4\eta_1$ .

4.  $J_{i+1}^{(k)}$  is a  $(1 + \epsilon/(10^4 t))$ -spectral sparsifier of  $J_i^{(k)}$ .

*Proof.* We prove all statements of this lemma by induction on  $i$ . For  $i = 0$ , since  $\delta_f = 0$  for all  $f$ , the while loop will not execute. Thus throughout this iteration we have  $Z = J_0^{(k)} = G^{(k)}$ . This immediately gives 1. By Lemma 6.7, the  $F^*$  we recover in this iteration contains all edges whose leverage score in  $G^{(k)}$  is at least  $\eta_1$ , and all edges in  $F^*$  have weight estimates satisfying 2. Since  $H^{(k-1)}$  is a 1.001-spectral sparsifier of  $G^{(k-1)}$ , and  $G^{(k-1)}$  is in turn a 1.001-spectral sparsifier of  $G^{(k)}$ , we have that  $H^{(k-1)}$  is a 1.003-spectral sparsifier of  $G^{(k)}$ . As a result, we know that, at the last step of the for loop iteration, all edges with leverage score at least  $\geq 10\eta_1$  in  $G^{(k)}$  will be added to  $H^{(k)}$ , and all edges added to  $H^{(k)}$  have leverage score at least  $\geq 4\eta_1$  in  $G^{(k)}$ , so we have 3. This means that  $J_1^{(k)}$  is obtained by sampling a set of edges in  $J_0^{(k)}$  whose leverage scores in  $J_0^{(k)}$  are at most  $10\eta_1$  with probability  $(1 + 1/1000)^{-1}$ , and multiply their weights by  $(1 + 1/1000)$  if sampled. Using Theorem 4.18, we have 4.

We now do an inductive step. Suppose all four statements hold for iterations  $0, 1, \dots, i - 1$  where  $1 < i < t$ . We show that they also hold for iteration  $i$ . We first need to analyze the while loop inside iteration  $i$ . Let us number a while loop iteration by the value of  $j$  at the *end* of the iteration.

**Claim 6.14.** At the end of while loop iteration  $j$  where  $j \leq t$ , we have for all  $f \in E_i^{(k)} \cap F_i^{(k)}$

$$\frac{1}{(1 + 2\epsilon_1)^j} \cdot (1 + 1/1000)^{\delta_f} (w_{J_i^{(k)}})_f \leq (w_Z)_f \leq (1 + 2\epsilon_1)^j (1 + 1/1000)^{\delta_f} (w_{J_i^{(k)}})_f.$$

*Proof.* We prove this claim by an induction on  $j$ . First we show that the statement is true for  $j = 0$  at the beginning of while loop iteration 1. Here all  $f \in E_i^{(k)} \cap F_i^{(k)}$  satisfy that  $(w_Z)_f = (1 + 1/1000)^i (w_{G^{(k)}})_f$ . Since we set  $\delta_f \leftarrow i - c_f$  before the while loop, and by (13)  $(w_{J_i^{(k)}})_f = (1 + 1/1000)^{c_f} (w_{G^{(k)}})_f$ , we have  $(w_Z)_f = (1 + 1/1000)^{\delta_f} (w_{J_i^{(k)}})_f$ , as desired.

Now suppose the statement is true at the end of iteration  $j - 1$  where  $1 < j \leq t$ . We then show that the statement is also true at the end of iteration  $j$ . Let  $Z_0$  be the  $Z$  before our updates to  $Z$  in iteration  $j$  and let  $Z_1$  be the  $Z$  after our updates. By Lemma 6.7, all edges recovered  $f \in F$  have their estimated edge weights  $\tilde{w}_f \in [\frac{1}{1+\epsilon_1}(w_{Z_0})_f, (1+\epsilon_1)(w_{Z_0})_f]$ . Therefore after our updates, we have for any  $f \in F$  such that  $\delta_f > 0$  that  $(w_{Z_1})_f \in [\frac{1}{1+2\epsilon_1}(1+1/1000)^{-1}(w_{Z_0})_f, (1+2\epsilon_1)(1+1/1000)^{-1}(w_{Z_0})_f]$ , and  $(w_{Z_1})_f = (w_{Z_0})_f$  for other edges  $f$ . Since we let  $\delta_f \leftarrow \delta_f - 1$  for such edges, and do not change the  $\delta_f$ 's of other edges, we have our desired statement for  $j$ .  $\square$

**Claim 6.15.** The while loop terminates after at most  $t$  iterations.

*Proof.* It suffices to show that  $\max_f \delta_f$  decreases by 1 in each while loop iteration. Since  $\delta_f \leq t$  for any  $f$ , this will imply that there can be at most  $t$  iterations. Then it boils down to showing that for all  $f^*$  with  $\delta_{f^*} = \max_f \delta_f$ ,  $f^*$  belongs to the recovered edge set  $F$ . Since  $f^* \in F_i^{(k)}$ , by 3 of our induction hypothesis, the leverage score of  $f^*$  in  $J_{i-1}^{(k)}$  is at least  $4\eta_3$ . Notice that by 4 of our induction hypothesis,  $J_{i-1}^{(k)}$  is a  $(1 + 1/1000)$ -spectral sparsifier of  $G^{(k)}$ . Then using

the fact that  $H^{(k-1)}$  is a 1.003-spectral sparsifier of  $G^{(k)}$  (which we proved at the beginning of the proof of this lemma), we have that  $H^{(k-1)}$  is a 1.005-spectral sparsifier of  $J_{i-1}^{(k)}$ .

By Claim 6.14, we have at the beginning of each while loop that, for all  $f$ ,

$$\begin{aligned} (w_Z)_f &\in \left[ \frac{1}{(1+2\epsilon_1)^t} (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f, (1+2\epsilon_1)^t (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f \right] \\ &\subseteq \left[ \frac{1}{1.01} (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f, 1.01 (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f \right]. \end{aligned} \quad (14)$$

Since  $\delta_{f^*} \geq \delta_f$  for all  $f$ , the above implies

$$L_Z \preceq 1.03 (1+1/1000)^{\delta_{f^*}} L_{J_{i-1}^{(k)}}. \quad (15)$$

By inverting both sides, we then get

$$L_Z^\dagger \succeq 1.03^{-1} (1+1/1000)^{-\delta_{f^*}} L_{J_{i-1}^{(k)}}^\dagger.$$

Combining this with (14), the leverage score of  $f^*$  in  $Z$  satisfies

$$(w_Z)_{f^*} b_{f^*} L_Z^\dagger b_{f^*} \geq \frac{1}{1.01 \cdot 1.03} (w_{J_i^{(k)}})_{f^*} b_{f^*}^T L_{J_{i-1}^{(k)}}^\dagger b_{f^*} \geq \frac{4\eta_3}{1.01 \cdot 1.03} \geq \eta_3,$$

as desired.  $\square$

By Claim 6.14, after the while loop terminates, we have that for all  $f$ ,

$$\begin{aligned} (w_Z)_f &\in \left[ \frac{1}{(1+2\epsilon_1)^t} (w_{J_i^{(k)}})_f, (1+2\epsilon_1)^t (w_{J_i^{(k)}})_f \right] \\ &\subseteq \left[ \frac{1}{(1+\epsilon/10000)} (w_{J_i^{(k)}})_f, (1+\epsilon/10000) (w_{J_i^{(k)}})_f \right], \end{aligned}$$

and thus we have 1. This also implies that  $Z$  is a  $(1+\epsilon/10000)$ -spectral sparsifier of  $J_i^{(k)}$ , and as result, for each edge  $f$ , its leverage scores in  $Z$  and  $J_i^{(k)}$  are within a  $(1+\epsilon/10000)^2 < 1.01$  factor of each other.

For all edges in  $E_i^{(k)} \setminus F_i^{(k)}$ , their weights in  $Z$  equal exactly their weights in  $J_i^{(k)}$ , therefore by Lemma 6.7, all edges recovered in  $F^*$  not in  $F_i^{(k)}$  have weight estimates satisfying 2.

Notice that by 4 of our induction hypothesis,  $J_i^{(k)}$  is a  $(1+1/1000)$ -spectral sparsifier of  $G^{(k)}$ . Then using the fact that  $H^{(k-1)}$  is a 1.003-spectral sparsifier of  $G^{(k)}$  (which we proved at the beginning of this proof), we have that  $H^{(k-1)}$  is a 1.01-spectral sparsifier of  $Z$ . Thus, at the last step of the for iteration, all edges added to  $H^{(k)}$  have leverage score at least  $\geq 5\eta_1$  in  $Z$ , and all edges with leverage score  $\geq 9\eta_1$  in  $Z$  will be added to  $H^{(k)}$ . Thus we also know that all edges added to  $H^{(k)}$  have leverage score  $\geq 4\eta_1$  in  $J_i^{(k)}$  (which gives 3), and all edges with leverage score  $\geq 10\eta_1$  in  $J_i^{(k)}$  will be added to  $H^{(k)}$ .

The above reasoning also implies that  $J_{i+1}^{(k)}$  is obtained by sampling a set of edges in  $J_i^{(k)}$  whose leverage score is at most  $10\eta_1$  with probability  $(1+1/1000)^{-1}$ , and multiply their weights by  $(1+1/1000)$  if sampled. Using Theorem 4.18, we have 4.  $\square$

*Proof of Theorem 6.8.* **Number of linear measurements.** Notice that each  $(S^{\text{hv}})_{i,j}^{(k)} w_{G^{(k)}[E_i^{(k)}]} \in \mathbb{R}^{n^{6/5} \eta_1^{-1} \epsilon_1^{-2} \text{polylog}(n, \frac{w_{\max}}{w_{\min}})}$ , so the total number of linear measurements is bounded by

$$t^3 n^{6/5} \eta_1^{-1} \epsilon_1^{-2} \text{polylog}(n, \frac{w_{\max}}{w_{\min}}) \leq n^{6/5} \epsilon^{-4} \text{polylog}(n, \frac{w_{\max}}{w_{\min}}, \epsilon^{-1}).$$

**Spectral sparsifier guarantee.** By Lemma 6.10,  $H^{(0)}$  is a 1.001-spectral sparsifier of  $G^{(0)}$ . We then show that whenever  $H^{(k-1)}$  is a 1.001-spectral sparsifier of  $G^{(k-1)}$ ,  $H^{(k)}$  is a  $(1 + \epsilon/1000)$ -spectral sparsifier of  $G^{(k)}$  with high probability. Notice that inside each iteration of the outermost for loop of  $k$ , for  $i = t$ , we have that with high probability  $E_t^{(k)} = \emptyset$ . This means that  $J_t^{(k)}$  consists of solely edges in  $F_t^{(k)}$ . Thus by Lemma 6.13,  $H_t^{(k)}$  is a  $(1 + \epsilon/(10^6 t))$ -spectral sparsifier of  $J_t^{(k)}$ . Also by Lemma 6.13,  $J_t^{(k)}$  is a  $(1 + \epsilon/(10^4 t))^t$ -spectral sparsifier of  $G^{(k)}$ . These combined imply that  $H^{(k)}$  is a  $(1 + \epsilon/1000)$ -spectral sparsifier of  $G^{(k)}$ . Now applying an induction on  $k$ , we have that  $H^{(t)}$  is a  $(1 + \epsilon/1000)$ -spectral sparsifier of  $G^{(t)}$ . Since  $G^{(t)}$  is a  $(1 + \epsilon/2)$ -spectral sparsifier of  $G$ ,  $H^{(t)}$  is a  $(1 + \epsilon)$ -spectral sparsifier of  $G$ , as desired.  $\square$

## 6.4 Sparsification of $G^{\text{sq}}$

### 6.4.1 Sparsification of $G^{\text{sq}}$ by heavy edge recovery

We first give in Figure 6 the linear sketch for sparsifying  $G^{\text{sq}}$  using the recovery of heavy edges in Lemma 6.6. We will then prove Lemma 6.6 later in Section 6.4.2. The ideas for the former linear sketch are the same as the ones we used in Section 6.3, since both are about how to sparsify a graph by repeatedly recovering heavy edges.

The performance of the linear sketch is characterized by Lemma 6.5.

**Lemma 6.5.** *For any parameter  $\epsilon_2 > 0$  and any integer  $n$ , there exists a linear sketch with sketching matrix  $S^{\text{sq}} \in \mathbb{R}^{n\epsilon_2^{-4} \text{polylog}(n, \epsilon_2^{-1}, \frac{w_{\max}}{w_{\min}}) \times \binom{n}{2}}$  and recovery algorithm `SQUARERECOVERY` such that, given an input graph  $G$  of  $n$  vertices with weight vector  $w_G$ , `SQUARERECOVERY`( $S^{\text{sq}} w_G$ ) returns a  $(1 + \epsilon_2)$ -spectral sparsifier of  $G^{\text{sq}}$  with high probability.*

The proof of Lemma 6.5 will also be largely similar to that of Theorem 6.8 in Section 6.3. However, we still include the proof here for completeness.

As in Section 6.3, we first prove some useful intermediate lemmas.

**Lemma 6.16.**  $(10^6 w_{\max} n^5 K_n)^{\text{sq}}$  is a 1.001-spectral sparsifier of  $(G^{(0)})^{\text{sq}}$ .

*Proof.* Let  $\Pi \in \mathbb{R}^{n \times n}$  be the projection matrix on the  $n - 1$ -dimensional subspace orthogonal to the all-one vector. Then  $\Pi$  has  $n - 1$  eigenvalues of 1 and one eigenvalue of 0 (with eigenvector being the all-one vector). It is known that  $L_{K_n} = n\Pi$ , so  $L_{(10^6 w_{\max} n^5 K_n)^{\text{sq}}} = 10^{12} w_{\max}^2 n^{11} \Pi$ , and has all  $n - 1$  nonzero eigenvalues equal to  $10^{12} w_{\max}^2 n^{11}$ . On the other hand, notice that by expanding

$$\begin{aligned} L_{(G^{(0)})^{\text{sq}}} &= L_{(G + 10^6 w_{\max} n^5 K_n)^{\text{sq}}} \\ &= L_{G^{\text{sq}}} + 2 \cdot 10^6 w_{\max} n^5 L_G + 10^{12} w_{\max}^2 n^{10} L_{K_n}. \end{aligned}$$

SQUARESKETCH( $G, \epsilon_2$ )

1. Let  $t = \lceil 10000 \log(10^6 \epsilon_2^{-1} \frac{w_{\max}}{w_{\min}} \cdot n^{10}) \rceil$ , and then for each  $k = 0, 1, 2, \dots, t$ :
  - (a) Let  $E_0^{(k)} \supseteq E_1^{(k)} \supseteq \dots \supseteq E_t^{(k)}$  be subsets of edge slots where  $E_0^{(k)} = \binom{V}{2}$  and  $E_i^{(k)}$  is obtained by sub-sampling each edge slot in  $E_{i-1}^{(k)}$  with probability  $(1 + 1/1000)^{-2}$ .
  - (b) For each pair  $0 \leq i, j \leq t$ , let  $(S^{\text{sqh}})_{i,j}^{(k)} \in \mathbb{R}^{n\eta_3^{-1}\epsilon_3^{-2}\text{polylog}(n) \times \binom{n}{2}}$  be a sketching matrix in Lemma 6.6 with  $\eta_3 = \epsilon_2^2/(10^{12}t^2 \log n)$ ,  $\epsilon_3 = \epsilon_2/(10^6t)$ .
2. For each  $k, i, j \in [0, t]$ , let  $G^{(k)} \leftarrow G + (1 + 10^{-4})^{-k} 10^6 w_{\max} n^5 K_n$  and compute the sketch  $(S^{\text{sqh}})_{i,j}^{(k)} w_{G^{(k)}[E_i^{(k)}]}$  where  $w_{G^{(k)}[E_i^{(k)}]} \in \mathbb{R}^{\binom{n}{2}}$  is weight vector of the subgraph  $G^{(k)}[E_i^{(k)}]$ .
3. Concatenate these sketches as  $S^{\text{sq}} w_G$ .

SQUARERECOVERY( $S^{\text{sq}} w_G$ )

1. Initially, let  $H^{(0)} \leftarrow 10^6 w_{\max} n^5 K_n$  (where  $(H^{(0)})^{\text{sq}}$  is a 1.001-spectral sparsifier of  $(G^{(0)})^{\text{sq}}$ ).
2. For  $k = 1, 2, \dots, t$ :
  - (a) Let  $H^{(k)} \leftarrow \emptyset$  (where  $(H^{(k)})^{\text{sq}}$  will be a  $(1 + \epsilon_2/1000)$ -spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ ).
  - (b) Set  $c_e \leftarrow 0$  for all  $e \in \binom{V}{2}$  (where  $c_e$  will be such that  $e$  is added to  $H^{(k)}$  when  $i = c_e$ ).
  - (c) For  $i = 0, 1, \dots, t$ :
    - i. Let  $Z \leftarrow (1 + 1/1000)^i G^{(k)}[E_i^{(k)}] + H^{(k)}[\binom{V}{2} \setminus E_i^{(k)}]$  ( $Z$  is to record what graph our linear sketches are taken over).
    - ii. Compute sketches  $s_j := (S^{\text{sqh}})_{i,j}^{(k)} w_Z, j \in [0, t]$ , where  $w_Z$  is the weight vector of  $Z$ .
    - iii. For each  $f \in H^{(k)} \cap E_i^{(k)}$ , let  $\delta_f \leftarrow i - c_f$ , and let  $\delta_f \leftarrow 0$  for all other edges.
    - iv. Let  $j \leftarrow 0$ . Then while  $\exists f : \delta_f > 0$ , do the following:
      - A. Use Lemma 6.6 to recover from  $s_j$  and  $(H^{(k-1)})^{\text{sq}}$  an edge set  $F$ , and  $j \leftarrow j+1$ .
      - B. For each  $f \in F$  such that  $\delta_f > 0$ , let  $\tilde{w}_f$  be the estimate of its weight:
        - I.  $Z \leftarrow Z - (1 - (1 + 1/1000)^{-1}) \tilde{w}_f f$ .
        - II.  $s_{j'} \leftarrow s_{j'} - (1 - (1 + 1/1000)^{-1}) (S^{\text{sqh}})_{i,j'}^{(k)} (\tilde{w}_f \chi_f)$  for all  $j \leq j' \leq t$
        - III.  $\delta_f \leftarrow \delta_f - 1$ .
      - v. Use Lemma 6.6 to recover from  $s_t$  and  $(H^{(k-1)})^{\text{sq}}$  a set  $F^*$  of edges.
      - vi. For each edge  $f \in F^*$  with estimated weight  $\tilde{w}_f$  such that  $\tilde{w}_f^2 b_f^T L_{(H^{(k-1)})^{\text{sq}}}^\dagger b_f \geq 8\eta_3$  and  $f$  is not already in  $H^{(k)}$ , add  $f$  to  $H^{(k)}$  with weight  $\tilde{w}_f$ , and let  $c_f \leftarrow i$ .
  3. Return  $(H^{(t)})^{\text{sq}}$ .

Figure 6: Linear sketch for sparsifying  $G^{\text{sq}}$ .

By standard bounds on the largest eigenvalue of the Laplacian matrix of a weighted graph, the largest eigenvalue of an  $n$ -vertex graph with maximum weight  $w_{\max}$  is at most  $n w_{\max}$ . Therefore, the largest eigenvalue of  $L_{(G^{(0)})^{\text{sq}} - (10^6 w_{\max} n^5 K_n)^{\text{sq}}}$  is at most  $3 \cdot 10^6 w_{\max}^2 n^6$ . This combined with  $L_{(G^{(0)})^{\text{sq}}} \geq L_{(10^6 w_{\max} n^5 K_n)^{\text{sq}}}$  shows that

$$-\frac{1}{10^5 n^5} L_{(G^{(0)})^{\text{sq}}} \preceq L_{(G^{(0)})^{\text{sq}}} - L_{(10^6 w_{\max} n^5 K_n)^{\text{sq}}} \preceq \frac{1}{10^5 n^5} L_{(G^{(0)})^{\text{sq}}}.$$

This implies that

$$\frac{1}{1.001} L_{(G^{(0)})^{\text{sq}}} \preceq L_{(10^6 w_{\max} n^5 K_n)^{\text{sq}}} \preceq 1.001 L_{(G^{(0)})^{\text{sq}}}$$

as desired.  $\square$

**Lemma 6.17.** *For all  $k \geq 1$ ,  $(G^{(k-1)})^{\text{sq}}$  is a 1.001-spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ .*

*Proof.* By definition

$$\begin{aligned} L_{(G^{(k-1)})^{\text{sq}}} &= L_{(G + (1+10^{-4})^{-(k-1)} 10^6 w_{\max} n^5 K_n)^{\text{sq}}} \\ &= L_{G^{\text{sq}}} + 2(1+10^{-4})^{-(k-1)} 10^6 w_{\max} n^5 L_G + (1+10^{-4})^{-2(k-1)} 10^{12} w_{\max}^2 n^{10} L_{K_n} \end{aligned}$$

and

$$\begin{aligned} L_{(G^{(k)})^{\text{sq}}} &= L_{(G + (1+10^{-4})^{-k} 10^6 w_{\max} n^5 K_n)^{\text{sq}}} \\ &= L_{G^{\text{sq}}} + 2(1+10^{-4})^{-k} 10^6 w_{\max} n^5 L_G + (1+10^{-4})^{-2k} 10^{12} w_{\max}^2 n^{10} L_{K_n}. \end{aligned}$$

The second terms of the above two expressions are  $(1+10^{-4})$ -sparsifiers of each other, and the third terms of the above two expressions are  $(1+10^{-4})^2$ -sparsifiers of each other. Therefore,  $L_{(G^{(k-1)})^{\text{sq}}}$  is a 1.001-spectral sparsifier of  $L_{(G^{(k)})^{\text{sq}}}$ .  $\square$

**Lemma 6.18.**  *$(G^{(t)})^{\text{sq}}$  is a  $(1+\epsilon_2/2)$ -spectral sparsifier of  $G^{\text{sq}}$ .*

*Proof.* By definition

$$\begin{aligned} L_{(G^{(t)})^{\text{sq}}} &= L_{(G + (1+10^{-4})^{-t} 10^6 w_{\max} n^5 K_n)^{\text{sq}}} \\ &= L_{G^{\text{sq}}} + 2(1+10^{-4})^{-t} 10^6 w_{\max} n^5 L_G + (1+10^{-4})^{-2t} 10^{12} w_{\max}^2 n^{10} L_{K_n} \\ &\leq L_{G^{\text{sq}}} + .1\epsilon_2 \frac{w_{\min}^2}{w_{\max}} n^{-5} L_G + .001\epsilon_2^2 \frac{w_{\min}^4}{w_{\max}^2} n^{-10} L_{K_n}, \end{aligned}$$

where the last line follows from our choice of  $t$ . Thus, the largest eigenvalue of the sum of the last two terms is bounded by  $.2\epsilon_2 w_{\min}^2 n^{-4}$ . By standard lower bounds on the second smallest eigenvalue, the second smallest eigenvalue of  $L_{G^{\text{sq}}}$  is at least  $w_{\min}^2/n^2$ . Therefore we have

$$-.2\epsilon_2 L_{G^{\text{sq}}} \preceq L_{(G^{(t)})^{\text{sq}}} - L_{G^{\text{sq}}} \preceq .2\epsilon_2 L_{G^{\text{sq}}},$$

which implies that  $(G^{(t)})^{\text{sq}}$  is a  $(1+\epsilon_2/2)$ -spectral sparsifier of  $G^{\text{sq}}$ .  $\square$

Fix an iteration of the outer for loop of  $k$ . Then for an iteration of the inner for loop of  $i$ , let  $H_i^{(k)}$  be the  $H^{(k)}$  at the *beginning* of the iteration, and let  $F_i^{(k)}$  be the edges in  $H_i^{(k)}$ . Define graph

$$J_i^{(k)} := (1 + 1/1000)^i G^{(k)}[E_i^{(k)} \setminus F_i^{(k)}] + \sum_{\ell=0}^{i-1} (1 + 1/1000)^\ell G^{(k)}[F_{\ell+1}^{(k)} \setminus F_\ell^{(k)}].$$

Notice that  $F_0^{(k)} = \emptyset$  and  $J_0^{(k)} = G^{(k)}$ . Also by the way we are assigning values to  $c_f$  in the algorithm, we have, at the beginning of the for loop (of  $i$ ) iteration,  $f \in F_{c_f+1}^{(k)} \setminus F_{c_f}^{(k)}$  for all  $f \in F_i^{(k)}$ . Thus we also have

$$J_i^{(k)} := (1 + 1/1000)^i G^{(k)}[E_i^{(k)} \setminus F_i^{(k)}] + \sum_{f \in F_i^{(k)}} (1 + 1/1000)^{c_f} (w_{G^{(k)}})_f f. \quad (16)$$

**Lemma 6.19.** *Suppose  $(H^{(k-1)})^{\text{sq}}$  is a 1.001-spectral sparsifier of  $(G^{(k-1)})^{\text{sq}}$ . Then with high probability, for all  $0 \leq i < t$ ,*

1. *After the while loop inside the  $i^{\text{th}}$  iteration terminates, for all  $f$ ,*

$$\frac{1}{1 + \epsilon_2/10000} (w_{J_i^{(k)}})_f \leq (w_Z)_f \leq (1 + \epsilon_2/10000) (w_{J_i^{(k)}})_f.$$

2. *For all  $f \in F_{i+1}^{(k)} \setminus F_i^{(k)}$ ,*

$$\frac{1}{1 + \epsilon_2/(10^6 t)} (w_{J_{i+1}^{(k)}})_f \leq (w_{H_{i+1}^{(k)}})_f \leq (1 + \epsilon_2/(10^6 t)) (w_{J_{i+1}^{(k)}})_f.$$

3. *All edges in  $F_{i+1}^{(k)}$  have leverage scores in  $(J_i^{(k)})^{\text{sq}}$  at least  $4\eta_3$ .*

4.  *$(J_{i+1}^{(k)})^{\text{sq}}$  is a  $(1 + \epsilon_2/(10^4 t))$ -spectral sparsifier of  $(J_i^{(k)})^{\text{sq}}$ .*

*Proof.* We prove all statements of this lemma by induction on  $i$ . For  $i = 0$ , since  $\delta_f = 0$  for all  $f$ , the while loop will not execute. Thus throughout this iteration we have  $Z = J_0^{(k)} = G^{(k)}$ . This immediately gives 1. Since  $(H^{(k-1)})^{\text{sq}}$  is a 1.001-spectral sparsifier of  $(G^{(k-1)})^{\text{sq}}$ , and  $(G^{(k-1)})^{\text{sq}}$  is in turn a 1.001-spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ , we have that  $(H^{(k-1)})^{\text{sq}}$  is a 1.003-spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ . Therefore by letting  $\tilde{G} = (H^{(k-1)})^{\text{sq}}$ , we have

$$\frac{((w_{G^{(k)}})_e b_e^T L_{\tilde{G}}^\dagger b_e)^2}{b_e^T L_{\tilde{G}}^\dagger L_{(G^{(k)})^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \approx_{1.003^4} \frac{((w_{G^{(k)}})_e b_e^T L_{(G^{(k)})^{\text{sq}}}^\dagger b_e)^2}{b_e^T L_{(G^{(k)})^{\text{sq}}}^\dagger b_e} = (w_{G^{(k)}})_e^2 b_e^T L_{(G^{(k)})^{\text{sq}}}^\dagger b_e, \quad (17)$$

where in the first step we have used

$$(w_{G^{(k)}})_e b_e^T L_{\tilde{G}}^\dagger b_e \approx_{1.003} (w_{G^{(k)}})_e b_e^T L_{(G^{(k)})^{\text{sq}}}^\dagger b_e$$

and

$$b_e^T L_{\tilde{G}}^\dagger L_{(G^{(k)})^{\text{sq}}} L_{\tilde{G}}^\dagger b_e \approx_{1.003} b_e^T L_{\tilde{G}}^\dagger L_{\tilde{G}} L_{\tilde{G}}^\dagger b_e = b_e^T L_{\tilde{G}}^\dagger b_e \approx_{1.003} b_e^T L_{(G^{(k)})^{\text{sq}}}^\dagger b_e.$$

Therefore by Lemma 6.6, the  $F^*$  we recover in this iteration contains all edges whose leverage score in  $(G^{(k)})^{\text{sq}}$  is at least  $1.1\eta_3$ , and all edges in  $F^*$  have weight estimates satisfying 2. Also, at the final step of this for loop iteration, since  $(H^{(k-1)})^{\text{sq}}$  is a 1.003-spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ , all edges with leverage score  $\geq 10\eta_3$  in  $(G^{(k)})^{\text{sq}}$  will be added to  $H^{(k)}$ , and all edges added to  $H^{(k)}$  have leverage score at least  $\geq 4\eta_3$  in  $(G^{(k)})^{\text{sq}}$ , so we have 3. This means that  $(J_1^{(k)})^{\text{sq}}$  is obtained by sampling a set of edges in  $(J_0^{(k)})^{\text{sq}}$  whose leverage scores in  $(J_0^{(k)})^{\text{sq}}$  are at most  $10\eta_3$  with probability  $(1 + 1/1000)^{-2}$ , and multiply their weights by  $(1 + 1/1000)^2$  if sampled. Using Theorem 4.18, we have 4.

We now do an inductive step. Suppose all four statements hold for iterations  $0, 1, \dots, i-1$  where  $1 < i < t$ . We show that they also hold for iteration  $i$ . We first need to analyze the while loop inside iteration  $i$ . Let us number a while loop iteration by the value of  $j$  at the *end* of the iteration.

**Claim 6.20.** At the end of while loop iteration  $j$  where  $j \leq t$ , we have for all  $f \in E_i^{(k)} \cap F_i^{(k)}$

$$\frac{1}{(1+2\epsilon_3)^j} \cdot (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f \leq (w_Z)_f \leq (1+2\epsilon_3)^j (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f.$$

*Proof.* We prove this claim by an induction on  $j$ . First we show that the statement is true for  $j = 0$  at the beginning of while loop iteration 1. Here all  $f \in E_i^{(k)} \cap F_i^{(k)}$  satisfy that  $(w_Z)_f = (1+1/1000)^i (w_{G^{(k)}})_f$ . Since we set  $\delta_f \leftarrow i - c_f$  before the while loop, and by (16)  $(w_{J_i^{(k)}})_f = (1+1/1000)^{c_f} (w_{G^{(k)}})_f$ , we have  $(w_Z)_f = (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f$ , as desired.

Now suppose the statement is true at the end of iteration  $j-1$  where  $1 < j \leq t$ . We then show that the statement is also true at the end of iteration  $j$ . Let  $Z_0$  be the  $Z$  before our updates to  $Z$  in iteration  $j$  and let  $Z_1$  be the  $Z$  after our updates. By Lemma 6.6, all edges recovered  $f \in F$  have their estimated edge weights  $\tilde{w}_f \in [\frac{1}{1+\epsilon_3} (w_{Z_0})_f, (1+\epsilon_3)(w_{Z_0})_f]$ . Therefore after our updates, we have for any  $f \in F$  such that  $\delta_f > 0$  that  $(w_{Z_1})_f \in [\frac{1}{1+2\epsilon_3} (1+1/1000)^{-1} (w_{Z_0})_f, (1+2\epsilon_3)(1+1/1000)^{-1} (w_{Z_0})_f]$ , and  $(w_{Z_1})_f = (w_{Z_0})_f$  for other edges  $f$ . Since we let  $\delta_f \leftarrow \delta_f - 1$  for such edges (those with  $\delta_f > 0$ ), and do not change the  $\delta_f$ 's of other edges, we have our desired statement for  $j$ .  $\square$

**Claim 6.21.** The while loop terminates after at most  $t$  iterations.

*Proof.* It suffices to show that  $\max_f \delta_f$  decreases by 1 in each while loop iteration. Since  $\delta_f \leq t$  for any  $f$ , this will imply that there can be at most  $t$  iterations. Then it boils down to showing that for all  $f^*$  with  $\delta_{f^*} = \max_f \delta_f$ ,  $f^*$  belongs to the recovered edge set  $F$ . Since  $f^* \in F_i^{(k)}$ , by 3 of our induction hypothesis, the leverage score of  $f^*$  in  $(J_{i-1}^{(k)})^{\text{sq}}$  is at least  $4\eta_3$ . Notice that by 4 of our induction hypothesis,  $(J_{i-1}^{(k)})^{\text{sq}}$  is a  $(1+1/1000)$ -spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ . Then using the fact that  $(H^{(k-1)})^{\text{sq}}$  is a 1.003-spectral sparsifier of  $(G^{(k)})^{\text{sq}}$  (which we proved at the beginning of the proof of this lemma), we have that  $(H^{(k-1)})^{\text{sq}}$  is a 1.005-spectral sparsifier of  $(J_{i-1}^{(k)})^{\text{sq}}$ .

By Claim 6.20, we have at the beginning of each while loop that, for all  $f$ ,

$$\begin{aligned} (w_Z)_f &\in \left[ \frac{1}{(1+2\epsilon_3)^t} (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f, (1+2\epsilon_3)^t (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f \right] \\ &\subseteq \left[ \frac{1}{1.01} (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f, 1.01 (1+1/1000)^{\delta_f} (w_{J_i^{(k)}})_f \right]. \end{aligned} \quad (18)$$

Since  $\delta_{f^*} \geq \delta_f$  for all  $f$ , the above implies

$$L_{Z^{\text{sq}}} \preceq 1.03 (1+1/1000)^{2\delta_{f^*}} L_{(J_{i-1}^{(k)})^{\text{sq}}} \preceq 1.04 (1+1/1000)^{2\delta_{f^*}} L_{(H^{(k-1)})^{\text{sq}}}, \quad (19)$$

where the second inequality follows from that  $(H^{(k-1)})^{\text{sq}}$  is a 1.005-spectral sparsifier of  $(J_{i-1}^{(k)})^{\text{sq}}$ . Let  $\tilde{G} = (H^{(k-1)})^{\text{sq}}$ . Now in order to show that  $f^*$  will be recovered as an edge in  $F^*$ , by Lemma 6.6, it suffices to show

$$\frac{((w_Z)_{f^*} b_{f^*}^T L_{\tilde{G}}^\dagger b_{f^*})^2}{b_{f^*}^T L_{\tilde{G}}^\dagger L_{Z^{\text{sq}}} L_{\tilde{G}}^\dagger b_{f^*}} \geq \eta_3. \quad (20)$$

By (19), the denominator satisfies

$$b_{f^*}^T L_{\tilde{G}}^\dagger L_{Z^{\text{sq}}} L_{\tilde{G}}^\dagger b_{f^*} \leq 1.04 (1+1/1000)^{2\delta_{f^*}} b_{f^*}^T L_{\tilde{G}}^\dagger b_{f^*}.$$

Therefore, the LHS of (20) is at least

$$(w_Z)_{f^*}^2 (1+1/1000)^{-2\delta_{f^*}} b_{f^*}^T L_{\tilde{G}}^\dagger b_{f^*} \geq 1.005^{-1} (w_Z)_{f^*}^2 (1+1/1000)^{-2\delta_{f^*}} b_{f^*}^T L_{(J_{i-1}^{(k)})^{\text{sq}}}^\dagger b_{f^*},$$

where the inequality follows from that  $(H^{(k-1)})^{\text{sq}}$  is a 1.005-spectral sparsifier of  $(J_{i-1}^{(k)})^{\text{sq}}$ . Finally, using (18) and that the leverage score of  $f^*$  is at least  $2\eta_3$  in  $(J_{i-1}^{(k)})^{\text{sq}}$ , we have that the above is at least  $\eta_3$ , proving (20).  $\square$

By Claim 6.20, after the while loop terminates, we have that for all  $f$ ,

$$\begin{aligned} (w_Z)_f &\in \left[ \frac{1}{(1+2\epsilon_3)^t} (w_{J_i^{(k)}})_f, (1+2\epsilon_3)^t (w_{J_i^{(k)}})_f \right] \\ &\subseteq \left[ \frac{1}{(1+\epsilon_2/10000)} (w_{J_i^{(k)}})_f, (1+\epsilon_2/10000) (w_{J_i^{(k)}})_f \right], \end{aligned}$$

and thus we have 1. This also implies that  $Z^{\text{sq}}$  is a  $(1+\epsilon_2/10000)^2$ -spectral sparsifier of  $(J_i^{(k)})^{\text{sq}}$ , and as result, for each edge  $f$ , its leverage scores in  $Z^{\text{sq}}$  and  $(J_i^{(k)})^{\text{sq}}$  are within a  $(1+\epsilon_2/10000)^3 < 1.01$  factor of each other.

For all edges in  $E_i^{(k)} \setminus F_i^{(k)}$ , their weights in  $Z$  equal exactly their weights in  $J_i^{(k)}$ , therefore by Lemma 6.6, all edges recovered in  $F^*$  not in  $F_i^{(k)}$  have weight estimates satisfying 2.

Notice that by 4 of our induction hypothesis,  $(J_i^{(k)})^{\text{sq}}$  is a  $(1+1/1000)$ -spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ . Then using the fact that  $(H^{(k-1)})^{\text{sq}}$  is a 1.003-spectral sparsifier of  $(G^{(k)})^{\text{sq}}$  (which we

proved at the beginning of this proof), we have that  $(H^{(k-1)})^{\text{sq}}$  is a 1.01-spectral sparsifier of  $Z^{\text{sq}}$ . By letting  $\tilde{G} = (H^{(k-1)})^{\text{sq}}$  we have, similar to (17),

$$\frac{((w_Z)_e b_e^T L_{\tilde{G}}^\dagger b_e)^2}{b_e^T L_{\tilde{G}}^\dagger L_{Z^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \approx_{1.01^4} \frac{((w_Z)_e b_e^T L_{Z^{\text{sq}}}^\dagger b_e)^2}{b_e^T L_{Z^{\text{sq}}}^\dagger b_e} = (w_Z)_e^2 b_e^T L_{Z^{\text{sq}}}^\dagger b_e. \quad (21)$$

Therefore by Lemma 6.6, the  $F^*$  we recover in this iteration contains all edges whose leverage scores in  $(Z)^{\text{sq}}$  are at least  $1.5\eta_3$ . Also, at the last step of the for iteration, since  $(H^{(k-1)})^{\text{sq}}$  is a 1.01-spectral sparsifier of  $Z^{\text{sq}}$ , all edges added to  $H^{(k)}$  have leverage scores at least  $\geq 5\eta_3$  in  $(Z)^{\text{sq}}$ , and all edges with leverage scores  $\geq 9\eta_3$  in  $(Z)^{\text{sq}}$  will be added to  $H^{(k)}$ . Thus we also know that all edges added to  $H^{(k)}$  have leverage scores at least  $4\eta_3$  in  $(J_i^{(k)})^{\text{sq}}$  (which gives 3), and all edges with leverage scores  $\geq 10\eta_3$  in  $(J_i^{(k)})^{\text{sq}}$  will be added to  $H^{(k)}$ ,

The above reasoning also implies that  $(J_{i+1}^{(k)})^{\text{sq}}$  is obtained by sampling a set of edges in  $(J_i^{(k)})^{\text{sq}}$  whose leverage score is at most  $10\eta_3$  with probability  $(1 + 1/1000)^{-2}$ , and multiply their weights by  $(1 + 1/1000)^2$  if sampled. Using Theorem 4.18, we have 4.  $\square$

*Proof of Lemma 6.5. Number of linear measurements.* Notice that each  $(S^{\text{sqh}})_{i,j}^{(k)} w_{G^{(k)}[E_i^{(k)}]} \in \mathbb{R}^{n\eta_3^{-1}\epsilon_3^{-2}\text{polylog}(n)}$ , so the total number of linear measurements is bounded by

$$t^3 n \eta_3^{-1} \epsilon_3^{-2} \text{polylog}(n) \leq n \epsilon_2^{-4} \text{polylog}(n, \frac{w_{\max}}{w_{\min}}, \epsilon_2^{-1}).$$

**Spectral sparsifier guarantee.** By Lemma 6.16,  $(H^{(0)})^{\text{sq}}$  is a 1.001-spectral sparsifier of  $(G^{(0)})^{\text{sq}}$ . We then show that whenever  $(H^{(k-1)})^{\text{sq}}$  is a 1.001-spectral sparsifier of  $(G^{(k-1)})^{\text{sq}}$ ,  $(H^{(k)})^{\text{sq}}$  is a  $(1 + \epsilon_2/1000)$ -spectral sparsifier of  $(G^{(k)})^{\text{sq}}$  with high probability. Notice that inside each iteration of the outermost for loop of  $k$ , for  $i = t$ , we have that with high probability  $E_t^{(k)} = \emptyset$ . This means that  $J_t^{(k)}$  consists of solely edges in  $F_t^{(k)}$ . Thus by Lemma 6.19,  $(H_t^{(k)})^{\text{sq}}$  is a  $(1 + \epsilon_2/(10^6 t))^2$ -spectral sparsifier of  $(J_t^{(k)})^{\text{sq}}$ . Also by Lemma 6.19,  $(J_t^{(k)})^{\text{sq}}$  is a  $(1 + \epsilon_2/(10^4 t))^t$ -spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ . These combined imply that  $(H^{(k)})^{\text{sq}}$  is a  $(1 + \epsilon_2/1000)$ -spectral sparsifier of  $(G^{(k)})^{\text{sq}}$ . Now applying an induction on  $k$ , we have that  $(H^{(t)})^{\text{sq}}$  is a  $(1 + \epsilon_2/1000)$ -spectral sparsifier of  $(G^{(t)})^{\text{sq}}$ . Since  $(G^{(t)})^{\text{sq}}$  is a  $(1 + \epsilon_2/2)$ -spectral sparsifier of  $G^{\text{sq}}$ ,  $(H^{(t)})^{\text{sq}}$  is a  $(1 + \epsilon_2)$ -spectral sparsifier of  $G^{\text{sq}}$ , as desired.  $\square$

#### 6.4.2 Recovery of heavy edges in $G^{\text{sq}}$

We now give in Figure 7 a linear sketch for recovering heavy edges in  $G^{\text{sq}}$  and its analysis. This linear sketch is essentially a direct application of  $\ell_2$ -heavy hitters.

The performance of the linear sketch is characterized in Lemma 6.6:

**Lemma 6.6.** *For any parameters  $\eta_3, \epsilon_3 \in (0, 1)$  and any integer  $n$ , there exists a linear sketch with sketching matrix  $S^{\text{sqh}} \in \mathbb{R}^{n\eta_3^{-1}\epsilon_3^{-2}\text{polylog}(n) \times \binom{n}{2}}$  and recovery algorithm **SQUAREHEAVYRECOVERY** such that, for an input graph  $G$  of  $n$  vertices with weight vector  $w_G$  and another graph  $\tilde{G}$ , **SQUAREHEAVYRECOVERY**( $S^{\text{sqh}} w_G, \tilde{G}$ ) recovers a set of edges  $F$  in  $G$  along with estimates of their weights  $\tilde{w}_f$ 's such that with high probability*

SQUAREHEAVYSKETCH( $G, \eta_3, \epsilon_3$ )

1. Let  $A \in \mathbb{R}^{\eta^{-2}\text{polylog}(n) \times \binom{n}{2}}$  be an  $\ell_2$ -heavy hitter sketching matrix with  $\eta = \sqrt{\eta_3}\epsilon_3/10000$  (Proposition 4.8).
2. Let  $J \in \mathbb{R}^{O(\delta^{-2}\log n) \times \binom{n}{2}}$  be an  $\ell_2$ -estimation sketching matrix with  $\delta = .01$  (Proposition 4.12)
3. Concatenate the sketches  $AW_G B_G$  and  $JW_G B_G$  as  $S^{\text{sqh}} w_G$ .

SQUAREHEAVYRECOVERY( $S^{\text{sqh}} w_G, \tilde{G}$ )

1. For each vertex pair  $s, t \in V$ :
  - (a) Let  $x = L_{\tilde{G}}^\dagger b_{s,t} \in \mathbb{R}^n$  be the set of vertex potentials induced by a unit electrical flow from  $s$  to  $t$  in  $\tilde{G}$ .
  - (b) Let  $z \in \mathbb{R}^{\binom{n}{2}}$  be an approximation to  $W_G B_G x$  that is recovered from  $AW_G B_G x$  using the recovery algorithm in Proposition 4.8.
  - (c) Let  $\beta$  be an estimate of  $\|W_G B_G x\|_2$  recovered from  $JW_G B_G x$  using Proposition 4.12.
  - (d) If  $z_{(s,t)} \geq .9\sqrt{\eta_3}\beta$ , mark edge  $(s, t)$  as heavy, and estimate its weight as  $\tilde{w}_{(s,t)} = \frac{z_{(s,t)}}{x^T b_{s,t}}$ .
2. Return all edges marked heavy along with the estimates of their weights.

Figure 7: Linear sketch for recovering heavy edges in  $G^{\text{sq}}$ .

1.  $F$  contains all edges  $e$  satisfying

$$\frac{((w_G)_e b_e^T L_{\tilde{G}}^\dagger b_e)^2}{b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \geq \eta_3.$$

2. All edges  $f \in F$  satisfy  $\frac{1}{1+\epsilon_3}(w_G)_f \leq \tilde{w}_f \leq (1 + \epsilon_3)(w_G)_f$ .

*Proof of Lemma 6.6. Number of linear measurements.* Notice that the sketches we compute satisfy  $AW_G B_G \in \mathbb{R}^{\eta_3^{-1}\epsilon_3^{-2}\text{polylog}(n) \times n}$  and  $JW_G B_G \in \mathbb{R}^{O(\log n) \times n}$ . Therefore by concatenating them, the total number of linear measurements is bounded by  $n\eta_3^{-1}\epsilon_3^{-2}\text{polylog}(n)$ .

**First guarantee.** By the guarantee of the  $\ell_2$ -heavy hitter sketch, we have

$$\begin{aligned} |(W_G B_G x)_{(s,t)} - z_{(s,t)}| &\leq \eta \|W_G B_G x\|_2 \\ &\leq (\sqrt{\eta_3}\epsilon_3/10000) \|W_G B_G x\|_2 \quad (\text{by the value of } \eta) \\ &= (\sqrt{\eta_3}\epsilon_3/10000) \sqrt{b_{s,t}^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_{s,t}} \end{aligned} \tag{22}$$

By the guarantee of the  $\ell_2$ -estimation sketch

$$\beta \in \left[ \frac{1}{1.01} \sqrt{b_{s,t}^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_{s,t}}, 1.01 \sqrt{b_{s,t}^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_{s,t}} \right]. \tag{23}$$

Now fix any  $e$  with

$$\frac{((w_e)_G b_e^T L_{\tilde{G}}^\dagger b_e)^2}{b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \geq \eta_3. \quad (24)$$

We then write

$$\begin{aligned} |z_e - (W_G B_G x)_e| &\leq \frac{\epsilon_3}{10000} \sqrt{\eta_3 b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \quad \text{(by (22))} \\ &\leq \frac{\epsilon_3}{10000} \sqrt{\eta_3} \sqrt{1/\eta_3} w_e b_e^T L_{\tilde{G}}^\dagger b_e \quad \text{(by (24))} \\ &\leq .001 \epsilon_3 w_e b_e^T L_{\tilde{G}}^\dagger b_e \\ &= .001 \epsilon_3 (W_G B_G x)_e. \end{aligned} \quad (25)$$

This implies that

$$\begin{aligned} z_e &\geq .999 w_e b_e^T L_{\tilde{G}}^\dagger b_e \\ &\geq .999 \sqrt{\eta_3 b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \quad \text{(by (24))} \\ &\geq .999 \sqrt{\eta_3} \frac{\beta}{1.01} \quad \text{(by (23))} \\ &\geq .9 \sqrt{\eta_3} \beta. \end{aligned}$$

Therefore  $e$  will be marked as heavy.

**Second guarantee.** For  $e$  such that  $z_e \geq .9 \sqrt{\eta_3} \beta$ , we have

$$\begin{aligned} (W_G B_G x)_e &\geq z_e - \frac{\epsilon_3}{10000} \sqrt{\eta_3 b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \quad \text{(by (22))} \\ &\geq .9 \sqrt{\eta_3} \frac{1}{1.01} \sqrt{b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} - \frac{\epsilon_3}{10000} \sqrt{b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \quad \text{(by (23))} \\ &\geq .87 \sqrt{\eta_3 b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \quad \text{(by } \epsilon_3 < 1\text{).} \end{aligned}$$

Since  $(W_G B_G x)_e = w_e b_e^T L_{\tilde{G}}^\dagger b_e$ , we have

$$\frac{((w_e)_G b_e^T L_{\tilde{G}}^\dagger b_e)^2}{b_e^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_e} \geq .7 \eta_3. \quad (26)$$

We then write

$$\begin{aligned} |z_e - (W_G B_G x)_e| &\leq \frac{\epsilon_3}{10000} \sqrt{\eta_3 b_{s,t}^T L_{\tilde{G}}^\dagger L_{G^{\text{sq}}} L_{\tilde{G}}^\dagger b_{s,t}} \quad \text{(by (22))} \\ &\leq \frac{\epsilon_3}{10000} \sqrt{\eta_3} \sqrt{1/(.7 \eta_3)} w_e b_e^T L_{\tilde{G}}^\dagger b_e \quad \text{(by (26))} \\ &\leq .01 \epsilon_3 w_e b_e^T L_{\tilde{G}}^\dagger b_e \\ &= .01 \epsilon_3 (W_G B_G x)_e. \end{aligned} \quad (27)$$

By dividing both sides by  $x^T b_e$ , we have

$$\left| \frac{z_e}{x^T b_e} - \frac{w_e b_e^T L_{\tilde{G}}^\dagger b_e}{x^T b_e} \right| \leq .01\epsilon_3 \frac{w_e b_e^T L_{\tilde{G}}^\dagger b_e}{x^T b_e}.$$

Since  $x^T b_e = b_e^T L_{\tilde{G}}^\dagger b_e$ , the above is equivalent to  $\left| \frac{z_e}{x^T b_e} - w_e \right| \leq .01\epsilon_3 w_e$ , and thus we conclude  $\frac{1}{1+\epsilon_3} w_e \leq \frac{w_e b_e^T L_{\tilde{G}}^\dagger b_e}{x^T b_e} \leq (1 + \epsilon_3) w_e$ , as desired.  $\square$

## 7 Preliminaries on matrix-weighted graphs

We consider an undirected,  $k \times k$  matrix-weighted graph  $G = (V, E)$  with  $|V| = n$  and  $|E| = m$ , where each edge  $(u, v)$ 's weight is given by  $\phi_{uv}\phi_{uv}^T$  for some  $\phi_{uv} \in \mathbb{R}^k$ . We will assume all  $\phi_{uv} \neq 0$ , since one could remove all zero weight edges from  $E$ .

**Remark 7.1** (Block matrices and vectors). Below, all  $nk \times nk$  matrices (or  $nk$ -dimensional vectors) are written in block form with block size  $k \times k$  (or  $k \times 1$ ). All subscripts also refer to row/column block numbers as opposed to row/column numbers. For example, for an  $nk \times nk$  matrix  $M$ , we write  $M_{ij}$  to denote the  $k \times k$  submatrix that is at the intersection of the  $i^{\text{th}}$  row block and the  $j^{\text{th}}$  column block.

**Definition 7.2** (Degree matrices). For a vertex  $u$ , its degree is given by

$$D_u = \sum_{u \sim v} \phi_{uv}\phi_{uv}^T \in \mathbb{R}^{k \times k}.$$

We then define the  $nk \times nk$  degree matrix  $D$  as a  $k \times k$ -block diagonal matrix:

$$D = \begin{pmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_n \end{pmatrix}.$$

**Definition 7.3.** For each edge  $(u, v) \in E$ , define two  $nk$ -dimensional vectors  $e_{u \leftarrow v}$  and  $e_{v \leftarrow u}$

$$e_{u \leftarrow v} = \begin{pmatrix} 0 \\ \vdots \\ \phi_{uv} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{$u^{\text{th}}$ block} \quad \text{and} \quad e_{v \leftarrow u} = \begin{pmatrix} 0 \\ \vdots \\ \phi_{uv} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{$v^{\text{th}}$ block}$$

**Proposition 7.1.**  $D = \sum_{u \sim v} (e_{u \leftarrow v} e_{u \leftarrow v}^T + e_{v \leftarrow u} e_{v \leftarrow u}^T)$ .

When defining almost regular graphs, we will be interested in the quantities  $\phi_{uv}^T D_u^\dagger \phi_{uv}$ . We will call  $\phi_{uv}^T D_u^\dagger \phi_{uv}$  the leverage score of edge  $(u, v)$  with respect to vertex  $u$ .

**Definition 7.4** (Leverage score of an edge w.r.t. a vertex). We call  $\phi_{uv}^T D_u^\dagger \phi_{uv}$  the leverage score of edge  $(u, v)$  w.r.t. vertex  $u$ .

**Proposition 7.2.** *For any edge  $(u, v)$ ,  $\phi_{uv}^T D_u^\dagger \phi_{uv} = e_{u \leftarrow v}^T D^\dagger e_{u \leftarrow v}$ .*

**Definition 7.5** (Adjacency matrices). The  $nk \times nk$  adjacency matrix  $A$  is given by  $A_{uv} = \phi_{uv}\phi_{uv}^T$  for  $u \sim v$  and  $A_{uv} = 0$  otherwise.

In the following definitions, we assume there is a (arbitrarily) fixed orientation of all edges.

**Definition 7.6** (Incidence vectors). For an edge  $(u, v)$  oriented as  $u \rightarrow v$ , its  $nk$ -dimensional incidence vector  $b_{uv}$  is given by

$$b_{uv} = \begin{pmatrix} 0 \\ \vdots \\ \phi_{uv} \\ 0 \\ \vdots \\ -\phi_{uv} \\ 0 \\ \vdots \end{pmatrix} \begin{array}{l} u^{\text{th}} \text{ block} \\ \cdot \\ v^{\text{th}} \text{ block} \end{array}$$

**Definition 7.7** (Laplacian matrices). The Laplacian matrix is given by  $L = D - A$ .

**Proposition 7.3.** *We have  $L = \sum_{u \sim v} b_{uv} b_{uv}^T$ .*

**Proposition 7.4.** *For any vector  $x \in \mathbb{R}^{nk}$ , we have*

$$x^T L x = \sum_{u \sim v} \langle x_u - x_v, \phi_{uv} \rangle^2.$$

**Definition 7.8** (Projection matrices). For an  $s$ -dimensional subspace  $\mathcal{S} \subseteq \mathbb{R}^d$  with orthonormal basis  $b_1, \dots, b_s \in \mathbb{R}^d$ , we write

$$\Pi_{\mathcal{S}} = \sum_{i=1}^s b_i b_i^T \in \mathbb{R}^{d \times d}$$

to denote the projection matrix onto  $\mathcal{S}$ . We will also write, for a symmetric matrix  $M \in \mathbb{R}^{d \times d}$ ,  $\Pi_M \in \mathbb{R}^{d \times d}$  to denote the projection matrix onto the range of  $M$ .

**Definition 7.9** (Normalized Laplacians). The normalized Laplacian matrix is given by

$$N \stackrel{\text{def}}{=} D^{\dagger/2} L D^{\dagger/2} \equiv \Pi_D - D^{\dagger/2} A D^{\dagger/2}.$$

**Proposition 7.5.**  $N = \sum_{u \sim v} (D^{\dagger/2} b_{uv})(D^{\dagger/2} b_{uv})^T$ .

**Proposition 7.6.** *For any vector  $x \in \mathbb{R}^{nk}$ , we have*

$$x^T N x = \sum_{u \sim v} \left( \left\langle x_u, D_u^{\dagger/2} \phi_{uv} \right\rangle - \left\langle x_v, D_v^{\dagger/2} \phi_{uv} \right\rangle \right)^2.$$

**Proposition 7.7.** *The eigenvalues of  $N$  is between  $[0, 2]$ .*

Sometimes, instead of using exact degrees, we will normalize the Laplacian matrix by approximate degrees. Thus we also define  $\tilde{D}$ -normalized Laplacian matrix where  $\tilde{D} \in \mathbb{R}^{nk \times nk}$  is a semidefinite,  $k \times k$ -block diagonal matrix that has the same range as  $D$ .

**Definition 7.10** ( $\tilde{D}$ -normalized Laplacians). For a semidefinite,  $k \times k$ -block diagonal matrix  $\tilde{D} \in \mathbb{R}^{nk \times nk}$  whose range is the same as  $D$ , the  $\tilde{D}$ -normalized Laplacian matrix is given by

$$\tilde{N} \stackrel{\text{def}}{=} \tilde{D}^{\dagger/2} L \tilde{D}^{\dagger/2}.$$

**Proposition 7.8.**  $\tilde{N} = \sum_{u \sim v} (\tilde{D}^{\dagger/2} b_{uv})(\tilde{D}^{\dagger/2} b_{uv})^T$ .

**Proposition 7.9.** *For  $\tilde{D}$  satisfying*

$$\frac{1}{\kappa} D \preceq \tilde{D} \preceq \kappa D$$

for some  $\kappa > 1$ , the eigenvalues of  $\tilde{N} = \tilde{D}^{\dagger/2} L \tilde{D}^{\dagger/2}$  is between  $[0, 2\kappa]$ .

Note that, since the degree matrix  $D$  is not necessarily full-rank, for either  $L$ ,  $N$ , or  $\tilde{N}$ , there will always be some “trivial” zero eigenvalues that correspond to the kernel of  $D$ . Therefore we shall focus only on the remaining “nontrivial” eigenvalues and their eigenvectors.

**Definition 7.11** (Nontrivial eigenvalues and eigenvectors). An eigenvalue  $\lambda$  and its associated eigenvector  $f \in \mathbb{R}^{nk}$  of either  $L$ ,  $N$ , or  $\tilde{N}$  are said to be *nontrivial* if we have

$$f_u \perp \ker(D_u), \quad \forall u \in V.$$

**Proposition 7.10.** *The number of nontrivial eigenvalues (counted with multiplicity) of either  $L$ ,  $N$ , or  $\tilde{N}$  is exactly  $\text{rank}(D)$ .*

## 8 Almost regular graphs have only few small eigenvalues

We first give the definition of almost regularity in matrix-weighted graphs.

**Definition 8.1** (Almost regular graphs). Let  $G = (V, E)$  be a  $k \times k$  matrix-weighted graph with edge weights  $\phi_{uv}\phi_{uv}^T$ . For a  $\gamma \geq 1$ , we say  $G$  is  $\gamma$ -almost regular if for all  $u \in V$  and  $(u, v) \in E$ , we have

$$\phi_{uv}^T D_u^\dagger \phi_{uv} \leq \frac{\gamma \cdot k}{n}. \tag{28}$$

We then consider the spectrum of the  $\tilde{D}$ -normalized Laplacian (Definition 7.10) of a graph  $G$  for some semidefinite,  $k \times k$ -block diagonal matrix  $\tilde{D}$  with the same range as  $D$ . Specifically, we show that when  $G$  is almost regular and  $\tilde{D}$  is close to  $D$ , the number of small nontrivial eigenvalues (Definition 7.11) is small.

**Theorem 8.1.** Let  $G = (V, E)$  be a  $k \times k$  matrix-weighted,  $\gamma$ -almost regular graph. Suppose  $\tilde{D}$  is a semidefinite,  $k \times k$ -block diagonal matrix satisfying that

$$\frac{1}{\kappa}D \preceq \tilde{D} \preceq \kappa D$$

for some  $\kappa \geq 1$ . Then for any  $\delta \in (0, 1)$ , the number of nontrivial eigenvalues of  $\tilde{N} = \tilde{D}^{\dagger/2}L\tilde{D}^{\dagger/2}$  that are at most  $\frac{1}{\kappa} - \delta$  is at most  $\frac{\gamma\kappa^2k^2}{\delta^2}$ .

This theorem will be a consequence a lemma that characterizes certain properties of the “spectral embedding” induced by the bottom eigenvectors of  $\tilde{N}$ . We note that similar spectral embeddings were previously used to prove higher-order Cheeger inequalities for scalar-weighted graphs [LRTV12, LGT14].

**Definition 8.2** (Spectral embeddings). Given orthonormal vectors  $f_1, \dots, f_\ell \in \mathbb{R}^{nk}$ , define an  $\ell \times nk$  matrix  $\mathcal{F}$  whose rows are transposes of  $f_1, \dots, f_\ell$ :

$$\mathcal{F} = \begin{array}{|c|c|c|c|c|} \hline & (f_1)_1^T & \cdots & (f_1)_u^T & \cdots & (f_1)_n^T \\ \hline & (f_2)_1^T & \cdots & (f_2)_u^T & \cdots & (f_2)_n^T \\ \hline & & & \vdots & & \\ \hline & (f_\ell)_1^T & \cdots & (f_\ell)_u^T & \cdots & (f_\ell)_n^T \\ \hline \end{array} \in \mathbb{R}^{\ell \times nk}.$$

Then define an embedding  $F : V \rightarrow \mathbb{R}^{\ell \times k}$  by letting  $F(u)$  equal the  $u^{\text{th}}$  column block of  $\mathcal{F}$ :

$$F(u) = \begin{array}{|c|} \hline (f_1)_u^T \\ \hline (f_2)_u^T \\ \hline \vdots \\ \hline (f_\ell)_u^T \\ \hline \end{array} \in \mathbb{R}^{\ell \times k}.$$

We call  $F$  the spectral embedding induced by  $f_1, \dots, f_\ell$ , and  $\mathcal{F}$  the embedding matrix induced by  $f_1, \dots, f_\ell$ .

Essentially, the following lemma says that the spectral embedding induced by the (nontrivial) bottom eigenvectors has norm spread out across a large number of vertices.

**Lemma 8.2.** Let  $G = (V, E)$  be a  $k \times k$  matrix-weighted,  $\gamma$ -almost regular graph. Suppose  $\tilde{D}$  is a semidefinite,  $k \times k$ -block diagonal matrix satisfying that

$$\frac{1}{\kappa}D \preceq \tilde{D} \preceq \kappa D$$

for some  $\kappa \geq 1$ . Fix a  $\delta \in (0, 1)$  and let  $0 \leq \tilde{\lambda}_1 \leq \dots \leq \tilde{\lambda}_\ell \leq \frac{1}{\kappa} - \delta$  be all nontrivial eigenvalues of  $\tilde{N} = \tilde{D}^{\dagger/2}L\tilde{D}^{\dagger/2}$  that are  $\leq \frac{1}{\kappa} - \delta$ . Let  $\tilde{f}_1, \dots, \tilde{f}_\ell$  be a corresponding set of orthonormal, nontrivial eigenvectors, which by the definition of nontriviality satisfies

$$(\tilde{f}_i)_u \perp \ker(D_u), \quad \forall i \in [\ell], u \in V.$$

Let  $\tilde{F}$  be the spectral embedding induced by  $\tilde{f}_1, \dots, \tilde{f}_\ell$ . Then we have for all  $u \in V$

$$\lambda_{\max}(\tilde{F}(u)^T \tilde{F}(u)) \leq \frac{\gamma \kappa^2 k}{\delta^2 n}, \quad (29)$$

where  $\lambda_{\max}(\cdot)$  denotes taking the largest eigenvalue.

*Proof of Theorem 8.1 using Lemma 8.2.* Let  $\ell$  be the number of nontrivial eigenvalues of  $\tilde{N}$  that are  $\leq \frac{1}{\kappa} - \delta$  and let  $\tilde{f}_1, \dots, \tilde{f}_\ell$  be a corresponding set of orthonormal, nontrivial eigenvectors. Let  $\tilde{F}$  be the spectral embedding induced by  $\tilde{f}_1, \dots, \tilde{f}_\ell$ . Since all  $\tilde{f}_i$ 's are unit vectors, we have

$$\sum_u \text{Tr}(\tilde{F}(u)^T \tilde{F}(u)) = \sum_u \|\tilde{F}(u)\|_F^2 = \sum_u \sum_{i=1}^{\ell} \|(\tilde{f}_i)_u\|^2 = \ell.$$

Therefore there must exist a vertex  $u$  with

$$\text{Tr}(\tilde{F}(u)^T \tilde{F}(u)) \geq \frac{\ell}{n}.$$

As  $\tilde{F}(u)^T \tilde{F}(u)$  is a  $k \times k$  positive semidefinite matrix, we have

$$\lambda_{\max}(\tilde{F}(u)^T \tilde{F}(u)) \geq \frac{1}{k} \cdot \text{Tr}(\tilde{F}(u)^T \tilde{F}(u)) \geq \frac{\ell}{nk}.$$

On the other hand, by Lemma 8.2,  $\lambda_{\max}(\tilde{F}(u)^T \tilde{F}(u)) \leq \frac{\gamma \kappa^2 k}{\delta^2 n}$ . Thus we must have  $\ell \leq \frac{\gamma \kappa^2 k^2}{\delta^2}$ .  $\square$

*Proof of Lemma 8.2.* Let  $\tilde{\mathcal{F}}$  be the embedding matrix induced by  $\tilde{f}_1, \dots, \tilde{f}_\ell$ . Suppose for the sake of contradiction (29) is violated by some  $u \in V$ . That is, we have

$$\lambda_{\max}(\tilde{F}(u)^T \tilde{F}(u)) > \frac{\gamma \kappa^2 k}{\delta^2 n}. \quad (30)$$

Define a diagonal matrix  $\tilde{\Lambda} \in \mathbb{R}^{\ell \times \ell}$  by

$$\tilde{\Lambda} = \begin{pmatrix} \tilde{\lambda}_1 & & \\ & \ddots & \\ & & \tilde{\lambda}_\ell \end{pmatrix} \in \mathbb{R}^{\ell \times \ell}.$$

Then, since  $\tilde{\lambda}_1, \dots, \tilde{\lambda}_\ell$  and  $\tilde{f}_1, \dots, \tilde{f}_\ell$  are corresponding eigenvalues and eigenvectors of  $\tilde{N}$ , we immediately have (recall that  $\tilde{\mathcal{F}}$ 's rows are  $\tilde{f}_1^T, \dots, \tilde{f}_\ell^T$ )

$$\tilde{\Lambda} \tilde{\mathcal{F}} = \tilde{\mathcal{F}} \tilde{N},$$

and in particular, by restricting to the  $u^{\text{th}}$  column block on both sides,

$$\tilde{\Lambda} \tilde{F}(u) = \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} D_u \tilde{D}_{uu}^{\dagger/2} - \sum_{v \sim u} \tilde{F}(v) \tilde{D}_{vv}^{\dagger/2} \phi_{uv} \phi_{uv}^T \tilde{D}_{uu}^{\dagger/2}. \quad (31)$$

Now consider the  $\ell \times \ell$  matrix  $\tilde{F}(u)\tilde{F}(u)^T$ , whose largest eigenvalue is equal to  $\lambda_{\max}(\tilde{F}(u)\tilde{F}(u)^T)$ . Let  $g \in \mathbb{R}^\ell$  be a unit eigenvector corresponding to  $\lambda_{\max}(\tilde{F}(u)\tilde{F}(u)^T)$ . Since  $f_1, \dots, f_\ell$  are orthonormal, we have

$$\begin{aligned} \sum_{v \in V} g^T \tilde{F}(v) \tilde{F}(v)^T g &= g^T \left( \sum_{v \in V} \tilde{F}(v) \tilde{F}(v)^T \right) g \\ &= g^T (\tilde{\mathcal{F}} \tilde{\mathcal{F}}^T) g \\ &= g^T I g \\ &= \|g\|^2 = 1. \end{aligned} \quad (32)$$

We will then show a contradiction by arguing that (31) cannot be true given (30). First by multiplying  $g^T$  to the left and  $\tilde{F}(u)^T g$  to the right to both sides of (31) we have

$$g^T \tilde{\Lambda} \tilde{F}(u) \tilde{F}(u)^T g = g^T \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} D_u \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g - \sum_{u \sim v} g^T \tilde{F}(v) \tilde{D}_{vv}^{\dagger/2} \phi_{uv} \phi_{uv}^T \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g. \quad (33)$$

Using the fact that  $g$  is a unit eigenvector corresponding to  $\lambda_{\max}(\tilde{F}(u)\tilde{F}(u)^T)$ , we can upper bound the LHS of (33) by

$$\begin{aligned} g^T \tilde{\Lambda} \tilde{F}(u) \tilde{F}(u)^T g &= \lambda_{\max}(\tilde{F}(u)\tilde{F}(u)^T) \cdot g^T \tilde{\Lambda} g \\ &\leq \left( \frac{1}{\kappa} - \delta \right) \cdot \lambda_{\max}(\tilde{F}(u)\tilde{F}(u)^T), \end{aligned}$$

where the last inequality holds since all eigenvalues  $\tilde{\lambda}_1, \dots, \tilde{\lambda}_\ell \leq \frac{1}{\kappa} - \delta$ . We can also lower bound the first term on the RHS of (33) by

$$g^T \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} D_u \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g \geq \frac{1}{\kappa} \cdot g^T \tilde{F}(u) \tilde{F}(u)^T g = \frac{1}{\kappa} \cdot \lambda_{\max}(\tilde{F}(u)\tilde{F}(u)^T),$$

where the first inequality follows from  $\tilde{D} \preceq \kappa D$ . Therefore, in order to contradict (33), it suffices to show that the second term on the RHS of (33) satisfies

$$\sum_{u \sim v} g^T \tilde{F}(v) \tilde{D}_{vv}^{\dagger/2} \phi_{uv} \phi_{uv}^T \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g < \delta \cdot \lambda_{\max}(\tilde{F}(u)\tilde{F}(u)^T).$$

We then do so by applying the Cauchy-Schwarz inequality:

$$\begin{aligned}
& \sum_{u \sim v} g^T \tilde{F}(v) \tilde{D}_{vv}^{\dagger/2} \phi_{uv} \phi_{uv}^T \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g \\
& \leq \sqrt{\sum_{u \sim v} g^T \tilde{F}(v) \tilde{F}(v)^T g} \sqrt{\sum_{u \sim v} g^T \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} \phi_{uv} \phi_{uv}^T \tilde{D}_{vv}^{\dagger} \phi_{uv} \phi_{uv}^T \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g} \\
& \leq 1 \cdot \sqrt{\sum_{u \sim v} g^T \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} \phi_{uv} \phi_{uv}^T \tilde{D}_{vv}^{\dagger} \phi_{uv} \phi_{uv}^T \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g} \quad (\text{by (32)}) \\
& \leq \sqrt{\frac{\gamma \kappa \cdot k}{n}} \cdot \sqrt{\sum_{u \sim v} g^T \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} \phi_{uv} \phi_{uv}^T \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g} \\
& \quad (\text{since } \phi_{uv}^T \tilde{D}_{vv}^{\dagger} \phi_{uv} \leq \frac{\gamma \kappa \cdot k}{n} \text{ by } \gamma\text{-almost regularity and } \tilde{D} \succeq \frac{1}{\kappa} D) \\
& = \sqrt{\frac{\gamma \kappa \cdot k}{n}} \sqrt{g^T \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} \left( \sum_{u \sim v} \phi_{uv} \phi_{uv}^T \right) \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g} \\
& = \sqrt{\frac{\gamma \kappa \cdot k}{n}} \sqrt{g^T \tilde{F}(u) \tilde{D}_{uu}^{\dagger/2} D_u \tilde{D}_{uu}^{\dagger/2} \tilde{F}(u)^T g} \\
& \leq \sqrt{\frac{\gamma \kappa \cdot k}{n}} \sqrt{\kappa} \sqrt{g^T \tilde{F}(u) \tilde{F}(u)^T g} \quad (\text{by } \tilde{D} \succeq \frac{1}{\kappa} D) \\
& = \sqrt{\frac{\gamma \kappa^2 k}{n}} \sqrt{\lambda_{\max}(\tilde{F}(u) \tilde{F}(u)^T)} \\
& < \delta \cdot \lambda_{\max}(\tilde{F}(u) \tilde{F}(u)^T) \quad (\text{since by our assumption (30) } \lambda_{\max}(\tilde{F}(u)^T \tilde{F}(u)) > \frac{\gamma \kappa^2 k}{\delta^2 n}).
\end{aligned}$$

Therefore it must be the case that all  $\lambda_{\max}(\tilde{F}(u)^T \tilde{F}(u)) \leq \frac{\gamma \kappa^2 k}{\delta^2 n}$ , as desired  $\square$

## 9 Almost regular graph decomposition

In this section, we show that every matrix-weighted graph that is sufficiently dense can be made into an almost regular graph (in the sense of Definition 8.1) by downscaling a small number of edges.

Let us first introduce some notations for rescaled graphs. For a  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv} \phi_{uv}^T$ 's and a scaling  $s : E \rightarrow [0, 1]$ , we will write  $G^s$  to denote the graph obtained from  $G$  by rescaling each edge  $(u, v)$ 's weight to  $(s_{uv} \phi_{uv})(s_{uv} \phi_{uv})^T$ . For simplicity we will use the superscript  $s$  when dealing with vectors and matrices associated with  $G^s$ . For example, we have  $\phi_{uv}^s = s_{uv} \phi_{uv}$ ,  $D_u^s = \sum_{u \sim v} (s_{uv} \phi_{uv})(s_{uv} \phi_{uv})^T$ , and  $L^s = \sum_{(u,v) \in E} s_{uv}^2 b_{uv} b_{uv}^T$ . Similarly, for a subset of edges  $F \subseteq E$ , we also use the superscript  $F$  to denote matrices associated with the induced subgraph  $G[F]$ , and thus, for example,  $D_u^F = \sum_{(u,v) \in F} \phi_{uv} \phi_{uv}^T$ .

Our main result in this section is a deterministic algorithm for finding a large almost regular subgraph within any given graph that is sufficiently dense.

**Theorem 9.1.** *There is a deterministic algorithm ALMOSTREGULARDECOMP that, given any  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv}\phi_{uv}^T$ 's and any  $\gamma \geq 1$ , outputs a scaling  $s : E \rightarrow [0, 1]$  such that*

1. *The rescaled graph  $G^s$  is  $\gamma$ -almost regular:*

$$(\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s \leq \frac{\gamma \cdot k}{n}, \quad \forall (u, v) \in E.$$

2. *The number of edges  $(u, v) \in E$  with  $s_{uv} \in (0, 1)$  is at most  $\frac{8n^2}{\gamma}$ .*
3. *The number of edges  $(u, v) \in E$  with  $s_{uv} = 0$  is at most  $\frac{8n}{\gamma \cdot k}(\text{rank}(D) - \text{rank}(D^s))$ .*

*The algorithm terminates in finite time.*

Note that since  $\text{rank}(D) \leq nk$ , the number of edges with  $s_{uv} = 0$  is at most  $\frac{8n^2}{\gamma}$ . Therefore, by setting  $\gamma$  to be  $\geq \frac{32n^2}{|E|}$ , the total fraction of edges with  $s_{uv} < 1$  is no more than  $1/2$ . Also note that our goal is to prove the existence of such a scaling for any given weights  $\phi_{uv}\phi_{uv}^T$ 's that can potentially have *infinite* precision, so we only focus on designing an algorithm that terminates in finite time, as opposed to giving an explicit bound on its running time.

Our approach for proving Theorem 9.1 is motivated by the one in [CLM<sup>+</sup>15], using which the authors showed that given a set of vectors, one can, by downscaling a small number of them, make every vector have small leverage score compared to the average. The main difference between our result and theirs is that we have an additional bound on the number of completely deleted edges that is proportional to the rank reduction (the third guarantee in Theorem 9.1), which will be useful when we incorporate Theorem 9.1 into our expander decomposition later in Section 10.

We now give the pseudocodes of our algorithm ALMOSTREGULARDECOMP and its subroutine WHACK-A-MOLE-I below. At a high level, in ALMOSTREGULARDECOMP, we first try to eliminate edges with high leverage scores<sup>8</sup>, by iteratively halving the weight of any such edge for some large number of iterations. The latter iterative process is essentially what WHACK-A-MOLE-I does. Then after WHACK-A-MOLE-I terminates, either (i) all edges have small leverage scores, in which case we are done, or (ii) the edge weights have become sufficiently divergent so that we can decrease the rank of  $D$  by deleting a small number of edges. If it is the second case, we then delete those edges and restart the process from beginning. We will show that the total number of deleted/rescaled edges is small, so that we end up with a large almost regular graph.

---

<sup>8</sup>Recall that in Definition 7.4 we call the quantity  $(\phi_{uv})^T (D_u)^\dagger \phi_{uv}$  the leverage of edge  $(u, v)$  w.r.t. vertex  $u$ .

---

**Algorithm 1:** WHACK-A-MOLE-I( $G, \gamma, T, s^0$ )

---

**Input :**  $G$ : a  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv}\phi_{uv}^T$ .  
 $\gamma$ : an almost regularity parameter.  
 $T$ : an integer denoting the iteration count.  
 $s^0$ : an initial scaling, mapping from  $E \rightarrow [0, 1]$ .

**Output:**  $s$ : a scaling, mapping from  $E \rightarrow [0, 1]$ .

1 Let  $s \leftarrow s^0$ .  
2 **for**  $t \leftarrow 1$  to  $T$  **do**  
3   **if**  $\exists u \in V, (u, v) \in E$  with  $(\phi_{uv}^s)^T(D_u^s)^\dagger\phi_{uv}^s > \frac{\gamma \cdot k}{n}$  **then**  
4     Pick an arbitrary pair of such  $u$  and  $(u, v)$ .  
5      $s_{uv} \leftarrow s_{uv}/2$ .  
6 **return**  $s$ .

---



---

**Algorithm 2:** ALMOSTREGULARDECOMP( $G, \gamma$ )

---

**Input :**  $G$ : a  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv}\phi_{uv}^T$ .  
 $\gamma$ : an almost regularity parameter.

**Output:**  $s$ : a scaling, mapping from  $E \rightarrow [0, 1]$ .

1 Initially, let  $s_{uv} \leftarrow 1$  for all  $(u, v) \in E$ .  
2 Let  $\ell \leftarrow \min_{(u,v) \in E} \|\phi_{uv}\|^2$  and  $r \leftarrow \max_{(u,v) \in E} \|\phi_{uv}\|^2$ .  
3 Let  $\alpha \leftarrow \max \left\{ 8n^4k, 4nk \cdot r \cdot \max_{F \subseteq E} \mu_{\max} \left( (D^F)^\dagger \right) \right\}$ .  
4 Let  $T \leftarrow n^2(4n^2k \log \alpha + \log(r/\ell))$ .  
5 **while**  $\text{rank}(D^s) > 0$  **do**  
6    $s \leftarrow \text{WHACK-A-MOLE-I}(G, \gamma, T, s)$ .  
7   **if**  $\forall u \in V, (u, v) \in E$  we have  $(\phi_{uv}^s)^T(D_u^s)^\dagger\phi_{uv}^s \leq \frac{\gamma \cdot k}{n}$  **then**  
8     Return  $s$  and halt.  
9   **else**  
10     Compute all eigenvalues of  $D^s$  and let  $0 < \mu_1 \leq \dots \leq \mu_p$  be the positive ones.  
11     Let  $\ell^s \leftarrow \min_{(u,v) \in E, s_{uv} > 0} \|\phi_{uv}^s\|^2$   
12     Find the smallest  $\rho \geq \ell^s$  such that all  $\|\phi_{uv}^s\|^2$  and  $\mu_i$  fall in  $\mathbb{R} \setminus (\rho, \rho \cdot \alpha)$ .  
13     Set  $s_{uv} \leftarrow 0$  for all  $(u, v) \in E$  with  $\|\phi_{uv}^s\|^2 \leq \rho$ .  
14     Set  $s_{uv} \leftarrow 1$  for all  $(u, v) \in E$  with  $s_{uv} > 0$ .  
15 **return**  $s$ .

---

**Analysis of Algorithm 1.** We first show that after WHACK-A-MOLE-I terminates, all edges that have been downscaled have large leverage scores w.r.t. at least one of their endpoints.

**Lemma 9.2.** Suppose the initial scaling  $s^0$  satisfies that  $s_{uv}^0 \in \{0, 1\}$  for all  $(u, v) \in E$ . Then after WHACK-A-MOLE-I terminates, for any  $(u, v) \in E$  such that  $s_{uv} \in (0, 1)$ , the leverage score

of  $(u, v)$  w.r.t. at least one of  $u, v$  is  $\geq \frac{\gamma \cdot k}{4n}$ . That is, at least one of the following two statements is true:

1.  $(\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s \geq \frac{\gamma \cdot k}{4n}$ .
2.  $(\phi_{uv}^s)^T (D_v^s)^\dagger \phi_{uv}^s \geq \frac{\gamma \cdot k}{4n}$ .

*Proof.* Fix an edge  $(u, v) \in E$  for which  $s_{uv} \in (0, 1)$  after WHACK-A-MOLE-I terminates. Since initially  $s_{uv}^0 \in \{0, 1\}$ , it must be the case that  $s_{uv}^0 = 1$ , and thus  $s_{uv}$  must change at least once during the process. Let  $t_{uv} \in [1, T]$  be the last for loop iteration in which  $s_{uv}$  changes. Then we know that at the beginning of iteration  $t_{uv}$ , we have either  $(\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s > \frac{\gamma \cdot k}{n}$  or  $(\phi_{uv}^s)^T (D_v^s)^\dagger \phi_{uv}^s > \frac{\gamma \cdot k}{n}$ . Suppose w.l.o.g. it is the first case. We first show that after we set  $s_{uv} \leftarrow s_{uv}/2$ , we have  $(\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s \geq \frac{\gamma \cdot k}{4n}$ . For clarity let us write  $s$  and  $s'$  to denote the scalings before and after the update in iteration  $t_{uv}$  respectively. Then we have

$$\begin{aligned} (\phi_{uv}^{s'})^T (D_u^{s'})^\dagger \phi_{uv}^{s'} &\geq (\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^{s'} \quad (D_u^{s'} = D_u^s - \frac{3s_{uv}^2}{4} \phi_{uv} \phi_{uv}^T \preceq D_u^s) \\ &= \frac{1}{4} (\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s \geq \frac{\gamma \cdot k}{4n} \quad ((\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s > \frac{\gamma \cdot k}{n}). \end{aligned}$$

After iteration  $t_{uv}$ , we only downscale the edge weights of other edges. Note that for any  $s''$  that is pointwise smaller than  $s'$ , we have  $D_u^{s''} \preceq D_u^{s'}$ , and thus the leverage score of  $(u, v)$  w.r.t.  $u$  can only increase afterwards. Therefore we have  $(\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s \geq \frac{\gamma \cdot k}{4n}$  in the end, as desired.  $\square$

As a consequence of Lemma 9.2, the total number of edges that have been downscaled during WHACK-A-MOLE-I is small.

**Corollary 9.3** (of Lemma 9.2). *Suppose the initial scaling  $s^0$  satisfies that  $s_{uv}^0 \in \{0, 1\}$  for all  $(u, v) \in E$ . Then after WHACK-A-MOLE-I terminates, the number of edges  $(u, v) \in E$  with  $s_{uv} \in (0, 1)$  is at most  $\frac{4n \cdot \text{rank}(D^s)}{\gamma \cdot k}$ .*

*Proof.* Note that

$$\begin{aligned} \sum_{(u, v) \in E} \left( (\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s + (\phi_{uv}^s)^T (D_v^s)^\dagger \phi_{uv}^s \right) &= \sum_{u \in V} \sum_{u \sim v} (\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s \\ &= \sum_{u \in V} \sum_{u \sim v} \text{Tr} \left( (D_u^s)^\dagger (\phi_{uv}^s) (\phi_{uv}^s)^T \right) \\ &= \sum_{u \in V} \text{Tr} \left( (D_u^s)^\dagger \left( \sum_{u \sim v} (\phi_{uv}^s) (\phi_{uv}^s)^T \right) \right) \\ &= \sum_{u \in V} \text{Tr} \left( (D_u^s)^\dagger D_u^s \right) \\ &= \sum_{u \in V} \text{rank}(D_u^s) = \text{rank}(D^s). \end{aligned}$$

By Lemma 9.2, for edges with  $s_{uv} \in (0, 1)$ , at least one of  $(\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s$ ,  $(\phi_{uv}^s)^T (D_v^s)^\dagger \phi_{uv}^s$  is  $\geq \frac{\gamma \cdot k}{4n}$ . Therefore there cannot be more than  $\text{rank}(D^s) / (\frac{\gamma \cdot k}{4n}) = \frac{4n \cdot \text{rank}(D^s)}{\gamma \cdot k}$  such edges.  $\square$

**Analysis of Algorithm 2.** We first show that all edges that are deleted at Line 13 have been downscaled during WHACK-A-MOLE-I, and thus by Lemma 9.2 have large leverage scores w.r.t. at least one of their endpoints.

**Lemma 9.4.** *The  $\rho$  found at Line 12 has the property that for all edges  $(u, v) \in E$  such that  $\|\phi_{uv}^s\|^2 \leq \rho$  at the moment (that is, they will be deleted immediately at the next line), we have  $s_{uv} < 1$ .*

*Proof.* First note that before Line 12 we have run WHACK-A-MOLE-I for  $T = n^2(4n^2k \log \alpha + \log(r/\ell))$  iterations. Therefore there must be an edge  $(u, v)$  with  $s_{uv} \in (0, 1)$  whose weight has been halved at least  $4n^2k \log \alpha + \log(r/\ell)$  times. Therefore for such an edge  $(u, v)$  we have

$$\begin{aligned} \ell^s &= \min_{(u', v') \in E, s_{u'v'} > 0} \|\phi_{u'v'}^s\|^2 \leq \|\phi_{uv}^s\|^2 \\ &\leq \frac{1}{4^{4n^2k \log \alpha + \log(r/\ell)}} \|\phi_{uv}\|^2 \\ &= \frac{1}{\alpha^{8n^2k}} \cdot \frac{\ell^2}{r^2} \cdot \|\phi_{uv}\|^2 \\ &\leq \frac{1}{\alpha^{8n^2k}} \cdot \min_{(u, v) \in E} \|\phi_{uv}\|^2. \end{aligned}$$

Therefore for any  $(u, v)$  whose  $\|\phi_{uv}^s\|^2$  falls in  $[\ell^s, \ell^s \cdot \alpha^{8n^2k}]$ , it must be the case that  $s_{uv} < 1$ . Also, note that the total number of different  $\|\phi_{uv}^s\|^2$ 's and  $\mu_i$ 's is at most  $n^2 + nk \leq 2n^2k$ . Therefore, as long as  $nk > 1$ , there must be an interval  $(\rho, \rho \cdot \alpha) \subseteq [\ell^s, \ell^s \cdot \alpha^{8n^2k}]$  where no  $\|\phi_{uv}^s\|^2$  or  $\mu_i$  resides. Therefore the  $\rho$  found at Line 12 must be  $< \ell^s \cdot \alpha^{8n^2k}$ . So for all edges  $(u, v) \in E$  such that  $\|\phi_{uv}^s\|^2 \leq \rho$  we must have  $s_{uv} < 1$ .  $\square$

We then show that the number of edges deleted at Line 13 of ALMOSTREGULARDECOMP is small compared to the rank reduction. To this end let us fix an iteration of the while loop where we go to the “else” branch. Let  $s$  denote the scaling returned by WHACK-A-MOLE-I, and let  $s'$  denote the scaling obtained after we delete the small weight edges at Line 13 (but before we reset the weights of other edges at Line 14).

**Lemma 9.5.** *The number of edges  $(u, v) \in E$  with  $\|\phi_{uv}^s\|^2 \in (0, \rho]$  at Line 13 is at most*

$$\frac{8n}{\gamma \cdot k} \left( \text{rank}(D^s) - \text{rank}(D^{s'}) \right).$$

**Lemma 9.6.** *We have*

$$\mu_{\max} \left( \left( D^{s'} \right)^{\dagger} \right) \leq \frac{1}{4nk \cdot \rho}.$$

*Proof.* Let  $s'_{\min} = \min \{s'_{uv} : s'_{uv} \neq 0\}$ . Note that we have

$$(s'_{\min})^2 = \min_{s'_{uv} > 0} \frac{\|\phi_{uv}^{s'}\|^2}{\|\phi_{uv}\|^2} \geq \frac{\min_{s'_{uv} > 0} \|\phi_{uv}^{s'}\|^2}{\max_{(u,v) \in E} \|\phi_{uv}\|^2} \geq \frac{\rho \cdot \alpha}{r}.$$

Let  $F' = \{(u, v) \in E : s'_{uv} > 0\}$ . Then we have

$$D^{s'} \succeq (s'_{\min})^2 D^F \succeq \frac{\rho \cdot \alpha}{r} \cdot D^F,$$

and as a result

$$(D^{s'})^\dagger \preceq \frac{r}{\rho \cdot \alpha} \cdot (D^F)^\dagger. \quad (34)$$

By Line 3, we have

$$\alpha \geq 4nk \cdot r \cdot \max_{F' \subseteq E} \mu_{\max} \left( (D^{F'})^\dagger \right) \geq 4nk \cdot r \cdot \mu_{\max} \left( (D^F)^\dagger \right).$$

Plugging this into (34) gives

$$\mu_{\max} \left( (D^{s'})^\dagger \right) \leq \frac{1}{4nk \cdot \rho}.$$

as desired.  $\square$

To prove Lemma 9.5, we first need to show that the rank of  $D^s$  decreases by a certain amount after deleting small weight edges.

**Lemma 9.7.** *Let  $i$  be such that  $\mu_i \leq \rho < \rho \cdot \alpha \leq \mu_{i+1}$ . Then we have*

$$\text{rank}(D^s) - \text{rank}(D^{s'}) \geq i.$$

*Proof.* Let  $f_1, \dots, f_i \in \mathbb{R}^{nk}$  be a set of orthonormal eigenvectors corresponding to  $\mu_1, \dots, \mu_i$ . Let  $\mathcal{S} \subseteq \mathbb{R}^{nk}$  be the column space of  $D^{s'}$ . Then it suffices to show that after projecting  $f_1, \dots, f_i$  onto the subspace  $\mathcal{S}^\perp$  (the subspace orthogonal to  $\mathcal{S}$ ), they are still linearly independent. To this end, we first note that for each  $f_j$  where  $j \leq i$ , its projection onto  $\mathcal{S}$  can be written as

$$\Pi_{\mathcal{S}} f_j = (D^{s'})^{\dagger/2} (D^{s'})^{1/2} f_j.$$

Using the fact that  $f_j$  is an eigenvector of  $D^s$  corresponding to eigenvalue  $\mu_j \leq \rho$ , we have

$$\begin{aligned} \left\| (D^{s'})^{1/2} f_j \right\|^2 &= f_j^T D^{s'} f_j \\ &\leq f_j^T D^s f_j \quad (D^{s'} \preceq D^s) \\ &= \mu_j \leq \rho. \end{aligned}$$

Then by Lemma 9.6,

$$\begin{aligned} \left\| (D^{s'})^{\dagger/2} (D^{s'})^{1/2} f_j \right\|^2 &= f_j^T (D^{s'})^{1/2} (D^{s'})^\dagger (D^{s'})^{1/2} f_j \\ &\leq \mu_{\max} \left( (D^{s'})^\dagger \right) f_j^T D^{s'} f_j \\ &\leq \frac{1}{4nk \cdot \rho} \cdot \rho = \frac{1}{4nk}. \end{aligned} \quad (35)$$

Let  $d$  be the dimension of the subspace  $\mathcal{S}^\perp$ , and let  $b_1, \dots, b_d$  be an orthonormal basis of  $\mathcal{S}^\perp$ . Also, let  $b_{d+1}, \dots, b_{nk} \in \mathbb{R}^{nk}$  be an orthonormal basis of  $\mathcal{S}$ . Then we can write each  $f_j$  as a linear combination of  $b_1, \dots, b_{nk}$ . That is, there exists a matrix  $C \in \mathbb{R}^{i \times nk}$  such that

$$\begin{pmatrix} f_1^T \\ \vdots \\ f_i^T \end{pmatrix} = C \begin{pmatrix} b_1^T \\ \vdots \\ b_{nk}^T \end{pmatrix}.$$

Since this is an orthogonal transformation, we have that the rows of  $C$  are orthonormal, namely  $CC^T = I$ . If we write the  $\tau^{\text{th}}$  column of  $C$  as  $c_\tau \in \mathbb{R}^i$ , then we have  $\sum_{\tau=1}^{nk} c_\tau c_\tau^T = CC^T = I$ . Moreover, we have

$$\sum_{\tau=d+1}^{nk} \|c_\tau\|^2 = \sum_{j=1}^i \|\Pi_{\mathcal{S}} f_j\|^2 \leq i \cdot \frac{1}{4nk} \leq nk \cdot \frac{1}{4nk} = \frac{1}{4},$$

where the first inequality follows from (35). Therefore

$$\text{Tr} \left( \sum_{\tau=1}^d c_\tau c_\tau^T \right) = \text{Tr} \left( \sum_{\tau=1}^{nk} c_\tau c_\tau^T \right) - \text{Tr} \left( \sum_{\tau=d+1}^{nk} c_\tau c_\tau^T \right) \geq i - \frac{1}{4}.$$

On the other hand, we have

$$\sum_{\tau=1}^d c_\tau c_\tau^T \preceq \sum_{\tau=1}^{nk} c_\tau c_\tau^T,$$

and thus all eigenvalues of  $\sum_{\tau=1}^d c_\tau c_\tau^T$  are at most 1. These two imply that  $\sum_{\tau=1}^d c_\tau c_\tau^T$  must be of full rank  $i$ . As a result,  $\Pi_{\mathcal{S}^\perp} f_1, \dots, \Pi_{\mathcal{S}^\perp} f_i$  are linearly independent.  $\square$

*Proof of Lemma 9.5.* By Lemma 9.7 we know that the rank reduction in  $D^s$  is at least  $i$ , where  $i$  is such that  $\mu_i \leq \rho < \rho \cdot \alpha \leq \mu_{i+1}$ . Let  $f_1, \dots, f_p$  be a set of orthonormal eigenvectors corresponding to  $\mu_1, \dots, \mu_p$ . By Lemma 9.4, we know that for any  $(u, v) \in E$  with  $\|\phi_{uv}^s\|^2 \in (0, \rho]$ , we have  $s_{uv} \in (0, 1)$ . Let  $F \stackrel{\text{def}}{=} \{(u, v) \in E : \|\phi_{uv}^s\|^2 \in (0, \rho]\}$ . As the initial scaling  $s^0$  that we send to WHACK-A-MOLE-I always has range  $\{0, 1\}$ , we have by Lemma 9.2 that for each  $e \in F$ , its leverage score is at least  $\frac{\gamma k}{4n}$  w.r.t. at least one of its endpoints.

For each edge  $(u, v)$ , define two  $nk$ -dimensional vectors  $x_{uv}^u$  and  $x_{uv}^v$ , such that  $(x_{uv}^u)_u = \phi_{uv}^s$  and  $(x_{uv}^u)_w = 0$  for all  $w \neq u$ , and  $(x_{uv}^v)_v = \phi_{uv}^s$  and  $(x_{uv}^v)_w = 0$  for all  $w \neq v$ . Then for any  $e \in F$  we have

$$(x_{uv}^u)^T (D^s)^\dagger x_{uv}^u + (x_{uv}^v)^T (D^s)^\dagger x_{uv}^v \geq \frac{\gamma \cdot k}{4n}.$$

Now if we restrict to only the eigenvalues smaller than  $\rho$ , we have

$$\begin{aligned} & (x_{uv}^u)^T \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^\dagger x_{uv}^u + (x_{uv}^v)^T \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^\dagger x_{uv}^v \\ &= (x_{uv}^u)^T \left( (D^s)^\dagger - \sum_{j=i+1}^p \frac{1}{\mu_j} f_j f_j^T \right) x_{uv}^u + (x_{uv}^v)^T \left( (D^s)^\dagger - \sum_{j=i+1}^p \frac{1}{\mu_j} f_j f_j^T \right) x_{uv}^v. \end{aligned} \quad (36)$$

Notice that for all  $e \in F$  we have  $\|x_{uv}^u\|^2 = \|x_{uv}^v\|^2 = \|\phi_{uv}^s\|^2 \leq \rho$ . And thus for any  $j \geq i + 1$ ,

$$\begin{aligned}\frac{1}{\mu_j} \langle x_{uv}^u, f_j \rangle^2 &\leq \frac{\rho}{\mu_j} \leq \frac{1}{\alpha}, \\ \frac{1}{\mu_j} \langle x_{uv}^v, f_j \rangle^2 &\leq \frac{\rho}{\mu_j} \leq \frac{1}{\alpha},\end{aligned}$$

where both second inequalities follow from  $\mu_j \geq \rho \cdot \alpha$  for  $j \geq i + 1$ . Now by Line 3,  $\alpha \geq 8n^4k$ . Therefore combining this with (36) we have

$$\begin{aligned}& \sum_{(u,v) \in F} (x_{uv}^u)^T \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^\dagger x_{uv}^u + (x_{uv}^v)^T \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^\dagger x_{uv}^v \\ &\geq \sum_{(u,v) \in F} (x_{uv}^u)^T (D^s)^\dagger x_{uv}^u + (x_{uv}^v)^T (D^s)^\dagger x_{uv}^v - 2nk \cdot \frac{1}{8n^4k} \\ &\geq |F| \cdot \left( \frac{\gamma \cdot k}{4n} - \frac{1}{4n^3} \right) \geq |F| \cdot \frac{\gamma \cdot k}{8n}.\end{aligned}$$

On the other hand, we have

$$\begin{aligned}& \sum_{(u,v) \in F} (x_{uv}^u)^T \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^\dagger x_{uv}^u + (x_{uv}^v)^T \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^\dagger x_{uv}^v \\ &= \text{Tr} \left( \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} \left( \sum_{(u,v) \in F} x_{uv}^u (x_{uv}^u)^T + x_{uv}^v (x_{uv}^v)^T \right) \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} \right) \\ &\leq \text{Tr} \left( \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} \left( \sum_{(u,v) \in E} x_{uv}^u (x_{uv}^u)^T + x_{uv}^v (x_{uv}^v)^T \right) \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} \right) \\ &= \text{Tr} \left( \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} D^s \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} \right) \\ &= \text{Tr} \left( \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} \left( \sum_{j=1}^p \mu_j f_j f_j^T \right) \left( \sum_{j=1}^i \mu_j f_j f_j^T \right)^{\dagger/2} \right) \quad (\text{spectral decomposition}) \\ &= \text{rank} \left( \sum_{j=1}^i \mu_j f_j f_j^T \right) \leq i.\end{aligned}$$

Combining the above two inequalities we have

$$|F| \leq \frac{8ni}{\gamma \cdot k} \leq \frac{8n}{\gamma \cdot k} \left( \text{rank}(D^s) - \text{rank}(D^{s'}) \right),$$

as desired.  $\square$

*Proof of Theorem 9.1.* Since in each while loop iteration we zero out the weight of at least one edge, the while loop must terminate in  $O(n^2)$  iterations, and thus the algorithm terminates in finite time. Also by the termination condition of the while loop, the resulting graph  $G^s$  is  $\gamma$ -almost regular. Finally, Corollary 9.3 and Lemma 9.5 give the desired bounds on the number of downscaled/deleted edges.  $\square$

## 10 Almost regular expander decomposition

In this section, we show that every matrix-weighted graph that is sufficiently dense can be made into an almost regular “expander” graph, by downscaling a small number of edges. This can be seen as a matrix-weighted analog of the celebrated expander decomposition of scalar-weighted graphs [GR99, KVV04, ST04].

We first give the definition of matrix-weighted expanders. Consider a  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv}\phi_{uv}^T$ . Let  $\lambda_1 \leq \dots \lambda_r$ , where  $r = \text{rank}(D)$ , be the nontrivial eigenvalues (Definition 7.11) of its normalized Laplacian  $N = D^{\dagger/2}LD^{\dagger/2}$ . Let  $f_1, \dots, f_r$  be a corresponding set of orthonormal, nontrivial eigenvectors, which by definition of nontriviality satisfies

$$(f_i)_u \perp \ker(D_u), \quad \forall i \in [r], u \in V.$$

**Definition 10.1** (Almost regular matrix-weighted expanders). For  $\gamma \geq 1$ ,  $\zeta \in (0, 1)$ , and  $\psi \geq 1$ , we say  $G$  is a  $(\gamma, \zeta, \psi)$ -almost regular expander if

1. ( $\gamma$ -almost regularity) For every vertex  $u$  and every incident edge  $(u, v) \in E$ , we have

$$\phi_{uv}^T D_u^\dagger \phi_{uv} \leq \frac{\gamma \cdot k}{n}.$$

2.  $((\zeta, \psi)$ -expander) for every edge  $(u, v) \in E$  we have

$$\left( D^{\dagger/2} b_{uv} \right)^T \left( \sum_{\lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} f_i f_i^T \right) \left( D^{\dagger/2} b_{uv} \right) \leq \frac{\psi \cdot k^2}{n^2} \quad (37)$$

**Notations for rescaled graphs.** In describing our main result in this section, we will use similar notations for rescaled graphs as in Section 9. Namely, for a  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv}\phi_{uv}^T$ 's and a scaling  $s : E \rightarrow [0, 1]$ , we will write  $G^s$  to denote the graph obtained from  $G$  by rescaling each edge  $(u, v)$ 's weight to  $(s_{uv}\phi_{uv})(s_{uv}\phi_{uv})^T$ . For ease of presentation we will use the superscript  $s$  when dealing with vectors and matrices associated with  $G^s$ . For instance, we write  $\phi_{uv}^s = s_{uv}\phi_{uv}$ ,  $D^s = \sum_{u \sim v} s_{uv}^2 (e_{u \leftarrow v} e_{u \leftarrow v}^T + e_{v \leftarrow u} e_{v \leftarrow u}^T)$ , and  $L^s = \sum_{u \sim v} s_{uv}^2 b_{uv} b_{uv}^T$ . Analogously, for a subset of edges  $F \subseteq E$ , we also use the superscript  $F$  when dealing with matrices associated with the induced subgraph  $G[F]$ , and thus, for instance,  $D^F = \sum_{(u, v) \in F} (e_{u \leftarrow v} e_{u \leftarrow v}^T + e_{v \leftarrow u} e_{v \leftarrow u}^T)$ .

We will write  $\lambda_1^s \leq \dots \leq \lambda_{\text{rank}(D^s)}^s$  to denote the nontrivial eigenvalues of the normalized Laplacian  $N^s = (D^s)^{\dagger/2} L^s (D^s)^{\dagger/2}$  of  $G^s$ , and write  $f_1^s, \dots, f_{\text{rank}(D^s)}^s$  to denote a corresponding

set of orthonormal, nontrivial eigenvectors. We then define a function  $H_\zeta^s : E \rightarrow \mathbb{R}$  by

$$H_\zeta^s(u, v) \stackrel{\text{def}}{=} \left( (D^s)^{\dagger/2} b_{uv}^s \right)^T \left( \sum_{\lambda_i^s \in (0, \zeta]} \frac{1}{\lambda_i^s} f_i^s (f_i^s)^T \right) \left( (D^s)^{\dagger/2} b_{uv}^s \right).$$

We also define another function  $R^s : E \rightarrow \mathbb{R}$  by

$$\begin{aligned} R^s(u, v) &\stackrel{\text{def}}{=} (e_{u \leftarrow v}^s)^T (D^s)^\dagger e_{u \leftarrow v}^s + (e_{v \leftarrow u}^s)^T (D^s)^\dagger e_{v \leftarrow u}^s \\ &= (\phi_{uv}^s)^T (D_u^s)^\dagger \phi_{uv}^s + (\phi_{uv}^s)^T (D_v^s)^\dagger \phi_{uv}^s. \end{aligned}$$

Then, for our goal, it suffices to find a scaling  $s$  such that for all  $(u, v) \in E$  we have

$$R^s(u, v) \leq \frac{\gamma \cdot k}{n} \quad \text{and} \quad H_\zeta^s(u, v) \leq \frac{\psi \cdot k^2}{n^2}.$$

Our main result here is a deterministic algorithm for finding a large almost regular expander subgraph within any given graph that is sufficiently dense.

**Theorem 10.1.** *There is a deterministic algorithm EXPANDERDECOMP that, given any  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv} \phi_{uv}^T$ 's, any  $\gamma \geq 1$ ,  $\zeta \in (0, 1)$ , and*

$$\psi \geq \frac{1024\gamma^2}{(1 - \zeta)^2},$$

*outputs a scaling  $s : E \rightarrow [0, 1]$  such that*

1. *The rescaled graph  $G^s$  is a  $(\gamma, \zeta, \psi)$ -almost regular expander.*
2. *The number of edges  $(u, v) \in E$  with  $s_{uv} < 1$  is at most*

$$\frac{1000n^2}{\gamma} + \frac{100000n^2\gamma^2k^2}{\psi(1 - \zeta)^2}.$$

*The algorithm terminates in finite time.*

Note that similar to Section 9, our goal is to prove the existence of such a scaling for any given weights  $\phi_{uv} \phi_{uv}^T$ 's that may potentially have *infinite* precision, so we only focus on designing an algorithm that terminates in finite time, as opposed to giving an explicit bound on its running time.

We give the pseudocodes of our algorithm EXPANDERDECOMP in Algorithm 4 and its subroutine WHACK-A-MOLE-II in Algorithm 3. At a high level, in EXPANDERDECOMP, we first try to eliminate edges with large  $H_\zeta^s(u, v)$  values by iteratively halving the weight of any such edge for some large number of iterations, while simultaneously maintaining the almost regularity of the graph, by also iteratively halving the weight of any edge with large  $R^s(u, v)$  value. This iterative process is essentially achieved by WHACK-A-MOLE-II. Then after WHACK-A-MOLE-II terminates, either (i) all edges have small  $H_\zeta^s(u, v)$  and  $R^s(u, v)$  values, in which case we are

done, or (ii) we can strictly increase<sup>9</sup> the number of nontrivial (Definition 7.11) zero eigenvalues of the normalized Laplacian by removing a small number of edges that have low weights after WHACK-A-MOLE-II. Since we maintain the almost regularity of the graph throughout, by Theorem 8.1, the total number of nontrivial zero eigenvalues is small, therefore we will only repeat (ii) for a limited number of times.

Since we want to ensure that the number of rescaled edges is small in WHACK-A-MOLE-II, we maintain the invariant that any edge that has been downscaled satisfies that either  $H_\zeta^s(u, v)$  is large or  $R^s(u, v)$  is large, by iteratively *doubling* the weight of any edge with both values small. In order to make sure that there is consistent progress in this process, we set up two thresholds  $\psi_1 < \psi_2$  for  $H_\zeta^s(u, v)$  and also two thresholds  $\gamma_1 < \gamma_2$  for  $R^s(u, v)$ . Then we halve the weight of an edge  $(u, v)$  if either  $H_\zeta^s(u, v) > \frac{\psi_2 k^2}{n^2}$  or  $R^s(u, v) > \frac{\gamma_2 k}{n}$ , but only double its weight if *both*  $H_\zeta^s(u, v) < \frac{\psi_1 k^2}{n^2}$  and  $R^s(u, v) < \frac{\gamma_1 k}{n}$ . To measure the progress in this elaborate process, we will need to use some global quantities as our potential functions. Specifically, we will design our potential functions based on certain variants of determinant, which we define below.

**Definition 10.2.** For a positive semidefinite matrix  $M \in \mathbb{R}^{N \times N}$  with eigenvalues

$$0 = \lambda_1 = \dots = \lambda_z < \lambda_{z+1} \leq \dots \leq \lambda_N,$$

define

$$\det_+(M) \stackrel{\text{def}}{=} \prod_{\lambda_i > 0} \lambda_i.$$

For an integer  $\ell$ , define

$$\det_\ell(M) \stackrel{\text{def}}{=} \prod_{i=z+1}^{z+\ell} \lambda_i.$$

We will need the following variational characterization of  $\det_\ell$ , whose proof is deferred to Appendix E.

**Lemma 10.2.** Let  $f_1, \dots, f_\ell$  be a set of orthonormal eigenvectors corresponding to the  $\ell$  smallest nonzero eigenvalues of  $N$ , and let  $F = (f_1 \ \dots \ f_\ell) \in \mathbb{R}^{nk \times \ell}$ . Let  $\mathcal{X} \subseteq \mathbb{R}^{nk \times \ell}$  be the set of all full-rank  $nk \times \ell$  matrices whose columns are in the range of  $N$ . Then

$$\det_\ell(N) = \min_{X \in \mathcal{X}} \frac{\det(X^T L X)}{\det(X^T D X)},$$

and a minimizer can be obtained by letting  $X^* = D^{\dagger/2} F$ .

The following matrix determinant lemma will be helpful in our analysis of the potential functions.

**Lemma 10.3** ([Har98]). For an invertible square matrix  $M \in \mathbb{R}^{d \times d}$  and vectors  $x, y \in \mathbb{R}^d$ , we have

$$\det(M + xy^T) = (1 + y^T M^{-1} x) \det(M).$$

---

<sup>9</sup>This statement is a slight simplification of our actual analysis, but helps with getting intuition.

---

**Algorithm 3:** WHACK-A-MOLE-II( $G, \gamma_1, \gamma_2, \zeta, \psi_1, \psi_2, \alpha, s^0$ )

---

**Input :**  $G$ : a  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv}\phi_{uv}^T$ .  
 $\gamma_1 < \gamma_2, \zeta, \psi_1 < \psi_2$ : parameters for almost regularity and expander.  
 $\alpha > 0$ : a gap parameter.  
 $s^0$ : an initial scaling, mapping from  $E \rightarrow [0, 1]$ , s.t.  $G^{s^0}$  is  $\gamma_1/2$ -almost regular.

**Output:**  $s$ : a scaling, mapping from  $E \rightarrow [0, 1]$ .  
 $\rho$ : a threshold.

1 Let  $s \leftarrow s^0$ .

2 **while** true **do**

3    Compute the eigenvalues  $\lambda_1 \leq \dots \leq \lambda_{nk}$  of  $(D^s)^{\dagger/2}L^s(D^s)^{\dagger/2}$ .

4    Compute the eigenvalues  $\mu_1 \leq \dots \leq \mu_{nk}$  of  $D^s$ .

5    **if** no  $\lambda_i$  is in the range  $(0, \zeta]$  **then**

6     Return  $s$  and  $\rho$ , and halt.

7    Let  $\lambda_{\min} = \min_{\lambda_i \in (0, \zeta]} \lambda_i$ .

8    Find the smallest  $\rho \geq \lambda_{\min}$  s.t. all  $(s_{uv}/s_{uv}^0)^2, \mu_i$ , and  $\lambda_i$  fall in  $\mathbb{R} \setminus (\rho, \rho \cdot \alpha)$ .

9    **if**  $\rho < 1$  **then**

10     Return  $s$  and  $\rho$ , and halt.

11    **else if**  $\nexists(u, v) \in E$  with  $H_\zeta^s(u, v) > \frac{\psi_2 \cdot k^2}{n^2}$  **then**

12     Return  $s$  and  $\rho$ , and halt.

13    **else**

14     Let  $s_{uv} \leftarrow s_{uv}/2$  for an arbitrary edge  $(u, v)$  with  $H_\zeta^s(u, v) > \frac{\psi_2 \cdot k^2}{n^2}$ .

15     **while**  $\exists(u, v) \in E$  with  $R^s(u, v) > \frac{\gamma_2 \cdot k}{n}$  **do**

16       Let  $s_{uv} \leftarrow s_{uv}/2$  for an arbitrary such edge  $(u, v)$ .

17     **while**  $\exists(u, v) : s_{uv} < s_{uv}^0$  satisfying both  $R^s(u, v) < \frac{\gamma_1 \cdot k}{n}$  and  $H_\zeta^s(u, v) < \frac{\psi_1 \cdot k^2}{n^2}$  **do**

18       Let  $s_{uv} \leftarrow s_{uv} \cdot 2$  for an arbitrary such edge  $(u, v)$ .

---

---

**Algorithm 4:** EXPANDERDECOMP( $G, \gamma, \zeta, \psi$ )

---

**Input :**  $G$ : a  $k \times k$  matrix-weighted graph  $G = (V, E)$  with edge weights  $\phi_{uv}\phi_{uv}^T$ .  
 $\gamma, \zeta, \psi$ : parameters for almost regularity and expander.

**Output:**  $s$ : a scaling, mapping from  $E \rightarrow [0, 1]$ .

```

1  $r_\phi \leftarrow \max_{(u,v) \in E} \|\phi_{uv}\|^2$ .
2 Let  $\alpha_1 \leftarrow \max_{F \subseteq E} \lambda_{\max} \left( (D^F)^\dagger \right)$ .
3 Let  $\alpha_2 \leftarrow \max_{F \subseteq E} \lambda_{\max} \left( (D^{\dagger/2} L^F D^{\dagger/2})^\dagger \right)$ .
4 Initially, let  $s_e \leftarrow 1$  for all  $e \in E$ .
5 while  $\text{rank}(D^s) > 0$  do
6   Let  $G' = (V, E')$  where  $E' = \{e : s_e > 0\}$ .
7   Let  $s' \leftarrow \text{ALMOSTREGULARDECOMP}(G', \gamma/32)$ , and then  $s_e \leftarrow s'_e$  for all  $e \in E'$ .
8    $s'_{\min} \leftarrow \min_{s'_{uv} > 0} s'_{uv}$ .
9   Let  $\alpha \leftarrow \max \{8n^4k, 128nk^3\gamma \cdot r_\phi \cdot \alpha_1 / (s'_{\min})^2, 128nk^3\gamma \cdot \alpha_2 / (s'_{\min})^2\}$ .
10   $(s, \rho) \leftarrow \text{WHACK-A-MOLE-II}(G, \gamma/16, \gamma, \zeta, \psi/1024, \psi, \alpha, s)$ .
11  if  $\forall (u, v) \in E$  we have  $R^s(u, v) \leq \frac{\gamma \cdot k}{n}$  and  $H_\zeta^s(u, v) \leq \frac{\psi \cdot k^2}{n^2}$  then
12    Return  $s$  and halt.
13  else
14    Set  $s_{uv} \leftarrow 0$  for all  $(u, v) \in E$  with  $(s_{uv}/s'_{uv})^2 \leq \rho$ .
15    Set  $s_{uv} \leftarrow 1$  for all  $(u, v) \in E$  with  $s_{uv} > 0$ .
16 return  $s$ .

```

---

### 10.1 Analysis of Algorithm 3

For a fixed scaling  $s : E \rightarrow [0, 1]$ , let  $\ell$  be the number of eigenvalues of the normalized Laplacian  $N^s$  of  $G^s$  that are in the range  $(0, \zeta]$ . Let  $\ell_0 = \left\lceil \frac{\gamma_2 k^2}{(1-\zeta)^2} \right\rceil$ , which, by Theorem 8.1, is an upper bound on  $\ell$  when the graph  $G^s$  is  $\gamma_2$ -almost regular. Then we consider the following potential function:

$$\Upsilon(s) \stackrel{\text{def}}{=} \begin{cases} \zeta^{\ell_0 - \ell} \det_\ell(N^s) & \ell \leq \ell_0 \\ \det_{\ell_0}(N^s) & \ell > \ell_0. \end{cases} \quad (38)$$

Thus,  $\Upsilon(s)$  is always a product of exactly  $\ell_0$  numbers between  $(0, \zeta]$ . It is not hard to see the following alternative form of  $\Upsilon(s)$ , which will be helpful for analyzing it.

**Proposition 10.4.** *Let  $\lambda_1, \dots, \lambda_{\ell_0}$  be the smallest  $\ell_0$  nonzero eigenvalues of  $N^s$ . Then*

$$\Upsilon(s) = \prod_{i=1}^{\ell_0} \min \{\lambda_i, \zeta\}.$$

**Corollary 10.5** (of Proposition 10.4). *For any  $\ell' \leq \ell_0$ , we have  $\Upsilon(s) \leq \det_{\ell'}(N^s) \zeta^{\ell_0 - \ell'}$ .*

Since we always maintain the  $\gamma_2$ -regularity, we will also need to consider the determinant of the degree matrix,  $\det_+(D)$ , as a potential function. We show that starting from an almost-regular graph,  $\det_+(D)$  can only decrease at a limited speed when edges are downscaled. The proof of the lemma below is deferred to Appendix E.

**Lemma 10.6.** *Let  $G = (V, E)$  be a  $\gamma$ -almost regular graph. Let  $s : E \rightarrow (0, 1]$  be a strictly positive scaling, and let  $S \stackrel{\text{def}}{=} \prod_{e \in E} \frac{1}{s_e}$ . Then we have*

$$\det_+(D^s) \geq \left(1 - \frac{2\gamma k}{n}\right)^{2 \log S} \det_+(D).$$

As a result of the above lemma, we then show that the while loop at Lines 15–16 of WHACK-A-MOLE-II will terminate after a bounded number of iterations. The proof of the following lemma is also deferred to Appendix E.

**Lemma 10.7.** *Suppose the input to WHACK-A-MOLE-II satisfies that  $G^{s_0}$  is  $\gamma_1/2$ -regular and  $\gamma_2 \geq 16\gamma_1$ . Consider the first  $t$  iterations of the outermost while loop. Then the total number of iterations executed so far by the inner while loop at Lines 15–16 is at most  $2t$ .*

We now prove our key lemma, which shows that our potential  $\Upsilon$  will decrease at least as fast as a geometric series with rate bounded away from 1. As a result, the smallest nonzero eigenvalue of the normalized Laplacian must also decrease at a steady rate. This implies that eventually we will be able to find a  $\rho < 1$  at Line 8, and therefore terminate the outermost while loop in finite time.

**Lemma 10.8.** *Suppose the input to WHACK-A-MOLE-II satisfies that  $G^{s_0}$  is  $\gamma_1/2$ -regular,  $\gamma_2 \geq 16\gamma_1$ ,  $\psi_2 \geq 1024\psi_1$ , and*

$$\psi_2 \geq \frac{1024\gamma_2^2}{(1 - \zeta)^2}.$$

*Consider the first  $t$  iterations of the outermost while loop, and let  $s^1$  be the scaling obtained at the end of the  $t^{\text{th}}$  iteration. Then*

$$\frac{\Upsilon(s^1)}{\Upsilon(s^0)} \leq \left(1 - \frac{\psi_2 k^2}{16n^2}\right)^t.$$

*Proof.* We first show that the potential reduces by a certain amount at Line 14.

**Claim 10.9.** Let the scalings before and after one execution of Line 14 be  $s$  and  $s'$  respectively. Then we have

$$\frac{\Upsilon(s')}{\Upsilon(s)} \leq 1 - \frac{\psi_2 k^2}{4n^2}.$$

*Proof of Claim 10.9.* Let  $\ell$  be the number of eigenvalues of  $N^s$  that are between  $(0, \zeta]$ . Since  $G^s$  is  $\gamma_2$ -almost regular, we have  $\ell \leq \ell_0$  by Theorem 8.1. Thus  $\Upsilon(s) = \det_\ell(N^s)\zeta^{\ell_0-\ell}$ . Also, by Corollary 10.5, we have  $\Upsilon(s') \leq \det_\ell(N^{s'})\zeta^{\ell_0-\ell}$ . Therefore, it suffices to show that

$$\frac{\det_\ell(N^{s'})}{\det_\ell(N^s)} \leq 1 - \frac{\psi_2 k^2}{4n^2}. \tag{39}$$

We then do so by considering the variational characterization  $\det_\ell$  from Lemma 10.2. Let  $X = (D^s)^{\dagger/2}F$  and  $X' = (D^{s'})^{\dagger/2}F'$  be the optimal matrix that minimizes the variational characterization of  $\det_\ell(N^s)$  and  $\det_\ell(N^{s'})$  respectively. Here  $F$  and  $F'$  are both  $nk \times \ell$  matrices whose columns are bottom nonzero eigenvectors of  $N^s$  and  $N^{s'}$  respectively.

It then suffices to show that the fraction in the characterization reduces by much after the execution of Line 14, even if we do not switch from  $X$  to  $X'$ , since switching to the latter can only decrease the potential function. Namely, since by the optimality of  $X'$  we have

$$\frac{\det(X^T L^{s'} X)}{\det(X^T D^{s'} X)} \geq \frac{\det((X')^T L^{s'} (X'))}{\det((X')^T D^{s'} (X'))},$$

it suffices to show

$$\left( \frac{\det(X^T L^{s'} X)}{\det(X^T D^{s'} X)} \right) / \left( \frac{\det(X^T L^s X)}{\det(X^T D^s X)} \right) \leq 1 - \frac{\psi_2 k^2}{4n^2}. \quad (40)$$

For the numerator we have by matrix determinant lemma

$$\begin{aligned} \frac{\det(X^T L^{s'} X)}{\det(X^T L^s X)} &= 1 - \frac{3}{4} \cdot (b_{uv}^s)^T X (X^T L^s X)^{-1} X^T (b_{uv}^s) \\ &= 1 - \frac{3}{4} \cdot (b_{uv}^s)^T (D^s)^{\dagger/2} F \left( F^T (D^s)^{\dagger/2} L^s (D^s)^{\dagger/2} F \right)^{-1} F^T (D^s)^{\dagger/2} b_{uv}^s \\ &= 1 - \frac{3}{4} \cdot H_\zeta^s(u, v) \leq 1 - \frac{3\psi_2 k^2}{4n^2}, \end{aligned} \quad (41)$$

where the last equality follows from that  $F$ 's columns are eigenvectors corresponding to all eigenvalues of  $N^s$  that are between  $(0, \zeta]$ . For the denominator, we have, again by matrix determinant lemma,

$$\begin{aligned} \frac{\det(X^T D^{s'} X)}{\det(X^T D^s X)} &\geq 1 - \frac{3}{4} (e_{u \leftarrow v}^s)^T X (X^T D^s X)^{-1} X^T e_{u \leftarrow v}^s - \frac{3}{4} (e_{v \leftarrow u}^s)^T X (X^T D^s X)^{-1} X^T e_{v \leftarrow u}^s \\ &= 1 - \frac{3}{4} (e_{u \leftarrow v}^s)^T X X^T e_{u \leftarrow v}^s - \frac{3}{4} (e_{v \leftarrow u}^s)^T X X^T e_{v \leftarrow u}^s \quad (\text{as } X^T D^s X = F^T F = I) \\ &= 1 - \frac{3}{4} (e_{u \leftarrow v}^s)^T (D^s)^{\dagger/2} F F^T (D^s)^{\dagger/2} (e_{u \leftarrow v}^s) - \frac{3}{4} (e_{v \leftarrow u}^s)^T (D^s)^{\dagger/2} F F^T (D^s)^{\dagger/2} (e_{v \leftarrow u}^s) \\ &= 1 - \frac{3}{4} \|F^T(u)(D_u^s)^{\dagger/2} \phi_{uv}\|^2 - \frac{3}{4} \|F^T(v)(D_v^s)^{\dagger/2} \phi_{uv}\|^2, \end{aligned}$$

where in the last equality we let  $F(u)$  be the  $u^{\text{th}}$  row block of  $F$ . Since  $G^s$  is  $\gamma_2$ -almost regular, we have  $\|(D_u^s)^{\dagger/2} \phi_{uv}\|^2 \leq \frac{\gamma_2 \cdot k}{n}$  and  $\|(D_v^s)^{\dagger/2} \phi_{uv}\|^2 \leq \frac{\gamma_2 \cdot k}{n}$ . By Lemma 8.2, we have  $\lambda_{\max}(F(u)F(u)^T) \leq \frac{\gamma_2 k}{(1-\zeta)^2 n}$ , and  $\lambda_{\max}(F(v)F(v)^T) \leq \frac{\gamma_2 k}{(1-\zeta)^2 n}$ , and therefore

$$\begin{aligned} \|F^T(u)(D_u^s)^{\dagger/2} \phi_{uv}\|^2 &\leq \frac{\gamma_2^2 \cdot k^2}{(1-\zeta)^2 n^2} \\ \|F^T(v)(D_v^s)^{\dagger/2} \phi_{uv}\|^2 &\leq \frac{\gamma_2^2 \cdot k^2}{(1-\zeta)^2 n^2}. \end{aligned}$$

This give us

$$\frac{\det(X^T D^{s'} X)}{\det(X^T D^s X)} \geq 1 - \frac{3\gamma_2^2 \cdot k^2}{2(1-\zeta)^2 n^2}. \quad (42)$$

(41), (42) coupled with  $\psi_2 \geq \frac{1024\gamma_2^2}{(1-\zeta)^2}$  imply (40), which finishes the proof of the claim.  $\square$

We next show that during the first while loop at Lines 15-16, the potential function cannot increase much.

**Claim 10.10.** Consider a fixed iteration of the outermost while loop. Suppose in this iteration of the outer while loop, the total number of iterations executed by the first inner while loop at Lines 15-16 is  $t_1$ . Let  $s, s'$  be the scalings before and after the while loop respectively. Then

$$\frac{\Upsilon(s')}{\Upsilon(s)} \leq \left(1 - \frac{96\gamma_2^2 k^2}{(1-\zeta)^2 n^2}\right)^{-t_1}.$$

*Proof of Claim 10.10.* We know that before Line 14, the graph is  $\gamma_2$ -almost regular. Then at Line 14 we halve the scale of a single edge, so the graph  $G^s$  is  $4\gamma_2$ -almost regular afterwards. We also know that after the while loop terminates at Line 16, the graph  $G^{s'}$  is  $\gamma_2$ -almost regular, by the termination condition of the while loop. Now consider the following process for obtaining  $s'$  from  $s$ .

1. While  $s \neq s'$ :

(a) For each  $(u, v) \in E$  such that  $s_{uv} > s'_{uv}$ , let  $s_{uv} \leftarrow s_{uv}/2$ .

We first argue, by induction, that at the end of each iteration of the while loop of the above process,  $G^s$  is  $4\gamma_2$ -almost regular. As noted above, initially, the graph  $G^s$  is  $4\gamma_2$ -almost regular. For the induction step, consider a fixed iteration, and let  $F$  be the edges  $(u, v)$  for which  $s_{uv} > s'_{uv}$ . Since we decrease the weights of all edges in  $F$  by a same factor, from the point view of leverage scores, it is equivalent to increase the weights of all other edges by a same multiple. Therefore the leverage scores of edges in  $F$  can only decrease, and thus can be at most  $\frac{4\gamma_2 k}{n}$  after this iteration. As for edges  $(u, v)$  not in  $F$ , they satisfy  $s_{uv} = s'_{uv}$ . We know that in  $G^{s'}$  their leverage scores are at most  $\frac{4\gamma_2 k}{n}$ , as  $G^{s'}$  is  $\gamma_2$ -almost regular. Since  $G^s$ 's weights always dominate those of  $G^{s'}$ , their leverage scores in  $G^s$  can only be smaller, and thus at most  $\frac{4\gamma_2 k}{n}$  as well.

We then argue that at any point of the process, the graph is  $16\gamma_2$ -almost regular. This follows by noting that at any point of the algorithm,  $G^s$ 's edge weights are within a factor 4 of those at the end of previous iteration, and those at the end of the current iteration.

We now show that in this process, each time we let  $s_{uv} \leftarrow s_{uv}/2$ , the potential function increases by at most  $(1 - \frac{400\gamma_2^2 k^2}{(1-\zeta)^2 n^2})^{-1}$ , which implies the statement of this claim. Let  $q, q'$  be the scalings before and after one execution of  $s_{uv} \leftarrow s_{uv}/2$ . Let  $\ell$  be the number of nonzero eigenvalues of  $N^q$  that are between  $(0, \zeta]$ , and let  $\ell_1 = \min(\ell, \ell_0)$ . Then we have  $\Upsilon(q) = \det_{\ell_1}(N^q) \zeta^{\ell_0 - \ell_1}$ , and, by Corollary 10.5,  $\Upsilon(q') \leq \det_{\ell_1}(N^{q'}) \zeta^{\ell_0 - \ell_1}$ . Thus it suffices to show

$$\frac{\det_{\ell_1}(N^{q'})}{\det_{\ell_1}(N^q)} \leq \left(1 - \frac{400\gamma_2^2 k^2}{(1-\zeta)^2 n^2}\right)^{-1}. \quad (43)$$

As in our proof of Claim 10.9, we once again consider the variational characterization in Lemma 10.2. Let  $X = (D^q)^{\dagger/2}F$  where  $F \in \mathbb{R}^{nk \times \ell_1}$ 's columns are nonzero bottom eigenvectors of  $N^q$ . The numerator of the characterization can only decrease. For the denominator, we have by matrix determinant lemma

$$\begin{aligned} \frac{\det(X^T D^{q'} X)}{\det(X^T D^q X)} &\geq 1 - \frac{3}{4}(e_{u \leftarrow v}^q)^T X (X^T D^q X)^{-1} X^T e_{u \leftarrow v}^q - \frac{3}{4}(e_{v \leftarrow u}^q)^T X (X^T D^q X)^{-1} X^T e_{v \leftarrow u}^q \\ &= 1 - \frac{3}{4}(e_{u \leftarrow v}^q)^T X X^T e_{u \leftarrow v}^q - \frac{3}{4}(e_{v \leftarrow u}^q)^T X X^T e_{v \leftarrow u}^q \quad (\text{as } X^T D^q X = F^T F = I) \\ &= 1 - \frac{3}{4}(e_{u \leftarrow v}^q)^T (D^q)^{\dagger/2} F F^T (D^q)^{\dagger/2} (e_{u \leftarrow v}^q) - \frac{3}{4}(e_{v \leftarrow u}^q)^T (D^q)^{\dagger/2} F F^T (D^q)^{\dagger/2} (e_{v \leftarrow u}^q) \\ &= 1 - \frac{3}{4} \|F^T(u)(D_u^q)^{\dagger/2} \phi_{uv}\|^2 - \frac{3}{4} \|F^T(v)(D_v^q)^{\dagger/2} \phi_{uv}\|^2, \end{aligned}$$

where in the last equality we let  $F(u)$  be the  $u^{\text{th}}$  row block of  $F$ . Since  $G^q$  is  $16\gamma_2$ -almost regular, we have  $\|(D_u^q)^{\dagger/2} \phi_{uv}\|^2 \leq \frac{16\gamma_2 \cdot k}{n}$  and  $\|(D_v^q)^{\dagger/2} \phi_{uv}\|^2 \leq \frac{16\gamma_2 \cdot k}{n}$ . By Lemma 8.2, we have  $\lambda_{\max}(F(u)F(u)^T) \leq \frac{16\gamma_2 k}{(1-\zeta)^2 n}$ , and  $\lambda_{\max}(F(v)F(v)^T) \leq \frac{16\gamma_2 k}{(1-\zeta)^2 n}$ , and therefore

$$\begin{aligned} \|F^T(u)(D_u^q)^{\dagger/2} \phi_{uv}\|^2 &\leq \frac{256\gamma_2^2 \cdot k^2}{(1-\zeta)^2 n^2} \\ \|F^T(v)(D_v^q)^{\dagger/2} \phi_{uv}\|^2 &\leq \frac{256\gamma_2^2 \cdot k^2}{(1-\zeta)^2 n^2}. \end{aligned}$$

This give us

$$\frac{\det(X^T D^{q'} X)}{\det(X^T D^q X)} \geq 1 - \frac{400\gamma_2^2 \cdot k^2}{(1-\zeta)^2 n^2}. \quad (44)$$

This then implies (43) and finishes the proof of the claim.  $\square$

We next show that during the second while loop at Lines 17-18, the potential function cannot increase much either.

**Claim 10.11.** Consider a fixed iteration of the outermost while loop. Suppose in this iteration of the outer while loop, the total number of iterations executed by the first inner while loop at Lines 17-18 is  $t_2$ . Let  $s, s'$  be the scalings before and after the while loop respectively. Then

$$\frac{\Upsilon(s')}{\Upsilon(s)} \leq \left(1 + \frac{3\psi_1 k^2}{n^2}\right)^{t_2}.$$

*Proof of Claim 10.11.* Let  $q, q'$  be the scalings before and after one execution of  $s_{uv} \leftarrow s_{uv} \cdot 2$ . Let  $\ell$  be the number of nonzero eigenvalues of  $N^q$  between  $(0, \zeta]$ , and let  $\ell_2 = \min(\ell, \ell_0)$ . Then we have  $\Upsilon(q) = \det_{\ell_2}(N^q) \zeta^{\ell_0 - \ell_2}$ , and, by Corollary 10.5,  $\Upsilon(q') \leq \det_{\ell_2}(N^{q'}) \zeta^{\ell_0 - \ell_2}$ . Thus it suffices to show

$$\frac{\det_{\ell_2}(N^{q'})}{\det_{\ell_2}(N^q)} \leq 1 + \frac{3\psi_1 k^2}{n^2}. \quad (45)$$

As in our proofs of Claims 10.9, 10.10, we also consider the variational characterization in Lemma 10.2. Let  $X = (D^q)^{\dagger/2}F$  where  $F \in \mathbb{R}^{nk \times \ell_2}$ 's columns are nonzero bottom eigenvectors of  $N^q$ . The denominator of the characterization can only increase. For the numerator we have by matrix determinant lemma

$$\begin{aligned} \frac{\det(X^T L^q X)}{\det(X^T L^q X)} &= 1 + 3(b_{uv}^q)^T X (X^T L^q X)^{-1} X^T (b_{uv}^q) \\ &= 1 + 3(b_{uv}^q)^T (D^q)^{\dagger/2} F \left( F^T (D^q)^{\dagger/2} L^q (D^q)^{\dagger/2} F \right)^{-1} F^T (D^q)^{\dagger/2} b_{uv}^q \\ &\leq 1 + 3H_\zeta^q(u, v) \leq 1 + \frac{3\psi_1 k^2}{n^2}. \end{aligned} \quad (46)$$

This implies (45) and finishes the proof the claim.  $\square$

For the first  $t$  iterations of the outermost while loop, let  $t_1$  be the total number of iterations executed by the first inner while loop, and  $t_2$  be the total number of iterations executed by the second inner while loop. Then we have  $t_1 \leq 2t$  by Lemma 10.7, and  $t_2 \leq t + t_1 \leq 3t$ . Therefore, by Claims 10.9, 10.10, 10.11, we have

$$\frac{\Upsilon(s^1)}{\Upsilon(s^0)} \leq \left(1 - \frac{\psi_2 k^2}{4n^2}\right)^t \left(1 - \frac{400\gamma_2^2 k^2}{(1-\zeta)^2 n^2}\right)^{-2t} \left(1 + \frac{3\psi_1 k^2}{n^2}\right)^{3t} \leq \left(1 - \frac{\psi_2 k^2}{16n^2}\right)^t,$$

where the last inequality follows from  $\psi_2 \geq 1024\psi_1$  and  $\psi_2 \geq \frac{1024\gamma_2^2}{(1-\zeta)^2}$ .  $\square$

Notice that by design, after WHACK-A-MOLE-II terminates, each edge  $(u, v)$  with  $s_{uv} < s_{uv}^0$  satisfies either  $R^s(u, v) \geq \frac{\gamma_1 k}{n}$  or  $H_\zeta^s(u, v) \geq \frac{\psi_1 k^2}{n^2}$ . We show that the total number of such edges is small.

**Lemma 10.12.** *After WHACK-A-MOLE-II terminates, the number of edges  $(u, v)$  with  $s_{uv} < s_{uv}^0$  such that  $R^s(u, v) \geq \frac{\gamma_1 k}{n}$  is at most  $\frac{n \cdot \text{rank}(D_s)}{\gamma_1 k}$ .*

*Proof.* Since  $R^s(u, v)$ 's are leverage scores, we have

$$\sum_{(u,v) \in E} R^s(u, v) = \text{rank}(D_s).$$

Then the desired bound follows.  $\square$

**Lemma 10.13.** *After WHACK-A-MOLE-II terminates, the number of edges  $(u, v)$  with  $s_{uv} < s_{uv}^0$  such that  $H_\zeta^s(u, v) \geq \frac{\psi_1 k^2}{n^2}$  is at most  $\frac{n^2 \gamma_2}{\psi_1 (1-\zeta)^2}$ .*

*Proof.* Let  $\ell$  be the number of eigenvalues of  $N^s$  between  $(0, \zeta]$ . Let  $\lambda_1, \dots, \lambda_\ell$  be all eigenvalues between  $(0, \zeta]$ , and let  $f_1, \dots, f_\ell$  be a set of orthonormal eigenvectors. By Theorem 8.1,  $\ell \leq \frac{\gamma_2 k^2}{(1-\zeta)^2}$ . Also we have by Proposition 3.12 that

$$\sum_{(u,v) \in E} H_\zeta^s(u, v) = \ell.$$

Thus our desired bound follows.  $\square$

## 10.2 Analysis of Algorithm 4

Our analysis of Algorithm 4 will mostly focus on bounding the total number of deleted edges at Line 14. Since we only delete edges with  $(s_{uv}/s'_{uv})^2 \leq \rho < 1$ , by the termination condition of the second inner while loop of WHACK-A-MOLE-II we know that each deleted edge satisfies either  $R^s(u, v) \geq \frac{\gamma_1 k}{n}$  or  $H_\zeta^s(u, v) \geq \frac{\psi_1 k^2}{n^2}$ , where  $\gamma_1 = \gamma/16$  and  $\psi_1 = \psi/1024$ . Thus we will show that the numbers of both types of edges are small.

Consider fixing a while loop iteration of EXPANDERDECOMP where we go the “else” branch. Let  $s'$  be the scaling we obtain after we invoke ALMOSTREGULARDECOMP at Line 7,  $s$  be the scaling returned by WHACK-A-MOLE-II, and  $\hat{s}$  be the scaling obtained after we delete the small weight edges at Line 14 (but before we reset the edge weights at Line 15). Then by an (almost) identical proof to that of Lemma 9.5, we have:

**Lemma 10.14.** *The number of edges  $(u, v) \in E$  with  $(s_{uv}/s'_{uv})^2 \leq \rho$  such that  $R^s(u, v) \geq \frac{\gamma_1 k}{n}$  is at most*

$$\frac{2n}{\gamma_1 k} \left( \text{rank}(D^{s'}) - \text{rank}(D^{\hat{s}}) \right).$$

This means that we can charge the number of deleted edges at Line 14 that satisfy  $R^s(u, v) \geq \frac{\gamma_1 k}{n}$  to the rank change of  $D$ . We will then bound the number of deleted edges that satisfy  $H_\zeta^s(u, v) \geq \frac{\psi_1 k^2}{n^2}$  by considering the number of nontrivial (Definition 7.11) zero eigenvalues of the normalized Laplacian, which, by Theorem 8.1, is at most  $2\gamma k^2$  when the graph is  $\gamma$ -almost regular.

To that end, for a normalized Laplacian matrix  $N$ , let  $\eta(N)$  denote the number of nontrivial zero eigenvalues of  $N$ . Additionally, let  $s''$  denote the scaling we obtain after we invoke ALMOSTREGULARDECOMP at Line 7 in the *next* iteration of the while loop. The following lemma characterizes how  $\eta(N)$  changes after a while loop iteration.

**Lemma 10.15.** *We have*

$$\eta(N^{s''}) \geq \eta(N^{s'}) + 1 - \left\lfloor \frac{2\gamma k}{n} \left( \text{rank}(D^{s'}) - \text{rank}(D^{s''}) \right) + \sqrt{\frac{1}{16nk}} \right\rfloor. \quad (47)$$

Before proving the lemma, we first give the proof of Theorem 10.1.

*Proof of Theorem 10.1.* Since we always maintain  $\gamma$ -almost regularity, the number of nontrivial zero eigenvalues can be at most  $2\gamma k^2$ . Therefore by Lemma 10.15 the total number of iterations of the while loop is at most  $6\gamma k^2$ . Thus, the algorithm terminates in finite time. Also, by the termination condition, the resulting graph  $G^s$  must be a  $(\gamma, \zeta, \psi)$ -almost regular expander.

By Lemma 10.13, the total number of deleted edges with  $H_\zeta^s(u, v) \geq \frac{\psi_1 k^2}{n^2}$  is at most

$$\frac{n^2 \gamma}{\psi_1 (1 - \zeta)^2} \cdot 6\gamma k^2 \leq \frac{10000 n^2 \gamma^2 k^2}{\psi (1 - \zeta)^2}.$$

By Lemma 10.14 and Theorem 9.1, the total number of deleted edges with  $R^s(u, v) \geq \frac{\gamma_1 k}{n}$ , plus those deleted by ALMOSTREGULARDECOMP, is at most

$$\frac{8n}{(\gamma/32)k} \cdot nk = \frac{256n^2}{\gamma}.$$

These two bounds coupled with Lemmas 10.12, 10.13 imply the desired bound on the number of rescaled edges, and thus finish the proof.  $\square$

It then remains to proof Lemma 10.15, for which we need the following lemma.

**Lemma 10.16.** *We have*

$$\lambda_{\max} \left( \left( D^{\dagger/2} L^{\hat{s}} D^{\dagger/2} \right)^{\dagger} \right) \leq \frac{1}{128nk^3\gamma \cdot \rho}.$$

*Proof.* Let  $\hat{s}_{\min} = \min \{ \hat{s}_{uv} : \hat{s}_{uv} \neq 0 \}$ . Since  $\hat{s}$  is obtained from  $s$  by zeroing out the edges  $(u, v)$  with  $(s_{uv}/s'_{uv})^2 \leq \rho$ , and no  $(s_{uv}/s'_{uv})^2$  is in the range  $(\rho, \rho \cdot \alpha)$ , we have

$$\hat{s}_{\min}^2 \geq \rho \cdot \alpha \cdot (s'_{\min})^2 \geq 128nk^3\gamma\alpha_2\rho,$$

where  $s'_{\min} = \min_{s'_{uv} > 0} s'_{uv}$  is defined at Line 8, and the last inequality follows from  $\alpha \geq 128nk^3\gamma \cdot \alpha_2/(s'_{\min})^2$ . Let  $F = \{(u, v) \in E : \hat{s}_{uv} > 0\}$ . Then we have

$$D^{\dagger/2} L^{\hat{s}} D^{\dagger/2} \succeq \hat{s}_{\min}^2 D^{\dagger/2} L^F D^{\dagger/2} \succeq 128nk^3\gamma\alpha_2\rho D^{\dagger/2} L^F D^{\dagger/2},$$

and therefore

$$\left( D^{\dagger/2} L^{\hat{s}} D^{\dagger/2} \right)^{\dagger} \preceq \frac{1}{128nk^3\gamma\alpha_2\rho} \left( D^{\dagger/2} L^F D^{\dagger/2} \right)^{\dagger}. \quad (48)$$

By definition,

$$\alpha_2 = \max_{F' \subseteq E} \lambda_{\max} \left( \left( D^{\dagger/2} L^{F'} D^{\dagger/2} \right)^{\dagger} \right) \geq \lambda_{\max} \left( \left( D^{\dagger/2} L^F D^{\dagger/2} \right)^{\dagger} \right).$$

This coupled with (48) implies that

$$\left( D^{\dagger/2} L^{\hat{s}} D^{\dagger/2} \right)^{\dagger} \preceq \frac{1}{128nk^3\gamma\rho} I$$

as desired.  $\square$

*Proof of Lemma 10.15.* Since  $s$  and  $s'$  have the same support, we have  $\eta(N^{s'}) = \eta(N^s)$ . Let  $z = \eta(N^s)$ . Let  $0 = \lambda_1 = \dots = \lambda_z < \lambda_{z+1} \leq \rho$  be the smallest  $z+1$  nontrivial eigenvalues of  $N^s$ , and let  $f_1, \dots, f_{z+1}$  be a corresponding set of orthonormal, nontrivial eigenvectors. Since  $G^{\hat{s}}$  is a subgraph of  $G^s$ ,  $f_1, \dots, f_z$  are also zero eigenvectors of  $(D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2}$ . We first show that  $f_{z+1}$  is not in the range of  $(D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2}$ . In particular, we will show that the projection of  $f_{z+1}$  onto the range of  $(D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2}$ , i.e.

$$\left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{\dagger/2} \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{1/2} f_{z+1},$$

has small norm. First note that

$$\begin{aligned} \left\| \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{1/2} f_{z+1} \right\|^2 &= f_{z+1}^T (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} f_{z+1} \\ &\leq f_{z+1}^T (D^s)^{\dagger/2} L^s (D^s)^{\dagger/2} f_{z+1} \\ &= \lambda_{z+1} \leq \rho. \end{aligned} \quad (49)$$

Then

$$\begin{aligned}
& \left\| \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{\dagger/2} \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{1/2} f_{z+1} \right\|^2 \\
&= f_{z+1}^T \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{1/2} \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{\dagger} \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{1/2} f_{z+1} \\
&\leq \lambda_{\max} \left( \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{\dagger} \right) \left\| \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{1/2} f_{z+1} \right\|^2 \\
&\leq \lambda_{\max} \left( \left( D^{\dagger/2} L^{\hat{s}} D^{\dagger/2} \right)^{\dagger} \right) \left\| \left( (D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2} \right)^{1/2} f_{z+1} \right\|^2 \\
&\leq \frac{1}{128nk^3\gamma\rho} \cdot \rho = \frac{1}{128nk^3\gamma}, \tag{50}
\end{aligned}$$

where the last inequality follows from Lemma 10.16 and (49).

Define a matrix  $\mathcal{F} \in \mathbb{R}^{(z+1) \times nk}$  by

$$\mathcal{F} := \begin{pmatrix} f_1^T \\ \vdots \\ f_{z+1}^T \end{pmatrix} \in \mathbb{R}^{(z+1) \times nk}$$

and a function  $F : V \rightarrow \mathbb{R}^{(z+1) \times k}$  by

$$F(u) = \begin{pmatrix} (f_1)_u^T \\ \vdots \\ (f_{z+1})_u^T \end{pmatrix} \in \mathbb{R}^{(z+1) \times nk}$$

Since  $G^s$  is  $\gamma$ -almost regular, we have by Lemma 8.2 that  $\lambda_{\max}(F(u)F(u)^T) \leq \frac{2\gamma k}{n}$  and by Theorem 8.1 that  $z+1 \leq 2\gamma k^2$ .

Now consider projecting the rows of  $\mathcal{F}$  twice: (i) project each row onto the null space of  $(D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2}$ , and then (ii) project each row onto the range of  $(D^s)^{\dagger/2} D^{s''} (D^s)^{\dagger/2}$ . Specifically, let  $\Pi_1, \Pi_2 \in \mathbb{R}^{nk \times nk}$  be the projection matrix onto the range of  $(D^s)^{\dagger/2} L^{\hat{s}} (D^s)^{\dagger/2}$  and  $(D^s)^{\dagger/2} D^{s''} (D^s)^{\dagger/2}$  respectively, and let  $\Pi_1^\perp, \Pi_2^\perp$  be the projection matrix onto the maximal subspace orthogonal to the range of  $\Pi_1$  and  $\Pi_2$  respectively. Then we consider the matrix  $\mathcal{F}' := \mathcal{F}\Pi_1^\perp\Pi_2$ .

In order to prove the lemma, it suffices to show that the rank of  $\mathcal{F}'$  is at least

$$z+1 - \left\lfloor \frac{2\gamma k}{n} \left( \text{rank}(D^{s'}) - \text{rank}(D^{s''}) \right) + \sqrt{\frac{1}{16nk}} \right\rfloor, \tag{51}$$

because this will imply that the number of zero eigenvalues of  $(D^s)^{\dagger/2} L^{s''} (D^s)^{\dagger/2}$  (and hence of  $N^{s''}$ ) is at least  $nk - \text{rank}(D^{s''}) + (51)$ .

Since

$$\mathcal{F}'(\mathcal{F}')^T = \mathcal{F}\Pi_1^\perp\Pi_2\Pi_1^\perp\mathcal{F}^T \preceq \mathcal{F}\Pi_1^\perp\Pi_1^\perp\mathcal{F}^T \preceq \mathcal{F}\mathcal{F}^T = I_{(z+1) \times (z+1)},$$

we know that all singular values of  $\mathcal{F}'$  is at most 1. Thus, it suffices to show that

$$\mathrm{Tr}(\mathcal{F}'(\mathcal{F}')^T) \geq z + 1 - \left( \frac{2\gamma k}{n} (\mathrm{rank}(D^{s'}) - \mathrm{rank}(D^{s''})) + \sqrt{\frac{1}{16nk}} \right),$$

which will imply that the number of nonzero singular values of  $\mathcal{F}'$  is at least (51). To this end, let us first write  $\mathrm{Tr}(\mathcal{F}\Pi_2\mathcal{F}^T)$  as

$$\begin{aligned} & \mathrm{Tr}(\mathcal{F}\Pi_2\mathcal{F}^T) \\ &= \mathrm{Tr}(\mathcal{F}(\Pi_1^\perp + \Pi_1)\Pi_2(\Pi_1^\perp + \Pi_1)\mathcal{F}^T) \\ &= \mathrm{Tr}(\mathcal{F}\Pi_1^\perp\Pi_2\Pi_1^\perp\mathcal{F}^T) + \mathrm{Tr}(\mathcal{F}\Pi_1\Pi_2\mathcal{F}^T) + \mathrm{Tr}(\mathcal{F}\Pi_1^\perp\Pi_2\Pi_1\mathcal{F}^T) \\ &\leq \mathrm{Tr}(\mathcal{F}\Pi_1^\perp\Pi_2\Pi_1^\perp\mathcal{F}^T) \\ &\quad + \sqrt{\mathrm{Tr}(\mathcal{F}\Pi_1\mathcal{F}^T)\mathrm{Tr}(\mathcal{F}\Pi_2\mathcal{F}^T)} + \sqrt{\mathrm{Tr}(\mathcal{F}\Pi_1^\perp\Pi_2\Pi_1^\perp\mathcal{F}^T)\mathrm{Tr}(\mathcal{F}\Pi_1\mathcal{F}^T)}, \end{aligned} \quad (52)$$

where the last equality follows from expanding, and the last inequality follows from Cauchy-Schwarz. This combined with (50) gives that

$$\begin{aligned} \mathrm{Tr}(\mathcal{F}'(\mathcal{F}')^T) &= \mathrm{Tr}(\mathcal{F}\Pi_1^\perp\Pi_2\Pi_1^\perp\mathcal{F}^T) \\ &\geq \mathrm{Tr}(\mathcal{F}\Pi_2\mathcal{F}^T) - 2\sqrt{\frac{z+1}{128nk^3\gamma}} \\ &\geq \mathrm{Tr}(\mathcal{F}\Pi_2\mathcal{F}^T) - 2\sqrt{\frac{2\gamma k^2}{128nk^3\gamma}} \\ &= \mathrm{Tr}(\mathcal{F}\Pi_2\mathcal{F}^T) - \sqrt{\frac{1}{16nk}}. \end{aligned} \quad (53)$$

Using the fact that  $\lambda_{\max}(F(u)F(u)^T) \leq \frac{2\gamma k}{n}$ , we have

$$\mathrm{Tr}(\mathcal{F}\Pi_2^\perp\mathcal{F}^T) \leq \frac{2\gamma k}{n} (\mathrm{rank}(D^{s'}) - \mathrm{rank}(D^{s''})),$$

and hence

$$\begin{aligned} \mathrm{Tr}(\mathcal{F}\Pi_2\mathcal{F}^T) &= \mathrm{Tr}(\mathcal{F}\mathcal{F}^T) - \mathrm{Tr}(\mathcal{F}\Pi_2^\perp\mathcal{F}^T) \\ &\geq z + 1 - \frac{2\gamma k}{n} (\mathrm{rank}(D^{s'}) - \mathrm{rank}(D^{s''})). \end{aligned} \quad (54)$$

Combining (53) and (54) finishes the proof of the lemma.  $\square$

## 11 Expanders are preserved under vertex sampling

### 11.1 Warm-up: ordinary expanders are preserved under vertex sampling

As a warm-up, we start with ordinary unweighted graphs, and prove that an expander with large minimum degree still has good expansion after uniform vertex sampling. We will use the algebraic definition of expansion, which could be translated to combinatorial expansion using Cheeger's inequality [AM85].

**Definition 11.1** (Expanders). An unweighted graph  $G = (V, E)$  is a  $\zeta$ -expander for some  $0 < \zeta \leq 1$  if the second smallest eigenvalue of the normalized Laplacian  $N$  is  $\lambda_2 \geq \zeta$ .

Note that any ordinary, unweighted graph  $G = (V, E)$  with minimum degree  $d_{\min}$  can be seen as a  $1 \times 1$ -matrix weighted,  $\frac{n}{d_{\min}}$ -almost regular graph. Therefore we can apply Lemma 8.2 and obtain similar properties of the spectral embedding induced by the bottom eigenvectors of the normalized Laplacian of  $G$ .

**Definition 11.2** (Spectral embeddings of scalar-weighted graphs). Given orthonormal vectors  $f_1, \dots, f_\ell \in \mathbb{R}^n$ , define an  $\ell \times n$  matrix  $\mathcal{F}$  whose rows are transposes of  $f_1, \dots, f_\ell$ :

$$\mathcal{F} = \begin{array}{|c|c|c|c|c|} \hline (f_1)_1 & \cdots & (f_1)_u & \cdots & (f_1)_n \\ \hline (f_2)_1 & \cdots & (f_2)_u & \cdots & (f_2)_n \\ \hline & & \vdots & & \\ \hline (f_\ell)_1 & \cdots & (f_\ell)_u & \cdots & (f_\ell)_n \\ \hline \end{array} \in \mathbb{R}^{\ell \times n}.$$

Then define an embedding  $F : V \rightarrow \mathbb{R}^\ell$  by letting  $F(u)$  equal the  $u^{\text{th}}$  column of  $\mathcal{F}$ :

$$F(u) = \begin{array}{|c|} \hline (f_1)_u \\ \hline (f_2)_u \\ \hline \vdots \\ \hline (f_\ell)_u \\ \hline \end{array} \in \mathbb{R}^\ell.$$

We call  $F$  the spectral embedding induced by  $f_1, \dots, f_\ell$ , and  $\mathcal{F}$  the embedding matrix induced by  $f_1, \dots, f_\ell$ .

By Lemma 8.2, we immediately have:

**Corollary 11.1** (of Lemma 8.2). *Let  $G = (V, E)$  be an ordinary, unweighted graph with minimum degree  $d_{\min}$ . Fix a  $\delta \in (0, 1)$  and let  $0 \leq \lambda_1 \leq \dots \leq \lambda_\ell \leq 1 - \delta$  be all eigenvalues of  $N = D^{-1/2}LD^{-1/2}$  that are  $\leq 1 - \delta$ . Let  $f_1, \dots, f_\ell$  be a corresponding set of orthonormal eigenvectors. Let  $F$  be the spectral embedding induced by  $f_1, \dots, f_\ell$ . Then we have for all  $u \in V$*

$$\|F(u)\|^2 \leq \frac{1}{\delta^2 d_{\min}}. \quad (55)$$

By Theorem 8.1 we also have a bound on the number of small eigenvalues:

**Corollary 11.2** (of Theorem 8.1). *Let  $G = (V, E)$  be an ordinary, unweighted graph with minimum degree  $d_{\min}$ . Then for any  $\delta \in (0, 1)$ , the number of eigenvalues of  $N = D^{-1/2}LD^{-1/2}$  that are at most  $1 - \delta$  is at most  $\frac{n}{\delta^2 d_{\min}}$ .*

**Theorem 11.3.** Let  $G = (V, E)$  be a  $\zeta$ -expander with minimum degree  $d_{\min} \geq 4 \cdot 10^6 \cdot \zeta^{-1} n^{\frac{32}{\log \log n}}$  for some  $\zeta \leq \frac{1}{\log n}$ . For an  $s \geq \frac{4 \cdot 10^6 \cdot \zeta^{-1} n^{\frac{32}{\log \log n}}}{d_{\min}} \cdot n$ , let  $C$  be a uniformly random vertex subset of size  $s$ . Then with probability  $1 - n^{-7}$ , the induced subgraph  $G[C]$  is a  $\zeta/n^{\frac{1024}{\log \log n}}$ -expander with minimum degree at least  $\frac{s}{2n} \cdot d_{\min}$ .

This theorem is consequence of applying the following lemma  $O(\frac{\log n}{\log \log n})$  times.

**Lemma 11.4.** Let  $G = (V, E)$  be a  $\zeta$ -expander with minimum degree  $d_{\min} \geq 2 \cdot 10^6 \cdot \zeta^{-1} \log^{10} n$  for some  $\zeta \leq \frac{1}{\log n}$ . For an  $s \geq \frac{n}{\log n}$ , let  $C$  be a uniformly random vertex subset of size  $s$ . Then with probability  $1 - n^{-8}$ , the induced subgraph  $G[C]$  is a  $\zeta/16$ -expander with minimum degree at least  $(1 - \frac{1}{2 \log n}) \cdot \frac{s}{n} \cdot d_{\min}$ .

We will prove Lemma 11.4 via a matrix Martingale argument, for which end we will need to set up the notions of Cholesky factorization and Schur complements. The following definitions and facts are from [KS16, DKP<sup>+</sup>17].

**Schur complements.** Let  $L \in \mathbb{R}^{n \times n}$  be the Laplacian matrix of an  $n$ -vertex, connected graph  $G$ , and let  $L(:, u)$  denote the column of  $L$  corresponding to vertex  $u$ . For a vertex  $u_1$ , we define the *Schur complement* of  $L$  with respect to  $u_1$  as

$$S^{(1)} = L - \frac{1}{L_{u_1 u_1}} L(:, u_1) L(:, u_1)^T \in \mathbb{R}^{n \times n}. \quad (56)$$

It is straightforward to see:

**Fact 11.5.** The entries on row  $u_1$  of  $S^{(1)}$  and the entries on column  $u_1$  of  $S^{(1)}$  are all zero.

The following fact is less straightforward, but well known:

**Fact 11.6.**  $S^{(1)}$  is a Laplacian matrix of a graph supported on  $V \setminus \{u_1\}$ .

We call the operation of subtracting  $\frac{1}{L_{u_1 u_1}} L(:, u_1) L(:, u_1)^T$  from  $L$  the *elimination of vertex  $u_1$* . Suppose we perform a sequence of eliminations of vertices  $u_1, u_2, \dots, u_t$  for some  $1 \leq t < n$ , and define for each  $i \in [t]$

$$\begin{aligned} \alpha_i &= S_{u_i u_i}^{(i-1)} \in \mathbb{R} \\ c_i &= S^{(i-1)}(:, u_i) \in \mathbb{R}^n \\ S^{(i)} &= S^{(i-1)} - \frac{1}{\alpha_i} c_i c_i^T \in \mathbb{R}^{n \times n} \end{aligned}$$

where  $S^{(0)} := L$ . We call  $S^{(t)}$  the *Schur complement* of  $L$  with respect to  $u_1, \dots, u_t$ . Then similar to Facts 11.5 and 11.6, we have:

**Fact 11.7.** The entries on rows  $u_1, \dots, u_t$  of  $S^{(t)}$  and the entries on columns  $u_1, \dots, u_t$  of  $S^{(t)}$  are all zero. Moreover,  $S^{(t)}$  is a Laplacian matrix of a graph supported on  $V \setminus \{u_1, \dots, u_t\}$ .

It is also known that:

**Fact 11.8.** Changing the order in which we eliminate  $u_1, \dots, u_t$  does not change the resulting Schur complement  $S^{(t)}$ .

Let  $C = V \setminus \{u_1, \dots, u_t\}$ . We define

$$\text{SC}(L, C) := S_{CC}^{(t)} \in \mathbb{R}^{|C| \times |C|} \quad (57)$$

as the *Schur complement of  $L$  onto  $C$* , where  $S_{CC}^{(t)}$  is  $S^{(t)}$  restricted to rows and columns in  $C$ . Note that by Fact 11.7,  $\text{SC}(L, C)$  contains all nonzero entries of  $S^{(t)}$ . We also write  $\text{SC}(G, C) = \text{SC}(L, C)$ . To connect Schur complements to the Laplacian  $L$ , we need to introduce partial Cholesky factorization.

**Partial Cholesky factorization.** Suppose we eliminate a sequence of vertices  $u_1, \dots, u_t$  as above. Let  $\mathcal{L}$  be the  $n \times t$  matrix whose  $i^{\text{th}}$  column is  $c_i$ , and let  $\mathcal{D}$  be the  $t \times t$  diagonal matrix with  $\mathcal{D}_{ii} = \alpha_i$ . Then by adding the matrices subtracted in the elimination steps back to  $S^{(t)}$ , we have

$$L = S^{(t)} + \sum_{i=1}^t \alpha_i c_i c_i^T = S^{(t)} + \mathcal{L} \mathcal{D} \mathcal{L}^T. \quad (58)$$

Let us define  $F = \{u_1, \dots, u_t\}$  and  $C = V \setminus F$ . Let  $S = \text{SC}(L, C)$ , where we recall that the latter is  $S^{(t)}$  restricted to rows and columns in  $C$ . Since  $S$  contains all nonzero entries of  $S^{(t)}$ , by writing  $\mathcal{L}$  in block form as  $\mathcal{L} = \begin{pmatrix} \mathcal{L}_{FF} \\ \mathcal{L}_{CF} \end{pmatrix}$ , we get

$$L = \begin{pmatrix} \mathcal{L}_{FF} \\ \mathcal{L}_{CF} \end{pmatrix} \mathcal{D} \begin{pmatrix} \mathcal{L}_{FF} \end{pmatrix}^T + \begin{pmatrix} 0_{FF} & 0_{FC} \\ 0_{CF} & S \end{pmatrix} = \begin{pmatrix} \mathcal{L}_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix} \begin{pmatrix} \mathcal{D} & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \mathcal{L}_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix}^T. \quad (59)$$

We call (59) a *partial Cholesky factorization* of  $L$ .

We now state a fact about matrix factorizations of the form in (59). For a set of vectors  $x_1, \dots, x_s$ , we will write  $\Pi_{x_1, \dots, x_s}$  to denote the projection matrix onto their linear span, and  $\Pi_{x_1, \dots, x_s}^\perp$  to denote the projection matrix onto the maximal subspace orthogonal to their linear span.

**Fact 11.9.** Given an  $n \times n$  positive semidefinite matrix  $M$  that can be factorized as

$$M = \begin{pmatrix} \mathcal{L}_{FF} \\ \mathcal{L}_{CF} \end{pmatrix} \mathcal{D} \begin{pmatrix} \mathcal{L}_{FF} \end{pmatrix}^T + \begin{pmatrix} 0_{FF} & 0_{FC} \\ 0_{CF} & S \end{pmatrix} = \begin{pmatrix} \mathcal{L}_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix} \begin{pmatrix} \mathcal{D} & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \mathcal{L}_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix}^T$$

for some bi-partition  $(C, F)$  of  $[n]$ , such that (i)  $\mathcal{L}_{FF} \in \mathbb{R}^{|F| \times |F|}$  is a full-rank matrix; (ii)  $\mathcal{L}_{CF} \in \mathbb{R}^{|C| \times |F|}$ ; (iii)  $\mathcal{D}$  is a diagonal matrix of non-negative entries. Let  $f_1, \dots, f_z$  be a basis of the null space of  $M$ . Then

$$S^\dagger = \Pi_{(f_1)_C, \dots, (f_z)_C}^\perp (M^\dagger)_{CC} \Pi_{(f_1)_C, \dots, (f_z)_C}^\perp,$$

where  $(M^\dagger)_{CC}$  is  $M^\dagger$  restricted to rows and columns in  $C$ .

As a direct corollary of the above fact, we have:

**Fact 11.10** (Corollary of Fact 11.9). Let  $f_1 \in \mathbb{R}^n$  be the all-one vector. For any  $C \subseteq V$ ,

$$\text{SC}(L, C) = \Pi_{(f_1)_C}^\perp (L^\dagger)_{CC} \Pi_{(f_1)_C}^\perp.$$

Below we will need to use the following lemma, which is a direct consequence of the Matrix Chernoff bound (Theorem 4.2).

**Lemma 11.11.** Let  $f_1, \dots, f_\ell \in \mathbb{R}^n$  be any  $\ell$  orthonormal vectors. Define  $F : [n] \rightarrow \mathbb{R}^\ell$  by letting  $F(i) = ((f_1)_i, \dots, (f_\ell)_i)^T \in \mathbb{R}^\ell$ . Suppose for any  $i \in [n]$ ,

$$\|F(i)\|^2 \leq \rho$$

for some  $\rho$ . For an  $s \geq 100\rho n \log n$ , let  $C \subseteq [n]$  be a uniformly random subset of indices of size  $s$ . Then we have with probability  $1 - n^{-10}$  that

$$\frac{1}{2} I_{\ell \times \ell} \preceq \frac{n}{s} \left( \sum_{i \in C} F(i) F(i)^T \right) \preceq 2I_{\ell \times \ell}$$

and therefore

$$\sum_{j=1}^{\ell} (f_j)_C (f_j)_C^T \preceq \frac{2s}{n} I_{n \times n}.$$

The following lemma shows that the Schur complement onto a random vertex subset is a better expander than the original graph. Here we use  $\Pi_x^\perp$  to denote the projection matrix onto the maximal subspace orthogonal to vector  $x$ .

**Lemma 11.12.** Let  $G = (V, E)$  be a  $\zeta$ -expander with minimum degree  $d_{\min} \geq 2 \cdot 10^6 \log^5 n$  where  $\zeta \leq \frac{1}{\log n}$ . Let the eigenvalues of  $N_G$  be  $0 = \lambda_1 < \zeta \leq \lambda_2 \leq \dots \leq \lambda_n$  and let  $f_1, \dots, f_n$  be a set of corresponding orthonormal eigenvectors. For an  $s \geq \frac{n}{\log n}$ , let  $C$  be a uniformly random vertex subset of size  $s$ . Then with probability  $1 - n^{-10}$

$$D_{CC}^{-1/2} \text{SC}(L_G, C) D_{CC}^{-1/2} \succeq \frac{n}{4s} \cdot \zeta \cdot \Pi_{(f_1)_C}^\perp.$$

*Proof.* By eliminating vertices outside of  $C$ , we get a partial Cholesky factorization of  $L_G$ :

$$L_G = \begin{pmatrix} \mathcal{L}_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix} \begin{pmatrix} \mathcal{D} & 0 \\ 0 & \text{SC}(L_G, C) \end{pmatrix} \begin{pmatrix} \mathcal{L}_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix}^T.$$

By multiplying  $D^{-1/2}$  on both sides, we then get a factorization of  $N_G$ :

$$\begin{aligned} N_G &= \begin{pmatrix} D_{FF}^{-1/2} \mathcal{L}_{FF} & 0 \\ D_{CC}^{-1/2} \mathcal{L}_{CF} & D_{CC}^{-1/2} \end{pmatrix} \begin{pmatrix} \mathcal{D} & 0 \\ 0 & \text{SC}(L_G, C) \end{pmatrix} \begin{pmatrix} D_{FF}^{-1/2} \mathcal{L}_{FF} & 0 \\ D_{CC}^{-1/2} \mathcal{L}_{CF} & D_{CC}^{-1/2} \end{pmatrix}^T \\ &= \begin{pmatrix} D_{FF}^{-1/2} \mathcal{L}_{FF} & 0 \\ D_{CC}^{-1/2} \mathcal{L}_{CF} & I_{CC} \end{pmatrix} \begin{pmatrix} \mathcal{D} & 0 \\ 0 & D_{CC}^{-1/2} \text{SC}(L_G, C) D_{CC}^{-1/2} \end{pmatrix} \begin{pmatrix} D_{FF}^{-1/2} \mathcal{L}_{FF} & 0 \\ D_{CC}^{-1/2} \mathcal{L}_{CF} & I_{CC} \end{pmatrix}^T. \end{aligned}$$

Then by Fact 11.9,

$$\left(D_{CC}^{-1/2} \text{SC}(L_G, C) D_{CC}^{-1/2}\right)^\dagger = \Pi_{(f_1)_C}^\perp \left(\sum_{i=2}^n \frac{1}{\lambda_i} (f_i)_C (f_i)_C^T\right) \Pi_{(f_1)_C}^\perp.$$

Let  $\ell$  be such that  $\lambda_\ell \leq \max\{\frac{n}{2s}, 1\} \cdot \zeta < \lambda_{\ell+1}$ . Define the embedding  $F : V \rightarrow \mathbb{R}^\ell$  as

$$F(u) = \begin{pmatrix} (f_1)_u \\ (f_2)_u \\ \vdots \\ (f_\ell)_u \end{pmatrix} \in \mathbb{R}^\ell.$$

By  $s \geq n/\log n$  and  $\zeta \leq 1/\log n$ , we have  $\frac{n}{2s} \cdot \zeta \leq 1/2$ , and thus by Corollary 11.1 we have that for each  $u \in V$

$$\|F(u)\|^2 \leq \frac{4}{d_{\min}} \leq 10^{-4} \cdot \log^{-3} n.$$

Then we have

$$\begin{aligned} \left(D_{CC}^{-1/2} \text{SC}(L_G, C) D_{CC}^{-1/2}\right)^\dagger &= \Pi_{(f_1)_C}^\perp \left(\sum_{i=2}^n \frac{1}{\lambda_i} (f_i)_C (f_i)_C^T\right) \Pi_{(f_1)_C}^\perp \\ &\preceq \Pi_{(f_1)_C}^\perp \left(\sum_{i=2}^{\ell} \frac{1}{\zeta} (f_i)_C (f_i)_C^T + \sum_{i=\ell+1}^n \frac{2s}{n} \cdot \frac{1}{\zeta} (f_i)_C (f_i)_C^T\right) \Pi_{(f_1)_C}^\perp \\ &\preceq \Pi_{(f_1)_C}^\perp \left(\sum_{i=2}^{\ell} \frac{1}{\zeta} (f_i)_C (f_i)_C^T + \frac{2s}{n} \cdot \frac{1}{\zeta} I\right) \Pi_{(f_1)_C}^\perp \\ &\preceq \Pi_{(f_1)_C}^\perp \left(\frac{2s}{n} \cdot \frac{1}{\zeta} I + \frac{2s}{n} \cdot \frac{1}{\zeta} I\right) \Pi_{(f_1)_C}^\perp \\ &= \frac{4s}{n} \cdot \frac{1}{\zeta} \Pi_{(f_1)_C}^\perp, \end{aligned}$$

where the third line follows from  $\lambda_{\max}(\sum_{i=\ell+1}^n (f_i)_C (f_i)_C^T) \leq \lambda_{\max}(\sum_{i=1}^n f_i f_i^T) = 1$ , and the second to last line holds with probability  $1 - n^{-10}$  by Lemma 11.11. By inverting both sides, we then get

$$D_{CC}^{-1/2} \text{SC}(L_G, C) D_{CC}^{-1/2} \succeq \frac{n}{4s} \cdot \zeta \cdot \Pi_{(f_1)_C}^\perp,$$

as desired.  $\square$

We now define the notion of graph squaring, which will be useful in setting up our martingale.

**Definition 11.3** (Graph squaring). For a graph  $G = (V, E)$  with degree matrix  $D$  and adjacency matrix  $A$ , define the square of  $G$  by

$$L_{G^2} \stackrel{\text{def}}{=} D - AD^{-1}A.$$

**Fact 11.13** ([CCL<sup>+</sup>15, JKPS17]).  $L_{G^2} \preceq 2L_G$  for any graph  $G$ .

Below, we will write  $\text{SC}_n(L, C)$  or  $\text{SC}_n(G, C)$  to denote an  $n \times n$  matrix obtained by augmenting  $\text{SC}(L, C)$  to  $n \times n$  by adding zeros on rows and columns in  $V \setminus C$ . We will also abuse the notation a bit and write  $L_{G[C]}$  to denote an  $n \times n$  matrix obtained by augmenting  $L_{G[C]}$  to  $n \times n$  by adding zeros on rows and columns in  $V \setminus C$ .

**Proposition 11.14.**  $\sum_{u \in V} \text{SC}_n(L_G, V \setminus \{u\}) = L_{G^2} + (n - 2)L_G$  for any graph  $G$ .

*Proof.* Notice that by definition

$$\begin{aligned} \sum_{u \in V} \text{SC}_n(L, V \setminus \{u\}) &= \sum_{u \in V} \left( L_G - \frac{1}{L_{uu}} L_G(:, u) L_G(:, u)^T \right) \\ &= nL_G - L_G D^{-1} L_G \\ &= n(D - A) - (D - A) D^{-1} (D - A) \\ &= nD - nA - (D - 2A + AD^{-1}A) \\ &= (n - 2)(D - A) + D - AD^{-1}A = (n - 2)L_G + L_{G^2} \end{aligned}$$

as desired.  $\square$

Let  $v_1, \dots, v_n$  be a uniformly random permutation of vertices in  $V$ . Define

$$V_i = \{v_1, \dots, v_i\}.$$

Consider the following sequence of matrices:

$$\begin{aligned} X_0 &= L_G \\ X_1 &= \frac{1}{d_{v_1}} L_G(:, v_1) L_G(:, v_1)^T + \left(1 + \frac{2}{n-2}\right) L_{G[V-V_1]} + \frac{1}{n} (L_{G^2} - 2L_G) \\ X_2 &= \frac{1}{d_{v_1}} L_G(:, v_1) L_G(:, v_1)^T + \frac{1}{n} (L_{G^2} - 2L_G) + \left(1 + \frac{2}{n-2}\right) \cdot \\ &\quad \left( \frac{1}{d_{v_2}^{G[V-V_1]}} L_{G[V-V_1]}(:, v_2) L_{G[V-V_1]}(:, v_2)^T + \left(1 + \frac{2}{n-3}\right) L_{G[V-V_2]} + \frac{1}{n-1} (L_{G[V-V_1]^2} - 2L_{G[V-V_1]}) \right) \end{aligned}$$

$X_{i+1}$  is obtained by replacing the  $L_{G[V-V_i]}$  in  $X_i$  by

$$\frac{1}{d_{v_{i+1}}^{G[V-V_i]}} L_{G[V-V_i]}(:, v_{i+1}) L_{G[V-V_i]}(:, v_{i+1})^T + \left(1 + \frac{2}{n-2-i}\right) L_{G[V-V_{i+1}]} + \frac{1}{n-i} (L_{G[V-V_i]^2} - 2L_{G[V-V_i]}). \quad (60)$$

Here, we have used  $d_v^H$  to denote the degree of vertex  $v$  in graph  $H$ .

**Lemma 11.15.**  $X_0, X_1, X_2, \dots$  is a matrix-valued martingale.

*Proof.* It suffices to prove that

$$\mathbb{E}_{v_{i+1}} [(\textcolor{red}{60}) \mid V_i] = L_{G[V-V_i]}.$$

Let us calculate the LHS term by term.

$$\begin{aligned} & \mathbb{E}_{v_{i+1}} \left[ \frac{1}{d_{v_{i+1}}^{G[V-V_i]}} L_{G[V-V_i]}(:, v_{i+1}) L_{G[V-V_i]}(:, v_{i+1})^T \mid V_i \right] \\ &= \frac{1}{n-i} L_{G[V-V_i]} D_{G[V-V_i]}^{-1} L_{G[V-V_i]} \\ &= \frac{1}{n-i} \left( D_{G[V-V_i]} - 2A_{G[V-V_i]} + A_{G[V-V_i]} D_{G[V-V_i]}^{-1} A_{G[V-V_i]} \right) \\ \\ & \mathbb{E}_{v_{i+1}} \left[ \left( 1 + \frac{2}{n-2-i} \right) L_{G[V-V_{i+1}]} \mid V_i \right] = \frac{n-2-i}{n-i} L_{G[V-V_i]} + \frac{2}{n-i} L_{G[V-V_i]} = L_{G[V-V_i]} \\ \\ & \mathbb{E}_{v_{i+1}} \left[ \frac{1}{n-i} (L_{G[V-V_i]^2} - 2L_{G[V-V_i]}) \mid V_i \right] \\ &= \frac{1}{n-i} (L_{G[V-V_i]^2} - 2L_{G[V-V_i]}) \\ &= \frac{1}{n-i} \left( D_{G[V-V_i]} - A_{G[V-V_i]} D_{G[V-V_i]}^{-1} A_{G[V-V_i]} \right) - \frac{2}{n-i} L_{G[V-V_i]}. \end{aligned}$$

One can verify that these three add up to  $L_{G[V-V_i]}$ . □

To analyze this matrix-valued martingale, we will resort to the following theorem:

**Theorem 11.16** (Matrix Freedman Inequality [[Tro11](#)]). *Consider a zero-mean matrix martingale  $\{Y_i : i = 0, 1, \dots\}$  whose values are symmetric matrices with dimensional  $n$ , and let  $\{Z_i : i = 0, 1, \dots\}$  be the difference sequence such that  $Z_i = Y_{i+1} - Y_i$ . Assume the difference sequence is uniformly bounded in the sense that for any  $i = 0, 1, \dots$*

$$\lambda_{\max}(Z_i) \leq R,$$

where  $\lambda_{\max}(Z_i)$  is the maximum absolute value of any eigenvalue of  $Z_i$ . Define the predictable quadratic variation process of the martingale as

$$W_i \stackrel{\text{def}}{=} \sum_{j=0}^i \mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_j].$$

Then for all  $i \geq 0$  and  $\sigma^2 > 0$ ,

$$\Pr [\exists i \geq 0 : \lambda_{\max}(Y_i) \geq t \quad \text{and} \quad \|W_i\| \leq \sigma^2] \leq d \cdot \exp \left\{ -\frac{t^2/2}{\sigma^2 + Rt/3} \right\}.$$

We now prove Lemma [11.4](#).

**Lemma 11.4.** Let  $G = (V, E)$  be a  $\zeta$ -expander with minimum degree  $d_{\min} \geq 2 \cdot 10^6 \cdot \zeta^{-1} \log^{10} n$  for some  $\zeta \leq \frac{1}{\log n}$ . For an  $s \geq \frac{n}{\log n}$ , let  $C$  be a uniformly random vertex subset of size  $s$ . Then with probability  $1 - n^{-8}$ , the induced subgraph  $G[C]$  is a  $\zeta/16$ -expander with minimum degree at least  $(1 - \frac{1}{2 \log n}) \cdot \frac{s}{n} \cdot d_{\min}$ .

*Proof.* The minimum degree guarantee follows from a direct application of Chernoff bounds. We then focus on showing that  $G[C]$  is a  $\zeta/4$ -expander.

Define  $Y_i = L_G^{\dagger/2} X_i L_G^{\dagger/2} - L_G^{\dagger/2} L_G L_G^{\dagger/2}$ . Since  $X_0, X_1, \dots$  is a martingale, so is  $Y_0, Y_1, \dots$ , by linearity of expectation. Moreover, since  $X_0 = L_G$ ,  $Y_0, Y_1, \dots$  has zero mean. Consider the first  $n - s + 1$  terms  $Y_0, Y_1, Y_2, \dots, Y_{n-s}$ . First let us calculate the difference sequence  $Z_i$ :

$$\begin{aligned} Z_i &= Y_{i+1} - Y_i \\ &= \prod_{j=1}^i \left( 1 + \frac{2}{n-1-j} \right) \cdot L_G^{\dagger/2} \\ &\quad \left( -\text{SC}_n(G[V - V_i], V - V_{i+1}) + \left( 1 + \frac{2}{n-2-i} \right) L_{G[V-V_{i+1}]} + \frac{1}{n-i} (L_{G[V-V_i]^2} - 2L_{G[V-V_i]}) \right) L_G^{\dagger/2}, \end{aligned}$$

where we have used Definitions (56), (57), and recall that  $\text{SC}_n(G, C)$  is obtained by augmenting  $\text{SC}(G, C)$  to  $n \times n$  by adding zeros. We then consider to bound the maximum (in absolute value) eigenvalue of  $Z_i$ . For the last term, we have

$$\lambda_{\max} \left( L_G^{\dagger/2} \frac{1}{n-i} (L_{G[V-V_i]^2} - 2L_{G[V-V_i]}) L_G^{\dagger/2} \right) \leq \frac{4}{n-i} \leq \frac{4}{s} \leq \frac{4 \log n}{d_{\min}} \leq \frac{4}{10^6 \log^4 n}, \quad (61)$$

where the first inequality follows from  $L_{G[V-V_i]^2} \preceq 2L_{G[V-V_i]} \preceq 2L_G$  by Fact 11.13. Now to bound the first two terms, we first rearrange them as

$$(-\text{SC}_n(G[V - V_i], V - V_{i+1}) + L_{G[V-V_{i+1}]}) + \frac{2}{n-2-i} L_{G[V-V_{i+1}]} \cdot \quad (62)$$

Then for the second term of (62), we have  $L_{G[V-V_{i+1}]} \preceq L_G$ , and hence

$$\lambda_{\max} \left( L_G^{\dagger/2} \frac{2}{n-2-i} L_{G[V-V_{i+1}]} L_G^{\dagger/2} \right) \leq \frac{2}{n-2-i} \leq \frac{2}{s-2} \leq \frac{4}{s} \leq \frac{4 \log n}{d_{\min}} \leq \frac{4}{10^6 \log^4 n}. \quad (63)$$

We then claim that the first term of (62) can be bounded by

$$\lambda_{\max} \left( L_G^{\dagger/2} (\text{SC}_n(G[V - V_i], V - V_{i+1}) - L_{G[V-V_{i+1}]}) L_G^{\dagger/2} \right) \leq \frac{1}{d_{\min} \zeta} \leq \frac{1}{8 \cdot 10^6 \log^5 n}. \quad (64)$$

To see why this is the case, let us define an edge set  $E_{i+1} = \{(v_{i+1}, w) \in G[V - V_i]\}$  and write  $L_{E_{i+1}}$  to denote the Laplacian of the subgraph of  $G$  induced by  $E_{i+1}$ . Notice that

$$L_{E_{i+1}}(:, v_{i+1}) = L_{G[V-V_i]}(:, v_{i+1}).$$

Therefore, we have

$$\text{SC}_n(G[V - V_i], V - V_{i+1}) - L_{G[V-V_{i+1}]} = \text{SC}_n(L_{E_{i+1}}, V - V_{i+1}).$$

Therefore to prove (64), it suffices to show that for any  $x$  such that  $x \in \text{range}(\text{SC}_n(L_{E_{i+1}}, V - V_{i+1}))$ , we have

$$x^T L^\dagger x \leq \frac{1}{\zeta d_{\min}} \cdot x^T \text{SC}_n(L_{E_{i+1}}, V - V_{i+1})^\dagger x = \frac{1}{\zeta d_{\min}} \cdot x^T L_{E_{i+1}}^\dagger x, \quad (65)$$

where the equality follows from Fact 11.10. Since  $G$  is a  $\zeta$ -expander with minimum degree  $d_{\min}$ , we know that  $x^T L^\dagger x \leq \frac{1}{\zeta d_{\min}} \|x\|^2$ . On the other hand, as  $G[E_{k+1}]$  is a star graph, and  $x \in \text{range}(\text{SC}_n(L_{E_{i+1}}, V - V_{i+1}))$  implies that  $x$  is supported on  $V - V_{i+1}$ , we have  $x^T L_{E_{i+1}}^\dagger x = \|x\|^2$ . This proves our desired inequality (65).

We also note that

$$\prod_{j=1}^i \left(1 + \frac{2}{n-1-j}\right) = \frac{n}{n-2} \frac{n-1}{n-3} \cdots \frac{n+1-i}{n-1-i} = \frac{n(n-1)}{(n-i)(n-1-i)} \leq \left(\frac{n}{s}\right)^2 \leq \log^2 n.$$

Therefore we have

$$\lambda_{\max}(Z_i) \leq \frac{1}{1000 \cdot \log^2 n}. \quad (66)$$

Fixing a  $j$ , we then consider bounding the spectral norm of  $\mathbb{E}[Z_j^2 | Y_1, \dots, Y_{j-1}]$ .

$$\begin{aligned} & \mathbb{E}[Z_j^2 | Y_1, \dots, Y_{j-1}] \\ & \leq \frac{2}{n-j} \cdot \left(\frac{n}{n-1-j}\right)^4 \sum_{v \in V - V_j} \left( L_G^{\dagger/2} \left( \text{SC}_n(G[V - V_j], V - V_j - \{v\}) - L_{G[V - V_j - \{v\}]} \right) L_G^{\dagger/2} \right)^2 + \\ & \quad \frac{2}{n-j} \cdot \left(\frac{n}{n-1-j}\right)^4 \sum_{v \in V - V_j} \left( L_G^{\dagger/2} \left( \frac{2}{n-2-j} L_{G[V - V_j - \{v\}]} + \frac{1}{n-j} \left( L_{G[V - V_j]^2} + 2L_{G[V - V_j]} \right) \right) L_G^{\dagger/2} \right)^2 \\ & \leq \frac{2 \log^5 n}{n-j} \cdot \frac{1}{\zeta d_{\min}} \sum_{v \in V - V_j} L_G^{\dagger/2} \left( \text{SC}_n(G[V - V_j], V - V_j - \{v\}) - L_{G[V - V_j - \{v\}]} \right) L_G^{\dagger/2} + \\ & \quad \frac{2 \log^5 n}{n-j} \cdot |V - V_j| \cdot \left(\frac{8}{s}\right)^2 I_{n \times n} \quad (\text{by (64), (63), (61)}) \\ & \leq \frac{2 \log^6 n}{n \zeta d_{\min}} L_G^{\dagger/2} L_{G[V - V_j]^2} L_G^{\dagger/2} + \frac{10000 \log^7 n}{n^2} I_{n \times n} \\ & \quad (\text{by Proposition 11.14 and } n-j \geq s \geq n/\log n) \\ & \leq \frac{64}{2 \cdot 10^6 \cdot n \log^3 n} I_{n \times n} + \frac{10000 \log^7 n}{n^2} I_{n \times n} \quad (\text{by Fact 11.13 and the lower bound on } d_{\min}) \\ & \leq \frac{1}{1000 n \log^2 n} I_{n \times n}. \end{aligned}$$

Therefore we have

$$\lambda_{\max}(\mathbb{E}[Z_j^2 | Y_1, \dots, Y_{j-1}]) \leq \frac{1}{1000 n \log^2 n}$$

and

$$\begin{aligned}
\lambda_{\max} \left( \sum_{j=0}^{n-s-1} \mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_{j-1}] \right) &\leq \sum_{j=0}^{n-s-1} \lambda_{\max} (\mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_{j-1}]) \\
&\leq n \cdot \frac{1}{1000n \log^2 n} \\
&= \frac{1}{1000 \log^2 n}.
\end{aligned} \tag{67}$$

We now invoke Theorem 11.16 with  $R = \frac{1}{1000 \log^2 n}$ ,  $\sigma^2 = \frac{1}{1000 \log^2 n}$ , and  $t = 1/2$ , and get

$$\Pr \left[ \exists i \in [0, n-s] : \lambda_{\max}(Y_i) \geq \frac{1}{10 \log n} \right] \leq n \cdot \exp \left( - \frac{\frac{1}{8}}{\frac{1}{1000 \log^2 n} + \frac{1}{6000 \log^2 n}} \right) \leq n^{-10}.$$

This coupled with Fact 11.9 implies that

$$L_{G[C]} \succeq \frac{1}{2} \cdot \frac{s^2}{n^2} \cdot \text{SC}(L_G, C).$$

Then we have

$$\begin{aligned}
\lambda_2 \left( D_{CC}^{-1/2} L_{G[C]} D_{CC}^{-1/2} \right) &\geq \frac{1}{2} \cdot \frac{s^2}{n^2} \cdot \lambda_2 \left( D_{CC}^{-1/2} \text{SC}(L_G, C) D_{CC}^{-1/2} \right) \\
&\geq \frac{1}{8} \cdot \frac{s}{n} \cdot \zeta,
\end{aligned}$$

where the last inequality holds with probability  $1 - n^{-10}$  by Lemma 11.12. Note that, by a Chernoff bound, we have with probability  $1 - n^{-10}$  that

$$\frac{1}{2} D_{CC} \preceq \frac{n}{s} D_{G[C]} \preceq 2 D_{CC}.$$

These two combined imply that  $G[C]$  is a  $\zeta/16$ -expander with probability  $1 - 3 \cdot n^{-10}$ .  $\square$

## 11.2 Matrix-weighted expanders are preserved under vertex sampling

We now introduce Schur complements and Cholesky factorization for Laplacian matrices of  $k \times k$  matrix-weighted graphs.

**Schur complements.** Let  $L \in \mathbb{R}^{nk \times nk}$  be the Laplacian of a  $k \times k$  matrix-weighted graph  $G$ , and let  $L(:, u) \in \mathbb{R}^{nk \times k}$  to denote the column block of  $L$  corresponding to vertex  $u$ . For a vertex  $u_1$ , we define the *Schur complement* of  $L$  with respect to  $u_1$  as

$$S^{(1)} = L - L(:, u_1) L_{u_1 u_1}^\dagger L(:, u_1)^T. \tag{68}$$

It is straightforward to see:

**Fact 11.17.** The entries on row block  $u_1$  of  $S^{(1)}$  and the entries on column block  $u_1$  of  $S^{(1)}$  are all zero.

We call the operation of subtracting  $L(:, u_1)L_{u_1 u_1}^\dagger L(:, u_1)^T$  from  $L$  the *elimination of vertex*  $u_1$ . Unlike the ordinary graph case, if we eliminate a sequence of vertices one by one, the resulting matrix is not necessarily the Laplacian of another  $k \times k$  matrix-weighted graph. So we alternatively define the *Schur complement with respect to a vertex set  $F$*  as

$$\text{SC}(L, C) = L_{CC} - L_{CF}L_{FF}^\dagger L_{FC}, \quad (69)$$

where  $C := V \setminus F$ . We also call  $\text{SC}(L, C)$  the *Schur complement of  $L$  (or  $G$ ) onto  $C$* . Note that when  $F = \{u_1\}$  consists of only a single vertex, we have  $S_{CC}^{(1)} = \text{SC}(L, C)$ . Also note that when  $k = 1$ , this definition matches the alternative definition of Schur complements in ordinary graphs [KLP<sup>+</sup>16]. We then connect  $\text{SC}(L, C)$  to the Laplacian  $L$  by introducing partial Cholesky factorization.

**Partial Cholesky factorization.** We show that  $L$  can be factorized in a similar way as in [KLP<sup>+</sup>16]. The proof of the following proposition is deferred to Appendix F.

**Proposition 11.18.** *We have*

$$L = \begin{pmatrix} I_{FF} & 0 \\ L_{CF}L_{FF}^\dagger & I_{CC} \end{pmatrix} \begin{pmatrix} L_{FF} & 0 \\ 0 & \text{SC}(L, C) \end{pmatrix} \begin{pmatrix} I_{FF} & 0 \\ L_{CF}L_{FF}^\dagger & I_{CC} \end{pmatrix}^T. \quad (70)$$

We call (70) a *partial Cholesky factorization* of  $L$ . We now state a fact about matrix factorizations of the form in (70). As in the previous subsection, for a set of vectors  $x_1, \dots, x_s$ , we will write  $\Pi_{x_1, \dots, x_s}$  to denote the projection matrix onto their linear span, and  $\Pi_{x_1, \dots, x_s}^\perp$  to denote the projection matrix onto the maximal subspace orthogonal to their linear span.

**Fact 11.19** (See e.g. [KLP<sup>+</sup>16]). Given an  $nk \times nk$  positive semidefinite matrix  $M$  divided into  $k \times k$  blocks that can be factorized as

$$M = \begin{pmatrix} I_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix} \begin{pmatrix} \mathcal{L}_{FF} & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} I_{FF} & 0 \\ \mathcal{L}_{CF} & I_{CC} \end{pmatrix}^T$$

for some bi-partition  $(C, F)$  of  $[n]$ . Let  $f_1, \dots, f_z \in \mathbb{R}^{nk}$  be a basis of the null space of  $M$ . Then

$$S^\dagger = \Pi_{(f_1)_C, \dots, (f_z)_C}^\perp (M^\dagger)_{CC} \Pi_{(f_1)_C, \dots, (f_z)_C}^\perp,$$

where  $(M^\dagger)_{CC}$  is  $M^\dagger$  restricted to row and column blocks in  $C$ .

As a direct corollary of the above fact, we have:

**Fact 11.20** (Corollary of Fact 11.9). Let  $f_1, \dots, f_z \in \mathbb{R}^{nk}$  be a basis of the null space of  $L$ . For any  $C \subseteq V$ ,

$$\text{SC}(L, C) = \Pi_{(f_1)_C, \dots, (f_z)_C}^\perp (L^\dagger)_{CC} \Pi_{(f_1)_C, \dots, (f_z)_C}^\perp.$$

Below we will need to use the following lemma, which is a direct consequence of the Matrix Chernoff bound (Theorem 4.2).

**Lemma 11.21.** Let  $f_1, \dots, f_\ell \in \mathbb{R}^{n^k}$  be any  $\ell$  orthonormal vectors. Define  $F : [n] \rightarrow \mathbb{R}^{\ell \times k}$  by

$$F(u) = \begin{pmatrix} (f_1)_u^T \\ \vdots \\ (f_\ell)_u^T \end{pmatrix} \in \mathbb{R}^{\ell \times k}.$$

Suppose for any  $u \in [n]$

$$\lambda_{\max}(F(u)^T F(u)) \leq R.$$

For an  $s \geq 100Rn \log n$ , let  $C \subseteq [n]$  be a uniformly random subset of indices of size  $s$ . Then we have with probability  $1 - n^{-10}$  that

$$\frac{1}{2} I_{\ell \times \ell} \preceq \frac{n}{s} \left( \sum_{u \in C} F(u) F(u)^T \right) \preceq 2I_{\ell \times \ell}$$

and therefore

$$\sum_{j=1}^{\ell} (f_j)_C (f_j)_C^T \preceq \frac{2s}{n} I_{nk \times nk}.$$

**Definition 11.4** (Subgraph preservation). Let  $G = (V, E)$  be a  $k \times k$  matrix-weighted graph. Let  $\lambda_1, \dots, \lambda_{nk}$  be eigenvalues of the normalized Laplacian  $N = D^{\dagger/2} L D^{\dagger/2}$  of  $G$  and let  $f_1, \dots, f_{nk}$  be a corresponding set of orthonormal eigenvectors. For a vertex subset  $U \subseteq V$  of size  $t$ , we say the vertex-induced subgraph  $G[U]$   $(\alpha, \beta, \zeta)$ -preserves  $G$  for some  $\alpha, \beta \geq 1$ ,  $\zeta \in (0, 1)$  iff

1. The null space of  $D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2}$  is exactly the linear span of  $\{(f_i)_U : \lambda_i = 0\}$ .
2. For all vectors  $x \in \mathbb{R}^{|U|k}$  such that  $x^T (f_i)_U = 0, \forall i : \lambda_i = 0$ ,

$$x^T \left( D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2} \right)^{\dagger} x \leq \cdot x^T \left( \alpha \cdot \frac{n^2}{t^2} \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} (f_i)_U (f_i)_U^T + \beta \cdot \frac{n}{t} \cdot \frac{1}{\zeta} I \right) x.$$

Therefore,  $G$   $(1, 1, \zeta)$ -preserves itself for any  $\zeta$ .

**Theorem 11.22.** Let  $G = (V, E)$  be a  $k \times k$  matrix-weighted  $(\gamma, \zeta, \psi)$ -almost regular expander with  $\zeta \leq 1/\log n$ . For an

$$s \geq 2 \cdot 10^6 \cdot \gamma \psi \zeta^{-1} k^2 n^{\frac{50000}{\log \log n}},$$

let  $C \subseteq V$  be a uniformly random vertex subset of size  $s$ . Then with probability at least  $1 - n^{-7}$ , the induced subgraph  $G[C]$   $(2, n^{\frac{4096}{\log \log n}}, \zeta)$ -preserves  $G$ .

The theorem is a consequence of applying the following lemma  $O(\frac{\log n}{\log \log n})$  times.

**Lemma 11.23.** Let  $G = (V, E)$  be a  $k \times k$  matrix-weighted  $(\gamma, \zeta, \psi)$ -almost regular expander with  $\zeta \leq 1/\log n$ . Suppose there exist an  $\alpha \in [1, 2]$ , a  $\beta \geq 1$ , and a  $\delta \in (0, 1)$ , and a

$$t \geq 2 \cdot 10^6 \cdot \alpha \beta \gamma \psi \zeta^{-1} k^2 \log^{10} n,$$

such that, a subgraph  $G[U]$  induced by a random vertex subset  $U$  of size  $t$   $(\alpha, \beta, \zeta)$ -preserves  $G$  with probability at least  $1 - \delta$ .

For an  $s \in [t/\log n, t]$ , let  $C \subseteq V$  be a uniformly random subset of size  $s$ . Then with probability at least  $1 - \delta - n^{-8}$ , the induced subgraph  $G[C]$   $((1 + \frac{1}{\log n})\alpha, 64\beta, \zeta)$ -preserves  $G$ .

*Proof of Lemma 11.23.* Let  $\lambda_1 \leq \dots \leq \lambda_{nk}$  be eigenvalues of the normalized Laplacian  $N = D^{\dagger/2} L D^{\dagger/2}$  of  $G$  and let  $f_1, \dots, f_{nk}$  be a corresponding set of orthonormal eigenvectors. Let  $z$  be such that  $\lambda_z = 0 < \lambda_{z+1}$ .

Consider generating  $C$  by first randomly sampling a subset  $U \subseteq V$  of size  $t$ , and then subsampling a subset  $C \subseteq U$  of size  $s$  from  $U$ . We first prove a claim about the Schur complement of  $L_{G[U]}$  onto  $C$ .

**Claim 11.24.** With probability at least  $1 - \delta - 2n^{-10}$ , we have for all  $x \in \mathbb{R}^{|C|k}$  such that  $x^T(f_i)_C = 0, \forall i : \lambda_i = 0$ ,

$$x^T \left( D_{CC}^{\dagger/2} \text{SC}(L_{G[U]}, C) D_{CC}^{\dagger/2} \right)^{\dagger} x \leq x^T \left( \alpha \cdot \frac{n^2}{t^2} \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} (f_i)_C (f_i)_C^T + \beta \cdot \frac{16s}{t} \cdot \frac{n}{t} \cdot \frac{1}{\zeta} I \right) x.$$

*Proof of Claim 11.24.* Let  $F := U \setminus C$ . By Proposition 11.18, we can factorize  $L_{G[U]}$  as

$$L_{G[U]} = \begin{pmatrix} I_{FF} & 0 \\ (L_{G[U]})_{CF} (L_{G[U]})_{FF}^{\dagger} & I_{CC} \end{pmatrix} \begin{pmatrix} (L_{G[U]})_{FF} & 0 \\ 0 & \text{SC}(L_{G[U]}, C) \end{pmatrix} \begin{pmatrix} I_{FF} & 0 \\ (L_{G[U]})_{CF} (L_{G[U]})_{FF}^{\dagger} & I_{CC} \end{pmatrix}^T.$$

By multiplying  $D_{UU}^{\dagger/2}$  on both sides, we get a factorization of  $D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2}$ :

$$\begin{aligned} & D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2} \\ &= \begin{pmatrix} D_{FF}^{\dagger/2} & 0 \\ D_{CC}^{\dagger/2} (L_{G[U]})_{CF} (L_{G[U]})_{FF}^{\dagger} & D_{CC}^{\dagger/2} \end{pmatrix} \begin{pmatrix} (L_{G[U]})_{FF} & 0 \\ 0 & \text{SC}(L_{G[U]}, C) \end{pmatrix} \begin{pmatrix} D_{FF}^{\dagger/2} & 0 \\ D_{CC}^{\dagger/2} (L_{G[U]})_{CF} (L_{G[U]})_{FF}^{\dagger} & D_{CC}^{\dagger/2} \end{pmatrix}^T \\ &= \begin{pmatrix} I_{FF} & 0 \\ D_{CC}^{\dagger/2} (L_{G[U]})_{CF} (L_{G[U]})_{FF}^{\dagger} D_{FF}^{1/2} & I_{CC} \end{pmatrix} \begin{pmatrix} D_{FF}^{\dagger/2} (L_{G[U]})_{FF} D_{FF}^{\dagger/2} & 0 \\ 0 & D_{CC}^{\dagger/2} \text{SC}(L_{G[U]}, C) D_{CC}^{\dagger/2} \end{pmatrix} \\ &\quad \begin{pmatrix} I_{FF} & 0 \\ D_{CC}^{\dagger/2} (L_{G[U]})_{CF} (L_{G[U]})_{FF}^{\dagger} D_{FF}^{1/2} & I_{CC} \end{pmatrix}^T, \end{aligned}$$

where in the last inequality we have used that the rows of  $(L_{G[U]})_{FF}$  are all in the range of  $D_{FF}$ . In the case that  $G[U]$   $(\alpha, \beta, \zeta)$ -preserves  $G$ , we have that the null space of  $D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2}$  is exactly the linear span of  $\{(f_i)_U : 1 \leq i \leq z\}$ . Then by Fact 11.19,

$$\left( D_{CC}^{\dagger/2} \text{SC}(L_{G[U]}, C) D_{CC}^{\dagger/2} \right)^{\dagger} = \Pi_{(f_1)_C, \dots, (f_z)_C}^{\perp} \left( \left( D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2} \right)^{\dagger} \right)_{CC} \Pi_{(f_1)_C, \dots, (f_z)_C}^{\perp}.$$

Let  $\tilde{\lambda}_1, \dots, \tilde{\lambda}_\ell$  be all eigenvalues of  $D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2}$  that are in the range  $(0, \frac{1}{2} \cdot \frac{t}{n}]$ . Let  $\tilde{f}_1, \dots, \tilde{f}_\ell$  be a set of orthonormal eigenvectors, and define the embedding  $\tilde{F} : V \rightarrow \mathbb{R}^{\ell \times k}$  by

$$\tilde{F}(u) = \begin{pmatrix} (\tilde{f}_1)_u^T \\ \vdots \\ (\tilde{f}_\ell)_u^T \end{pmatrix} \in \mathbb{R}^{\ell \times k}.$$

Since  $s \in [t/\log n, t]$  and  $\zeta \leq 1/\log n$ , we have  $\frac{t}{2s} \cdot \frac{t}{n} \cdot \zeta \leq \frac{1}{2} \cdot \frac{t}{n}$ . Then we have

$$\begin{aligned} & \left( D_{UU}^{\dagger/2} L_{G[U]} D_{UU}^{\dagger/2} \right)^\dagger \preceq \\ & \Pi_{(f_1)_U, \dots, (f_z)_U}^\perp \left( \alpha \cdot \frac{n^2}{t^2} \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} (f_i)_U (f_i)_U^T + \beta \cdot \frac{4n}{t} \sum_{j=1}^{\ell} \frac{1}{\zeta} \tilde{f}_j \tilde{f}_j^T + \beta \cdot \frac{4}{\frac{t}{2s} \cdot \frac{t}{n} \cdot \zeta} I \right) \Pi_{(f_1)_U, \dots, (f_z)_U}^\perp. \end{aligned} \quad (71)$$

Using the matrix Chernoff bound and that  $t \geq 2 \cdot 10^6 \gamma k \log^5 n$ , we have with probability  $1 - n^{-10}$  that

$$\frac{1}{2} D_{G[U]} \preceq \frac{t}{n} D_{UU} \preceq 2 D_{G[U]}. \quad (72)$$

As a result,  $G[U]$  is  $2\gamma$ -almost regular. Therefore by applying Lemma 8.2 to  $G[U]$  with  $\tilde{D} = \frac{t}{n} D_{UU}$  with  $\kappa = 2$  and  $\delta = \frac{1}{2}$ , we have for each vertex  $u$

$$\lambda_{\max} \left( \tilde{F}(u)^T \tilde{F}(u) \right) \leq \frac{32\gamma k}{n}.$$

By Lemma 11.21, with probability  $1 - n^{-10}$ ,

$$\sum_{j=1}^{\ell} (\tilde{f}_j)_C (\tilde{f}_j)_C^T \preceq \frac{2s}{t} I. \quad (73)$$

Now by restricting both sides of (71) to vertices in  $C$  and then plugging in (73), we have our desired claim.  $\square$

We now set up a matrix-valued martingale. To that end, we also need the notion of graph squaring as we did in the ordinary graph case.

**Definition 11.5** (Graph squaring). For a graph  $G = (V, E)$ , define the square of  $G$  by

$$L_{G^2} \stackrel{\text{def}}{=} D - AD^\dagger A.$$

The proof of the following claim is deferred to Appendix F.

**Claim 11.25.**  $L_{G^2} \preceq 2L_G$  for any graph  $G$ .

Below, we will write  $\text{SC}_n(L, C)$  or  $\text{SC}_n(G, C)$  to denote an  $nk \times nk$  matrix obtained by augmenting  $\text{SC}(L, C)$  to  $nk \times nk$  by adding zeros on row and column blocks in  $V \setminus C$ . We will also abuse the notation a bit and write  $L_{G[C]}$  to denote an  $nk \times nk$  matrix obtained by augmenting  $L_{G[C]}$  to  $nk \times nk$  by adding zeros on row and column blocks in  $V \setminus C$ . The proof of the following proposition is also deferred to Appendix F.

**Proposition 11.26.**  $\sum_{u \in V} \text{SC}_n(L_G, V \setminus \{u\}) = L_{G^2} + (n - 2)L_G$  for any graph  $G$ .

Let  $v_1, \dots, v_t$  be a uniformly random permutation of vertices in  $U$ . Define

$$V_i = \{v_1, \dots, v_i\}.$$

Let  $H = G[U]$ . Consider the following sequence of matrices:

$$\begin{aligned} X_0 &= L_H \\ X_1 &= L_H(:, v_1) D_{v_1}^\dagger L_H(:, v_1)^T + \left(1 + \frac{2}{t-2}\right) L_{H[U-V_1]} + \frac{1}{t} (L_{H^2} - 2L_H) \\ X_2 &= L_H(:, v_1) D_{v_1}^\dagger L_H(:, v_1)^T + \frac{1}{t} (L_{H^2} - 2L_H) + \left(1 + \frac{2}{t-2}\right) \cdot \\ &\quad \left( L_{H[U-V_1]}(:, v_2) \left(D_{v_2}^{H[U-V_1]}\right)^\dagger L_{H[U-V_1]}(:, v_2)^T + \left(1 + \frac{2}{t-3}\right) L_{H[U-V_2]} \right. \\ &\quad \left. + \frac{1}{t-1} (L_{H[U-V_1]^2} - 2L_{H[U-V_1]}) \right) \end{aligned}$$

$X_{i+1}$  is obtained by replacing the  $L_{H[U-V_i]}$  in  $X_i$  by

$$\begin{aligned} L_{H[U-V_i]}(:, v_{i+1}) \left(D_{v_{i+1}}^{H[U-V_i]}\right)^\dagger L_{H[U-V_i]}(:, v_{i+1})^T &+ \left(1 + \frac{2}{t-2-i}\right) L_{H[U-V_{i+1}]} \\ &+ \frac{1}{t-i} (L_{H[U-V_i]^2} - 2L_{H[U-V_i]}). \end{aligned} \tag{74}$$

Here, we have used  $D_v^H$  to denote the degree of vertex  $v$  in graph  $H$ . The proof of the following claim is also deferred to Appendix F.

**Claim 11.27.**  $X_0, X_1, X_2, \dots$  is a matrix-valued martingale.

Define  $Y_i = L_H^{\dagger/2} X_i L_H^{\dagger/2} - L_H^{\dagger/2} L_H L_H^{\dagger/2}$ . As  $X_0, X_1, \dots$  is a martingale, so is  $Y_0, Y_1, \dots$ , by the linearity of expectation. Additionally, since  $X_0 = L_H$ ,  $Y_0, Y_1, \dots$  has zero expectation. We focus on the first  $t-s+1$  terms  $Y_0, Y_1, Y_2, \dots, Y_{t-s}$ . First we calculate the difference sequence  $Z_i = Y_{i+1} - Y_i$ :

$$\begin{aligned} Z_i &= \prod_{j=1}^i \left(1 + \frac{2}{t-1-j}\right) \cdot \\ &\quad L_H^{\dagger/2} \left( -\text{SC}_n(H[U-V_i], U-V_{i+1}) + \left(1 + \frac{2}{t-2-i}\right) L_{H[U-V_{i+1}]} \right. \\ &\quad \left. - \frac{1}{t-i} (L_{H[U-V_i]^2} - 2L_{H[U-V_i]}) \right) L_H^{\dagger/2}, \end{aligned}$$

where we have used the definition of Schur complements. We then consider to bound the maximum (in absolute value) eigenvalue of  $Z_i$ . For the last term, we have

$$\lambda_{\max} \left( L_H^{\dagger/2} \frac{1}{t-i} (L_{H[U-V_i]^2} - 2L_{H[U-V_i]}) L_H^{\dagger/2} \right) \leq \frac{4}{t-i} \leq \frac{4}{s} \leq \frac{4 \log n}{t} \leq \frac{4}{10^6 \log^6 n}, \quad (75)$$

where the first inequality follows from  $L_{H[U-V_i]^2} \preceq 2L_{H[U-V_i]} \preceq 2L_H$  by Claim 11.25. Now to bound the first two terms, we first rearrange them as

$$(-\text{SC}_n(H[U-V_i], U-V_{i+1}) + L_{H[U-V_{i+1}]}) + \frac{2}{t-2-i} L_{H[U-V_{i+1}]} \cdot \quad (76)$$

Then for the second term of (76), we have  $L_{H[U-V_{i+1}]} \preceq L_H$ , and hence

$$\lambda_{\max} \left( L_H^{\dagger/2} \frac{2}{t-2-i} L_{H[U-V_{i+1}]} L_H^{\dagger/2} \right) \leq \frac{2}{t-2-i} \leq \frac{2}{s-2} \leq \frac{4}{s} \leq \frac{4 \log n}{t} \leq \frac{4}{10^6 \log^6 n}. \quad (77)$$

We then claim that the first term of (76) can be bounded by

$$\lambda_{\max} \left( L_H^{\dagger/2} (\text{SC}_n(H[U-V_i], U-V_{i+1}) - L_{H[U-V_{i+1}]}) L_H^{\dagger/2} \right) \leq \alpha \frac{\psi k^2}{t} + \beta \frac{\gamma k}{\zeta t} \leq \frac{1}{8 \cdot 10^6 \log^6 n}. \quad (78)$$

To see why this is the case, let us define an edge set  $E_{i+1} = \{(v_{i+1}, w) \in E : w \in U - V_{i+1}\}$ , and write  $L_{E_{i+1}}$  to denote the Laplacian of the subgraph of  $H$  induced by  $E_{i+1}$ . Notice that

$$L_{E_{i+1}}(:, v_{i+1}) = L_{H[U-V_i]}(:, v_{i+1}).$$

Therefore, we have

$$\text{SC}(H[U-V_i], U-V_{i+1}) - L_{H[U-V_{i+1}]} = \text{SC}(L_{E_{i+1}}, U-V_{i+1}).$$

Therefore to prove (78), it suffices to show that for any vector  $x$  such that  $x \in \text{range}(\text{SC}_n(L_{E_{i+1}}, U-V_{i+1}))$ , we have

$$x^T L_H^\dagger x \leq \left( \alpha \frac{\psi k^2}{t} + \beta \frac{\gamma k}{\zeta t} \right) \cdot x^T \text{SC}(L_{E_{i+1}}, U-V_{i+1})^\dagger x = \left( \alpha \frac{\psi k^2}{t} + \beta \frac{\gamma k}{\zeta t} \right) \cdot x^T L_{E_{i+1}}^\dagger x. \quad (79)$$

Since  $x \in \text{range}(\text{SC}_n(L_{E_{i+1}}, U-V_{i+1}))$ , we can write  $x$  as a liner combination of the incidence vectors of the edges in  $E_{i+1}$ :

$$x = \sum_{e \in E_{i+1}} c_e b_e$$

for some coefficients  $c_e$ 's, where we have  $x_{v_{i+1}} = \sum_{e \in E_{i+1}} c_e (b_e)_{v_{i+1}} = 0$ . Then since  $G[E_{i+1}]$  is a star graph, we have

$$x^T L_{E_{i+1}}^\dagger x = \sum_{e \in E_{i+1}} c_e^2.$$

We now calculate  $x^T L_H^\dagger x$  by

$$\begin{aligned}
x^T L_H^\dagger x &= x^T D_{UU}^{\dagger/2} \left( D_{UU}^{\dagger/2} L_H D_{UU}^{\dagger/2} \right)^\dagger D_{UU}^{\dagger/2} x \\
&\leq x^T D_{UU}^{\dagger/2} \left( \alpha \cdot \frac{n^2}{t^2} \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} (f_i)_U (f_i)_U^T + \beta \cdot \frac{n}{t} \cdot \frac{1}{\zeta} I \right) D_{UU}^{\dagger/2} x \\
&\quad (\text{since } H = G[U] \text{ } (\alpha, \beta, \zeta)\text{-preserves } G) \\
&\leq |E_{i+1}| \cdot \alpha \cdot \frac{n^2}{t^2} \sum_{e \in E_{i+1}} c_e^2 \frac{\psi k^2}{n^2} + \beta \frac{n}{t} \cdot \frac{1}{\zeta} \left\| \sum_{e \in E_{i+1}} c_e D_{UU}^{\dagger/2} b_e \right\|^2 \\
&\quad (\text{by expanding and that } G \text{ is a } (\gamma, \zeta, \psi)\text{-almost regular expander}) \\
&\leq \alpha \frac{\psi k^2}{t} \left( \sum_{e \in E_{i+1}} c_e^2 \right) + \beta \frac{n}{t} \cdot \frac{1}{\zeta} \left( \sum_{(w, v_{i+1}) \in E_{i+1}} c_e^2 \|D_w^{\dagger/2} \phi_{v_{i+1} w}\|^2 \right) \\
&\quad (\text{by } |E_{i+1}| \leq t \text{ and that } \sum_{e \in E_{i+1}} c_e D_{UU}^{\dagger/2} b_e \text{ is zero on } v_{i+1}) \\
&\leq \left( \alpha \frac{\psi k^2}{t} + \beta \frac{\gamma k}{\zeta t} \right) \left( \sum_{e \in E_{i+1}} c_e^2 \right) \\
&\quad (\text{by } \gamma\text{-almost regularity of } G).
\end{aligned}$$

This proves our desired inequality (79).

We also note that

$$\prod_{j=1}^i \left( 1 + \frac{2}{t-1-j} \right) = \frac{t}{t-2} \frac{t-1}{t-3} \cdots \frac{t+1-i}{t-1-i} = \frac{t(t-1)}{(t-i)(t-1-i)} \leq \left( \frac{t}{s} \right)^2 \leq \log^2 n.$$

Therefore

$$\lambda_{\max}(Z_i) \leq \frac{1}{1000 \cdot \log^4 n}. \tag{80}$$

Fixing a  $j$ , we then consider bounding the spectral norm of  $\mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_{j-1}]$ .

$$\begin{aligned}
& \mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_{j-1}] \\
& \preceq \frac{2}{t-j} \cdot \left( \frac{t}{t-1-j} \right)^4 \sum_{v \in U-V_j} \left( L_H^{\dagger/2} \left( \text{SC}_n(H[U-V_j], U-V_j - \{v\}) - L_{H[U-V_j-\{v\}]} \right) L_H^{\dagger/2} \right)^2 + \\
& \quad \frac{2}{t-j} \cdot \left( \frac{t}{t-1-j} \right)^4 \sum_{v \in U-V_j} \left( L_H^{\dagger/2} \left( \frac{2}{t-2-j} L_{H[U-V_j-\{v\}]} + \frac{1}{t-j} \left( L_{H[U-V_j]^2} + 2L_{H[U-V_j]} \right) \right) L_H^{\dagger/2} \right)^2 \\
& \preceq \frac{2 \log^5 n}{t-j} \cdot \left( \alpha \frac{\psi k^2}{t} + \beta \frac{\gamma k}{\zeta t} \right) \sum_{v \in U-V_j} L_H^{\dagger/2} \left( \text{SC}_n(H[U-V_j], U-V_j - \{v\}) - L_{H[U-V_j-\{v\}]} \right) L_H^{\dagger/2} + \\
& \quad \frac{2 \log^5 n}{t-j} \cdot |U-V_j| \cdot \left( \frac{8}{s} \right)^2 I_{tk \times tk} \quad (\text{by (78),(77),(75)}) \\
& \preceq \frac{2 \log^6 n}{t} \left( \alpha \frac{\psi k^2}{t} + \beta \frac{\gamma k}{\zeta t} \right) L_H^{\dagger/2} L_{H[U-V_j]^2} L_H^{\dagger/2} + \frac{10000 \log^7 n}{t^2} I_{tk \times tk} \\
& \quad (\text{by Proposition 11.26 and } t-j \geq s \geq t/\log n) \\
& \preceq \frac{64}{2 \cdot 10000 \cdot t \log^4 n} I_{tk \times tk} + \frac{10000 \log^7 n}{t^2} I_{tk \times tk} \quad (\text{by Claim 11.25 and the values of } \gamma, \zeta, \psi) \\
& \preceq \frac{1}{1000t \log^4 n} I_{tk \times tk}.
\end{aligned}$$

Therefore we have

$$\lambda_{\max} (\mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_{j-1}]) \leq \frac{1}{1000t \log^4 n}$$

and

$$\begin{aligned}
\lambda_{\max} \left( \sum_{j=0}^{t-s-1} \mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_{j-1}] \right) & \leq \sum_{j=0}^{t-s-1} \lambda_{\max} (\mathbb{E} [Z_j^2 \mid Y_1, \dots, Y_{j-1}]) \\
& \leq t \cdot \frac{1}{1000t \log^4 n} \\
& = \frac{1}{1000 \log^4 n}.
\end{aligned} \tag{81}$$

We now invoke Theorem 11.16 with  $R = \frac{1}{1000 \log^4 n}$ ,  $\sigma^2 = \frac{1}{1000 \log^4 n}$ , and  $t = \frac{1}{10 \log n}$ , and get

$$\Pr [\exists i \in [0, n-s] : \lambda_{\max}(Y_i) \geq 1/2] \leq n \cdot \exp \left( -\frac{\frac{1}{200 \log^2 n}}{\frac{1}{1000 \log^4 n} + \frac{1}{30000 \log^5 n}} \right) \leq n^{-10}.$$

This coupled with Fact 11.19 implies that

$$L_{G[C]} \succeq \left( 1 - \frac{1}{5 \log n} \right) \cdot \frac{s^2}{t^2} \cdot \text{SC}(L_{G[U]}, C). \tag{82}$$

Then using Claim 11.24, for all  $x \in \mathbb{R}^{|C|^k}$  such that  $x^T(f_i)_C = 0, \forall i : \lambda_i = 0$ ,

$$\begin{aligned} & x^T \left( D_{CC}^{\dagger/2} L_{G[C]} D_{CC}^{\dagger/2} \right)^\dagger x \\ & \leq \left( 1 + \frac{1}{2 \log n} \right) \cdot \frac{t^2}{s^2} \cdot x^T \left( \alpha \cdot \frac{n^2}{t^2} \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} (f_i)_C (f_i)_C^T + \beta \cdot \frac{16s}{t} \cdot \frac{n}{t} \cdot \frac{1}{\zeta} I \right) x \\ & \leq x^T \left( \left( 1 + \frac{1}{\log n} \right) \alpha \cdot \frac{n^2}{s^2} \sum_{i: \lambda_i \in (0, \zeta]} \frac{1}{\lambda_i} (f_i)_C (f_i)_C^T + 64\beta \cdot \frac{n}{s} \cdot \frac{1}{\zeta} I \right) x. \end{aligned} \quad (83)$$

(82), (84) together imply that  $G[C]$   $((1 + \frac{1}{\log n})\alpha, 64\beta, \zeta)$ -preserves  $G$ , as desired.  $\square$

## 12 A lower bound for weighted spectral sparsification

In this section, we prove a superlinear lower bound of  $n^{21/20-o(1)}$  for computing some  $O(1)$ -spectral sparsifier.

**Theorem 1.5** (Lower bound for weighted spectral sparsification). *There exist constants  $\epsilon, \delta \in (0, 1)$  such that any incidence sketch of  $N$  measurements that computes a  $(1+\epsilon)$ -spectral sparsifier with probability  $\geq 1 - \delta$  on any  $w$  must satisfy  $N \geq n^{21/20-o(1)}$ .*

By Proposition 3.3, in order to prove Theorem 1.5, it suffices to prove Theorem 3.8. Recall that  $B^w \in \mathbb{R}^{\binom{n}{2} \times n}$  is the weighted signed edge-vertex incidence matrix of the input graph generated from the distribution specified in Section 3.

**Theorem 3.8.** *For any fixed sketching matrix  $\Phi \in \mathbb{R}^{k \times \binom{n}{2}}$  where  $k \leq n^{1/20-\epsilon}$  for some constant  $\epsilon > 0$ , we have*

$$\mathbb{E}_\pi [d_{\text{TV}}((\Phi B^w)_{\pi, \text{yes}}, (\Phi B^w)_{\pi, \text{no}})] \leq o(1).$$

*Proof of Theorem 3.8.* Similar to Section 3.3, we can assume w.l.o.g. the number of  $(u, v) \in E$  for which  $\phi_{uv} \neq 0$  is at least  $\Omega(n^2)$ , since otherwise

$$\mathbb{E}_\pi [d_{\text{TV}}((\Phi B^w)_{\pi, \text{yes}}, (\Phi B^w)_{\pi, \text{no}})] \leq O(1) \cdot \Pr_\pi [\phi_{\pi(1)\pi(n/2+1)} \neq 0] \leq o(1),$$

and we would already have our desired result. Let us then define the  $k \times k$  matrix-weighted graph  $H_\phi = (V, E_\phi)$ , where  $E_\phi$  contains *all* edges  $(u, v)$  whose  $\phi_{uv} \neq 0$  (including the ones not present in the input graph), and each edge  $(u, v)$  has matrix weight  $\phi_{uv} \phi_{uv}^T$ . Thus by the above assumption we have  $|E_\phi| \geq \Omega(n^2)$ .

By Theorem 3.13, there exists a scaling  $s : E_\phi \rightarrow [0, 1]$  s.t.  $H_\phi^s$  is a  $(\gamma, \zeta, \psi)$ -almost regular expander, where  $(\gamma, \zeta, \psi) = (8 \log n, 1/\log n, 16k^2 \log^3 n)$ , and  $|(u, v) \in E_\phi : s_{uv} < 1| \leq o(n^2)$ . Let  $\lambda_1 \leq \dots \leq \lambda_{nk}$  be the eigenvalues of the normalized Laplacian matrix of  $H_\phi^s$ , and let  $f_1, \dots, f_{nk} \in \mathbb{R}^{nk}$  be a corresponding set of orthonormal eigenvectors.

For  $i = 0, \dots, n^{4/5} - 1$ , let  $B_i$  denote the vertices in the  $i^{\text{th}}$  block of the input graph:

$$B_i \stackrel{\text{def}}{=} \left\{ \pi(n^{1/5}i + 1), \dots, \pi(n^{1/5}i + n^{1/5}) \right\}.$$

Since  $k \leq n^{1/20-\epsilon}$ , we have  $2 \cdot 10^6 \gamma \psi \zeta^{-1} k^2 n^\epsilon \leq n^{1/5}$ . Thus by Theorem 3.14, with probability  $1 - 1/n^4$ , all vertex-induced subgraphs  $H_\phi^s[B_i \cup B_{i+1}]$  preserve  $H_\phi^s$  in the sense of (7). Using this fact, we prove the following claim:

**Claim 12.1.** For each edge  $(x, y)$  with  $s_{xy} = 1$ , conditioned on  $(x, y)$  being the crossing edge (i.e.  $\{\pi(1), \pi(n/2 + 1)\} = \{x, y\}$ ), with probability at least  $1 - 1/n^2$  over  $\pi$ , we have that  $b_{xy}$  is in the range of  $L_\pi$ , and

$$b_{xy}^T L_\pi^\dagger b_{xy} \leq \frac{k^2}{n^{2/5-o(1)}}. \quad (85)$$

Once again by  $k \leq n^{1/20-\epsilon}$ , the RHS of (85) is  $o(1)$ . Since  $s_{uv} = 1$  holds for a  $1 - o(1)$  fraction of the edges, we have, by Proposition 3.4,  $\mathbb{E}_\pi [d_{\text{TV}}((\Phi B^w)_{\pi, \text{yes}}, (\Phi B^w)_{\pi, \text{no}})] = o(1)$ .  $\square$

*Proof of Claim 12.1.* Consider drawing  $\pi$  from the conditional distribution on  $\pi(1) = x$  and  $\pi(n/2 + 1) = y$ . Suppose w.l.o.g.  $(x, y)$  is oriented  $x \rightarrow y$ , so  $(b_{xy})_x = \phi_{xy}$  and  $(b_{xy})_y = -\phi_{xy}$ . Let  $E_\pi$  be the set of edges of  $\mathcal{H}_\pi$ , where we recall that  $\mathcal{H}_\pi$  consists of the non-crossing edges, such that each edge  $(u, v)$  has matrix weight  $n^{4/5} \log^{-1} n \phi_{uv} \phi_{uv}^T \in \mathbb{R}^{k \times k}$ . Let us call a function  $\mathbf{f} : E_\pi \rightarrow \mathbb{R}$  a *flow* in the graph  $\mathcal{H}_\pi$ , and say  $\mathbf{f}$  routes a *demand*  $d \in \mathbb{R}^{nk}$  if  $\sum_{e \in E_\pi} \mathbf{f}_e b_e = d$ .

By Fact 4.3, to prove (85), it suffices to show that there exists a flow  $\mathbf{f}$  in  $\mathcal{H}_\pi$  that routes demand  $n^{-2/5} b_{xy}$  satisfying  $\|\mathbf{f}\|_2^2 \leq \frac{k^2}{n^{2/5-o(1)}}$ . By scaling, it then suffices to show that there is a flow  $\mathbf{f}$  that routes demand  $b_{xy}$  satisfying  $\|\mathbf{f}\|_2^2 \leq k^2 n^{2/5+o(1)}$ . To construct such a flow, we consider first breaking the demand  $b_{xy}$  into demands  $d_0, d_1, \dots, d_{n^{4/5}/2-1} \in \mathbb{R}^{nk}$  such that

1.  $\sum_{0 \leq i < n^{4/5}/2} d_i = b_{xy}$ .
2.  $d_i$  is supported on  $B_i \cup B_{i+1}$ .

We will then route each  $d_i$  in the vertex induced subgraph  $H_\phi[B_i \cup B_{i+1}]$  using a flow  $\mathbf{f}_i$ , and finally obtain a flow  $\mathbf{f} = \sum_{0 \leq i < n^{4/5}/2} \mathbf{f}_i$  that routes  $b_{xy}$  and satisfies  $\|\mathbf{f}\|_2^2 \leq 2 \sum_{0 \leq i < n^{4/5}/2} \|\mathbf{f}_i\|_2^2$ .

To show how these  $d_i$ 's are constructed, let us first define vectors  $\tilde{d}_0, \dots, \tilde{d}_{n^{4/5}/2} \in \mathbb{R}^{nk}$  where  $\tilde{d}_i$  is supported on  $B_i$ . We will later define each  $d_i$  by letting  $d_i = \tilde{d}_i - \tilde{d}_{i+1}$ . Let  $D \in \mathbb{R}^{nk \times nk}$  be the degree matrix of  $H_\phi^s$ . For now we would like  $\tilde{d}_i$  to satisfy

$$\forall j : 0 \leq \lambda_j \leq \zeta, \quad f_j^T D^{\dagger/2} \tilde{d}_i = (1 - \mu_i)(f_j)_x^T D_{xx}^{\dagger/2} \phi_{xy} + \mu_i (f_j)_y^T D_{yy}^{\dagger/2} \phi_{xy}. \quad (86)$$

where we define

$$\mu_i := \frac{i}{n^{4/5}/2}.$$

In particular, we would like  $f_j^T D^{\dagger/2} \tilde{d}_0 = (f_j)_x^T D_{xx}^{\dagger/2} \phi_{xy}$  and  $f_j^T D^{\dagger/2} \tilde{d}_{n^{4/5}/2} = (f_j)_y^T D_{yy}^{\dagger/2} \phi_{xy}$  for all  $j$  such that  $0 \leq \lambda_j \leq \zeta$ . Thus, we let  $\tilde{d}_0$  be such that

$$(\tilde{d}_0)_u = \begin{cases} \phi_{xy} & u = x \\ 0 & \text{otherwise} \end{cases}$$

and let  $\tilde{d}_{n^{4/5}/2}$  be such that

$$(\tilde{d}_{n^{4/5}/2})_u = \begin{cases} \phi_{xy} & u = y \\ 0 & \text{otherwise.} \end{cases}$$

We would then like to construct the remaining  $\tilde{d}_i$ 's such that each  $\|D^{\dagger/2}\tilde{d}_i\|_2$  is small. First let us write the linear equations (86) in matrix form. Let  $\ell$  be s.t.  $\lambda_\ell \leq \zeta < \lambda_{\ell+1}$ . Define a function  $F : V \rightarrow \mathbb{R}^{\ell \times k}$  such that

$$F(u) = \begin{pmatrix} (f_1)_u^T \\ \vdots \\ (f_\ell)_u^T \end{pmatrix}$$

and define another function  $G : \{B_0, \dots, B_{n^{4/5}/2}\} \rightarrow \mathbb{R}^{\ell \times n^{1/5}k}$  such that

$$G(B_i) = \begin{pmatrix} (f_1)_{B_i}^T \\ \vdots \\ (f_\ell)_{B_i}^T \end{pmatrix}$$

Then we can write (86) as

$$G(B_i)D_{B_i B_i}^{\dagger/2}(\tilde{d}_i)_{B_i} = (1 - \mu_i)F(x)D_{xx}^{\dagger/2}\phi_{xy} + \mu_i F(y)D_{yy}^{\dagger/2}\phi_{xy} \quad (87)$$

Let us write the RHS as  $r_i := (1 - \mu_i)F(x)D_{xx}^{\dagger/2}\phi_{xy} + \mu_i F(y)D_{yy}^{\dagger/2}\phi_{xy}$ . Then by Fact 4.3, we can find a  $(\tilde{d}_i)_{B_i}$  satisfying (87) such that  $\|D_{B_i B_i}^{\dagger/2}(\tilde{d}_i)_{B_i}\|_2^2 = r_i^T(G(B_i)G(B_i)^T)^{-1}r_i$ . Now notice

$$G(B_i)G(B_i)^T = \sum_{u \in B_i} F(u)F(u)^T.$$

Since  $f_1, \dots, f_\ell$  are orthonormal, we have  $\sum_{u \in V} F(u)F(u)^T = I_{\ell \times \ell}$ . By Lemma 8.2, we have for all  $u \in V$  that  $\lambda_{\max}(F(u)F(u)^T) \leq \frac{\gamma k}{(1-\zeta)^2 n} \leq \frac{16k \log n}{n}$ . Since the vertices in  $B_i$  are chosen uniformly at random, by a Matrix Chernoff bound (Theorem 4.2), we have with probability  $\geq 1 - n^{-5}$  that  $\frac{1}{2}I \preceq n^{4/5}G(B_i)G(B_i)^T \preceq 2I$ , which implies  $r_i^T(G(B_i)G(B_i)^T)^{-1}r_i \leq 2n^{4/5} \|r_i\|^2$ . As a result, we can find  $\tilde{d}_1, \dots, \tilde{d}_{n^{4/5}/2-1}$  such that

$$\|D^{\dagger/2}\tilde{d}_i\|_2^2 \leq 4n^{4/5} \left( \|F(x)D_{xx}^{\dagger/2}\phi_{xy}\|_2^2 + \|F(y)D_{yy}^{\dagger/2}\phi_{xy}\|_2^2 \right).$$

For simplicity let  $C_i := B_i \cup B_{i+1}$ . We then construct each  $d_i$  for  $0 \leq i < n^{4/5}/2$  by letting  $d_i := \tilde{d}_i - \tilde{d}_{i+1}$ . Thus  $d_i$  is indeed supported on  $C_i$ , and we have  $\sum_{0 \leq i < n^{4/5}/2} d_i = b_{xy}$ . Now we consider how to route each  $d_i$  in  $H_\phi[B_i \cup B_{i+1}]$ . Once again by Fact 4.3, there is a flow  $\mathbf{f}_i$  with

$\ell_2^2$ -norm  $(d_i)_{C_i}^T L_{H_\phi[C_i]}^\dagger (d_i)_{C_i} \leq (d_i)_{C_i}^T L_{H_\phi^s[C_i]}^\dagger (d_i)_{C_i}$ . We then write

$$\begin{aligned}
& (d_i)_{C_i}^T L_{H_\phi^s[C_i]}^\dagger (d_i)_{C_i} \\
&= (d_i)_{C_i}^T D_{C_i C_i}^{\dagger/2} \left( D_{C_i C_i}^{\dagger/2} L_{H_\phi^s[C_i]} D_{C_i C_i}^{\dagger/2} \right)^\dagger D_{C_i C_i}^{\dagger/2} (d_i)_{C_i} \\
&\leq n^{o(1)} (d_i)_{C_i}^T D_{C_i C_i}^{\dagger/2} \left( \frac{n^2}{n^{2/5}} \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} (f_j)_{C_i} (f_j)_{C_i}^T + \frac{n}{n^{1/5}} (\log n) I \right) D_{C_i C_i}^{\dagger/2} (d_i)_{C_i} \\
&\leq n^{8/5+o(1)} \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} \left( f_j^T D^{\dagger/2} (\tilde{d}_i - \tilde{d}_{i+1}) \right)^2 + n^{4/5+o(1)} d_i^T D^\dagger d_i \\
&= n^{8/5+o(1)} \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} \left( \frac{1}{n^{4/5}/2} f_j^T D^{\dagger/2} b_{xy} \right)^2 + n^{4/5+o(1)} \left( \tilde{d}_i^T D^\dagger \tilde{d}_i + \tilde{d}_{i+1}^T D^\dagger \tilde{d}_{i+1} \right) \\
&= n^{o(1)} b_{xy}^T D^{\dagger/2} \left( \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} f_j f_j^T \right) D^{\dagger/2} b_{xy} + n^{4/5+o(1)} \left( \tilde{d}_i^T D^\dagger \tilde{d}_i + \tilde{d}_{i+1}^T D^\dagger \tilde{d}_{i+1} \right).
\end{aligned}$$

Notice that the first term in the above is at most  $\frac{n^{o(1)} \psi k^2}{n^2} \leq \frac{n^{o(1)} k^4}{n^2}$  by the expander property and  $s_{xy} = 1$ . The second term for each  $0 < i < n^{4/5}/2 - 1$  is at most

$$n^{8/5+o(1)} \left( \|F(x) D_{xx}^{\dagger/2} \phi_{xy}\|_2^2 + \|F(y) D_{yy}^{\dagger/2} \phi_{xy}\|_2^2 \right) \leq \frac{k^2}{n^{2/5-o(1)}},$$

where the inequality follows from  $\gamma$ -regularity and that  $\lambda_{\max}(F(u)^T F(u)) \leq \frac{16k \log n}{n}$ . For  $i = 0$  or  $i = n^{4/5}/2 - 1$ , the second term is at most

$$\begin{aligned}
& n^{4/5+o(1)} \left( \|D_{xx}^{\dagger/2} \phi_{xy}\|_2^2 + \|D_{yy}^{\dagger/2} \phi_{xy}\|_2^2 \right) + \\
& n^{8/5+o(1)} \left( \|F(x) D_{xx}^{\dagger/2} \phi_{xy}\|_2^2 + \|F(y) D_{yy}^{\dagger/2} \phi_{xy}\|_2^2 \right) \leq \frac{k}{n^{1/5-o(1)}}.
\end{aligned}$$

Finally by letting  $\mathbf{f} = \sum_{0 \leq i < n^{4/5}/2} \mathbf{f}_i$ , we have  $\|\mathbf{f}\|_2^2 \leq k^2 n^{2/5+o(1)}$  as desired.  $\square$

### 13 A lower bound for weighted spanner computation

In this section, we prove a superlinear lower bound of  $n^{1+\alpha-o(1)}$  for computing an  $O(n^{\frac{2}{3}(1-\alpha)})$ -spanner for any constant  $\alpha \in (0, 1/10)$ .

**Theorem 1.8** (Lower bound for weighted spanner computation). *For any constant  $\alpha \in (0, 1/10)$ , there exist constants  $C \geq 1, \delta \in (0, 1)$  such that any incidence sketch of  $N$  measurements that computes an  $o(n^{\frac{2}{3}(1-\alpha)})$ -spanner with probability  $\geq 1 - \delta$  on any  $w$  with  $\frac{w_{\max}}{w_{\min}} \leq C$  must satisfy  $N \geq n^{1+\alpha-o(1)}$ .*

Our proof will be very similar to that of our lower bound for spectral sparsifiers in Sections 3 and 12. In our proof, we will also use the block cycle graph as our hard instance as in Section 3.1. However, we make the following modifications.

1. We let the number of blocks be  $\ell = n^{\frac{2}{3}(1-\alpha)}$ , and the number of vertices in each block be  $s = n^{\frac{1}{3} + \frac{2}{3}\alpha}$ .
2. We draw the weights of all non-crossing edges independently from  $\mathcal{N}(8 \log n, 1)$ . and the weight of the crossing edge from  $\mathcal{N}(0, \log^2 n)$ .

Similarly, if the crossing edge has negative weight, we say the input is invalid, and accept any sketch as a valid sketch.

Similar to Propositions 3.1 and 3.2, we have:

**Proposition 13.1.** *With probability at least  $1 - 1/n$ , all non-crossing edges have weights in the range  $[0.1 \log n, 10 \log n]$ . If the crossing edge is present, then its weight is between  $[0.1 \log n, 10 \log n]$  with probability at least 0.41.*

**Proposition 13.2.** *Any linear sketch that can compute, with probability 0.9, a  $\frac{1}{1000}n^{\frac{2}{3}(1-\alpha)}$ -spanner in a graph where the edge weights are within a factor of 100 of each other, can distinguish between the Yes and No distributions with probability at least 0.6.*

We will prove the following theorem, which by Proposition 3.3 implies Theorem 1.8. Recall that  $B^w \in \mathbb{R}^{\binom{n}{2} \times n}$  is the weighted signed edge-vertex incidence matrix of the input graph, as defined in Section 3.2.

**Theorem 13.3.** *For any fixed sketching matrix  $\Phi \in \mathbb{R}^{k \times \binom{n}{2}}$  where  $k \leq n^{\alpha-\epsilon}$  for some constant  $\epsilon > 0$ , we have*

$$\mathbb{E}_\pi [d_{\text{TV}}((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}})] \leq o(1).$$

We also have the following proposition similar to Proposition 3.4.

**Proposition 13.4.** *For any permutation  $\pi$  such that  $b_{\pi(1)\pi(n/2+1)}$  is in the range of  $L_\pi$ ,*

$$d_{\text{TV}}((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}}) \leq O(1) \cdot \min \left\{ 1, (\log^2 n) b_{\pi(1)\pi(n/2+1)} L_\pi^\dagger b_{\pi(1)\pi(n/2+1)} \right\},$$

where  $L_\pi = \sum_{\text{non-crossing } (u,v)} b_{uv} b_{uv}^T$ , and  $b_{uv}$ 's are defined in (1).

*Proof of Theorem 13.3.* We can assume w.l.o.g. that the number of  $(u,v) \in E$  with  $\phi_{uv} \neq 0$  is  $\Omega(n^2)$ , as otherwise we would already have

$$\mathbb{E}_\pi [d_{\text{TV}}((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}})] \leq O(1) \cdot \Pr_\pi [\phi_{\pi(1)\pi(n/2+1)} \neq 0] \leq o(1).$$

Then we define the  $k \times k$  matrix-weighted graph  $H_\phi = (V, E_\phi)$ , where  $E_\phi$  consists of all edges  $(u,v)$  whose  $\phi_{uv} \neq 0$  (including those not present in the input graph), and each edge  $(u,v)$  has matrix weight  $\phi_{uv} \phi_{uv}^T$ . Therefore by the assumption above we have  $|E_\phi| \geq \Omega(n^2)$ .

By Theorem 3.13, there exists a scaling  $s : E_\phi \rightarrow [0, 1]$  s.t.  $H_\phi^s$  is a  $(\gamma, \zeta, \psi)$ -almost regular expander, where  $(\gamma, \zeta, \psi) = (8 \log n, 1/\log n, 16k^2 \log^3 n)$ , and  $|(u,v) \in E_\phi : s_{uv} < 1| \leq o(n^2)$ .

Let  $\lambda_1 \leq \dots \leq \lambda_{nk}$  be the eigenvalues of the normalized Laplacian matrix of  $H_\phi^s$ , and let  $f_1, \dots, f_{nk} \in \mathbb{R}^{nk}$  be a corresponding set of orthonormal eigenvectors.

For  $i = 0, \dots, n^{\frac{2}{3}(1-\alpha)} - 1$ , let  $B_i$  denote the vertices in the  $i^{\text{th}}$  block of the input graph:

$$B_i \stackrel{\text{def}}{=} \left\{ \pi(n^{\frac{1}{3}+\frac{2}{3}\alpha}i + 1), \dots, \pi(n^{\frac{1}{3}+\frac{2}{3}\alpha}i + n^{\frac{1}{3}+\frac{2}{3}\alpha}) \right\}.$$

Since  $k \leq n^{\alpha-\epsilon}$  and  $\alpha < 1/10$ , we have  $2 \cdot 10^6 \gamma \psi \zeta^{-1} k^2 n^\epsilon \leq n^{\frac{1}{3}+\frac{2}{3}\alpha}$ . Thus by invoking Theorem 3.14, with probability  $1 - 1/n^4$ , all vertex-induced subgraphs  $H_\phi^s[B_i \cup B_{i+1}]$  preserve  $H_\phi^s$  in the sense of (7). Using this fact, we prove the following claim:

**Claim 13.5.** For each edge  $(x, y)$  with  $s_{xy} = 1$ , conditioned on  $(x, y)$  being the crossing edge (i.e.  $\{\pi(1), \pi(n/2 + 1)\} = \{x, y\}$ ), with probability at least  $1 - 1/n^2$  over  $\pi$ , we have that  $b_{xy}$  is in the range of  $L_\pi$ , and

$$b_{xy}^T L_\pi^\dagger b_{xy} \leq \frac{k^2}{n^{2\alpha-o(1)}}. \quad (88)$$

Once again by  $k \leq n^{\alpha-\epsilon}$ , the RHS of (88) is  $o(1)$ . Since  $s_{uv} = 1$  holds for a  $1 - o(1)$  fraction of the edges, we have, by Proposition 13.4,  $\mathbb{E}_\pi [d_{\text{TV}}((\Phi B^w)_{\pi,\text{yes}}, (\Phi B^w)_{\pi,\text{no}})] = o(1)$ .  $\square$

*Proof of Claim 13.5.* Consider drawing  $\pi$  from the conditional distribution on  $\pi(1) = x$  and  $\pi(n/2 + 1) = y$ . Suppose w.l.o.g.  $(x, y)$  is oriented  $x \rightarrow y$ , so  $(b_{xy})_x = \phi_{xy}$  and  $(b_{xy})_y = -\phi_{xy}$ . Let  $E_\pi$  be the set of edges of  $\mathcal{H}_\pi$ , where we let  $\mathcal{H}_\pi$  consist of the non-crossing edges, such that each edge  $(u, v)$  has matrix weight  $\phi_{uv} \phi_{uv}^T \in \mathbb{R}^{k \times k}$ . Let us call a function  $\mathbf{f} : E_\pi \rightarrow \mathbb{R}$  a *flow* in the graph  $\mathcal{H}_\pi$ , and say  $\mathbf{f}$  routes a *demand*  $d \in \mathbb{R}^{nk}$  if  $\sum_{e \in E_\pi} \mathbf{f}_e b_e = d$ .

By Fact 4.3, to prove (88), it suffices to show that there exists a flow  $\mathbf{f}$  in  $\mathcal{H}_\pi$  that routes demand  $b_{xy}$  satisfying  $\|\mathbf{f}\|_2^2 \leq \frac{k^2}{n^{2\alpha-o(1)}}$ . To construct such a flow, we consider first breaking the demand  $b_{xy}$  into demands  $d_0, d_1, \dots, d_{n^{\frac{2}{3}(1-\alpha)}/2-1} \in \mathbb{R}^{nk}$  such that

1.  $\sum_{0 \leq i < n^{\frac{2}{3}(1-\alpha)}/2} d_i = b_{xy}$ .
2.  $d_i$  is supported on  $B_i \cup B_{i+1}$ .

We will then route each  $d_i$  in the vertex induced subgraph  $H_\phi[B_i \cup B_{i+1}]$  using a flow  $\mathbf{f}_i$ , and finally obtain a flow  $\mathbf{f} = \sum_{0 \leq i < n^{\frac{2}{3}(1-\alpha)}/2} \mathbf{f}_i$  that routes  $b_{xy}$  and satisfies  $\|\mathbf{f}\|_2^2 \leq 2 \sum_{0 \leq i < n^{\frac{2}{3}(1-\alpha)}/2} \|\mathbf{f}_i\|_2^2$ .

To show how these  $d_i$ 's are constructed, let us first define vectors  $\tilde{d}_0, \dots, \tilde{d}_{n^{\frac{2}{3}(1-\alpha)}/2} \in \mathbb{R}^{nk}$  where  $\tilde{d}_i$  is supported on  $B_i$ . We will later define each  $d_i$  by letting  $d_i = \tilde{d}_i - \tilde{d}_{i+1}$ . Let  $D \in \mathbb{R}^{nk \times nk}$  be the degree matrix of  $H_\phi^s$ . For now we would like  $\tilde{d}_i$  to satisfy

$$\forall j : 0 \leq \lambda_j \leq \zeta, \quad f_j^T D^{\dagger/2} \tilde{d}_i = (1 - \mu_i)(f_j)_x^T D_{xx}^{\dagger/2} \phi_{xy} + \mu_i(f_j)_y^T D_{yy}^{\dagger/2} \phi_{xy}. \quad (89)$$

where we define

$$\mu_i := \frac{i}{n^{\frac{2}{3}(1-\alpha)}/2}.$$

In particular, we would like  $f_j^T D^{\dagger/2} \tilde{d}_0 = (f_j)_x^T D_{xx}^{\dagger/2} \phi_{xy}$  and  $f_j^T D^{\dagger/2} \tilde{d}_{n^{2/3(1-\alpha)}/2} = (f_j)_y^T D_{yy}^{\dagger/2} \phi_{xy}$  for all  $j$  such that  $0 \leq \lambda_j \leq \zeta$ . Thus, we let  $\tilde{d}_0$  be such that

$$(\tilde{d}_0)_u = \begin{cases} \phi_{xy} & u = x \\ 0 & \text{otherwise} \end{cases}$$

and let  $\tilde{d}_{n^{2/3(1-\alpha)}/2}$  be such that

$$(\tilde{d}_{n^{2/3(1-\alpha)}/2})_u = \begin{cases} \phi_{xy} & u = y \\ 0 & \text{otherwise.} \end{cases}$$

We would then like to construct the remaining  $\tilde{d}_i$ 's such that each  $\|D^{\dagger/2} \tilde{d}_i\|_2$  is small. First let us write the linear equations (89) in matrix form. Let  $\ell$  be s.t.  $\lambda_\ell \leq \zeta < \lambda_{\ell+1}$ . Define a function  $F : V \rightarrow \mathbb{R}^{\ell \times k}$  such that

$$F(u) = \begin{pmatrix} (f_1)_u^T \\ \vdots \\ (f_\ell)_u^T \end{pmatrix}$$

and define another function  $G : \{B_0, \dots, B_{n^{2/3(1-\alpha)}/2}\} \rightarrow \mathbb{R}^{\ell \times n^{1/3 + 2/3\alpha} k}$  such that

$$G(B_i) = \begin{pmatrix} (f_1)_{B_i}^T \\ \vdots \\ (f_\ell)_{B_i}^T \end{pmatrix}$$

Then we can write (89) as

$$G(B_i) D_{B_i B_i}^{\dagger/2} (\tilde{d}_i)_{B_i} = (1 - \mu_i) F(x) D_{xx}^{\dagger/2} \phi_{xy} + \mu_i F(y) D_{yy}^{\dagger/2} \phi_{xy}. \quad (90)$$

Let us write the RHS as  $r_i := (1 - \mu_i) F(x) D_{xx}^{\dagger/2} \phi_{xy} + \mu_i F(y) D_{yy}^{\dagger/2} \phi_{xy}$ . Then by Fact 4.3, we can find a  $(\tilde{d}_i)_{B_i}$  satisfying (90) such that  $\|D_{B_i B_i}^{\dagger/2} (\tilde{d}_i)_{B_i}\|_2^2 = r_i^T (G(B_i) G(B_i)^T)^{-1} r_i$ . Now notice

$$G(B_i) G(B_i)^T = \sum_{u \in B_i} F(u) F(u)^T.$$

Since  $f_1, \dots, f_\ell$  are orthonormal, we have  $\sum_{u \in V} F(u) F(u)^T = I_{\ell \times \ell}$ . By Lemma 8.2, we have for all  $u \in V$  that  $\lambda_{\max}(F(u) F(u)^T) \leq \frac{\gamma k}{(1-\zeta)^2 n} \leq \frac{16k \log n}{n}$ . Since the vertices in  $B_i$  are chosen uniformly at random, by a Matrix Chernoff bound (Theorem 4.2), we have with probability  $\geq 1 - n^{-5}$  that  $\frac{1}{2} I \preceq n^{2/3(1-\alpha)} G(B_i) G(B_i)^T \preceq 2I$ , which implies  $r_i^T (G(B_i) G(B_i)^T)^{-1} r_i \leq 2n^{2/3(1-\alpha)} \|r_i\|^2$ . As a result, we can find  $\tilde{d}_1, \dots, \tilde{d}_{n^{2/3(1-\alpha)}/2-1}$  such that

$$\|D^{\dagger/2} \tilde{d}_i\|_2^2 \leq 4n^{2/3(1-\alpha)} \left( \|F(x) D_{xx}^{\dagger/2} \phi_{xy}\|_2^2 + \|F(y) D_{yy}^{\dagger/2} \phi_{xy}\|_2^2 \right).$$

For simplicity let  $C_i := B_i \cup B_{i+1}$ . We then construct each  $d_i$  for  $0 \leq i < n^{\frac{2}{3}(1-\alpha)}/2$  by letting  $d_i := \tilde{d}_i - \tilde{d}_{i+1}$ . Thus  $d_i$  is indeed supported on  $C_i$ , and we have  $\sum_{0 \leq i < n^{\frac{2}{3}(1-\alpha)}/2} d_i = b_{xy}$ . Now we consider how to route each  $d_i$  in  $H_\phi[B_i \cup B_{i+1}]$ . Once again by Fact 4.3, there is a flow  $\mathbf{f}_i$  with  $\ell_2^2$ -norm  $(d_i)_{C_i}^T L_{H_\phi[C_i]}^\dagger (d_i)_{C_i} \leq (d_i)_{C_i}^T L_{H_\phi^s[C_i]}^\dagger (d_i)_{C_i}$ . We then write

$$\begin{aligned}
& (d_i)_{C_i}^T L_{H_\phi^s[C_i]}^\dagger (d_i)_{C_i} \\
&= (d_i)_{C_i}^T D_{C_i C_i}^{\dagger/2} \left( D_{C_i C_i}^{\dagger/2} L_{H_\phi^s[C_i]} D_{C_i C_i}^{\dagger/2} \right)^\dagger D_{C_i C_i}^{\dagger/2} (d_i)_{C_i} \\
&\leq n^{o(1)} (d_i)_{C_i}^T D_{C_i C_i}^{\dagger/2} \left( n^{\frac{4}{3}(1-\alpha)} \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} (f_j)_{C_i} (f_j)_{C_i}^T + n^{\frac{2}{3}(1-\alpha)} (\log n) I \right) D_{C_i C_i}^{\dagger/2} (d_i)_{C_i} \\
&\leq n^{\frac{4}{3}(1-\alpha)+o(1)} \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} \left( f_j^T D^{\dagger/2} (\tilde{d}_i - \tilde{d}_{i+1}) \right)^2 + n^{\frac{2}{3}(1-\alpha)+o(1)} d_i^T D^\dagger d_i \\
&= n^{\frac{4}{3}(1-\alpha)+o(1)} \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} \left( \frac{1}{n^{\frac{2}{3}(1-\alpha)}/2} f_j^T D^{\dagger/2} b_{xy} \right)^2 + n^{\frac{2}{3}(1-\alpha)+o(1)} \left( \tilde{d}_i^T D^\dagger \tilde{d}_i + \tilde{d}_{i+1}^T D^\dagger \tilde{d}_{i+1} \right) \\
&= n^{o(1)} b_{xy}^T D^{\dagger/2} \left( \sum_{j: \lambda_j \in (0, \zeta]} \frac{1}{\lambda_j} f_j f_j^T \right) D^{\dagger/2} b_{xy} + n^{\frac{2}{3}(1-\alpha)+o(1)} \left( \tilde{d}_i^T D^\dagger \tilde{d}_i + \tilde{d}_{i+1}^T D^\dagger \tilde{d}_{i+1} \right).
\end{aligned}$$

Notice that the first term in the above is at most  $\frac{n^{o(1)} \psi k^2}{n^2} \leq \frac{n^{o(1)} k^4}{n^2}$  by the expander property and  $s_{xy} = 1$ . The second term for each  $0 < i < n^{\frac{2}{3}(1-\alpha)}/2 - 1$  is at most

$$n^{\frac{4}{3}(1-\alpha)+o(1)} \left( \|F(x) D_{xx}^{\dagger/2} \phi_{xy}\|_2^2 + \|F(y) D_{yy}^{\dagger/2} \phi_{xy}\|_2^2 \right) \leq \frac{k^2 n^{\frac{4}{3}(1-\alpha)+o(1)}}{n^2},$$

where the inequality follows from  $\gamma$ -regularity and that  $\lambda_{\max}(F(u)^T F(u)) \leq \frac{16k \log n}{n}$ . For  $i = 0$  or  $i = n^{\frac{2}{3}(1-\alpha)}/2 - 1$ , the second term is at most

$$\begin{aligned}
& n^{\frac{2}{3}(1-\alpha)+o(1)} \left( \|D_{xx}^{\dagger/2} \phi_{xy}\|_2^2 + \|D_{yy}^{\dagger/2} \phi_{xy}\|_2^2 \right) + \\
& n^{\frac{4}{3}(1-\alpha)+o(1)} \left( \|F(x) D_{xx}^{\dagger/2} \phi_{xy}\|_2^2 + \|F(y) D_{yy}^{\dagger/2} \phi_{xy}\|_2^2 \right) \leq \frac{k n^{\frac{2}{3}(1-\alpha)+o(1)}}{n}.
\end{aligned}$$

Finally by letting  $\mathbf{f} = \sum_{0 \leq i < n^{\frac{2}{3}(1-\alpha)}/2} \mathbf{f}_i$ , we have  $\|\mathbf{f}\|_2^2 \leq k^2 n^{-2\alpha+o(1)}$  as desired.  $\square$

## A Proof of Proposition 1.6

*Proof of Proposition 1.6.* Consider generating a weighted graph as follows. Let  $0 < a < b$  be some given parameters. First, we pick a uniformly random edge slot  $e \in \binom{V}{2}$ . Then we let the weight of each edge  $f \neq e$  be an independent  $\mathcal{N}(b, b^2/(100 \log n))$ , a Gaussian distribution with mean  $b$  and variance  $b^2/(100 \log n)$ . Finally, we include edge  $e$  in the graph with probability  $1/2$ ,

and if we do include it, we draw its weight from  $\mathcal{N}(0, a^2)$ . If  $e$  is present has negative weight, we say the input is *invalid*, and accept any sketch as a valid sketch. We will call the conditional distribution on the presence of  $e$  the **Yes** distribution, and the conditional distribution on the absence of  $e$  the **No** distribution.

We then show that any linear sketch that computes an  $o(w_{\max}/w_{\min})$ -spanner must detect the presence or absence of  $e$  with the edge slot  $e$  with good probability.

**Claim A.1.** Any linear sketch that can compute an  $o(w_{\max}/w_{\min})$ -spanner with probability .9 can distinguish between the **Yes** and **No** distributions with probability .6.

*Proof.* By Chernoff bounds (Theorem 4.7), with probability  $1 - 1/n$ , all edges other than  $e$  have weights in  $[b/100, 100b]$ . If we do add  $e$ , it has weight  $\geq a/100$  with probability at least .4, by standard properties of Gaussian distributions. Then conditioned on both of these happening, one can detect  $e$  by looking at an  $o(w_{\max}/w_{\min}) = o(b/a)$  spanner of the graph. Therefore, if a linear sketch can compute an  $o(w_{\max}/w_{\min})$ -spanner with probability 0.9, it can detect if  $e$  is present with probability at least

$$\geq (1 - 1/n)(.5 + .5 \cdot .4).9 \geq .6.$$

□

Suppose for the sake of contradiction, there is a sketching matrix  $\Phi \in \mathbb{R}^{k \times \binom{n}{2}}$  where  $\phi_e$  is the  $e^{\text{th}}$  column of  $\Phi$  and  $k = o(n^2)$ , such that one can recover from  $\Phi w$  an  $o(w_{\max}/w_{\min})$ -spanner with probability .9. Let  $(\Phi w)_{\text{yes}}$  and  $(\Phi w)_{\text{no}}$  be the sketches obtained conditioned on the presence/absence of  $e$ . By Theorem 4.6, conditioned on a fixed choice of the edge slot  $e$ , if  $\phi_e$  is in the linear span of  $\{\phi_f : f \neq e\}$ , then the total variation distance between  $(\Phi w)_{\text{yes}}$  and  $(\Phi w)_{\text{no}}$  is bounded by

$$D_{\text{TV}}((\Phi w)_{\text{yes}}|e, (\Phi w)_{\text{no}}|e) \leq O(1) \cdot a^2 \phi_e^T \left( \sum_{e \neq f \in \binom{V}{2}} \frac{b^2}{\log n} \phi_f \phi_f^T \right)^{\dagger} \phi_e,$$

Since  $k = o(n^2)$ , for at least  $(1 - o(1))\binom{n}{2}$  edge slots  $e$ , we have  $\phi_e^T \left( \sum_{f \in \binom{V}{2}} \phi_f \phi_f^T \right)^{\dagger} \phi_e \leq o(1)$ , which implies that  $\phi_e$  is in the linear span of  $\{\phi_f : f \neq e\}$ . Let  $E^*$  with  $|E| \geq (1 - o(1))\binom{n}{2}$

denote the set of such edges. Thus we have

$$\begin{aligned}
D_{\text{TV}}((\Phi w)_{\text{yes}}, (\Phi w)_{\text{no}}) &\leq \frac{1}{\binom{n}{2}} \left( \sum_{e \in E^*} D_{\text{TV}}((\Phi w)_{\text{yes}}|e, (\Phi w)_{\text{no}}|e) + \sum_{e \notin E^*} 1 \right) \\
&\leq O(1/n^2) \cdot a^2 \sum_{e \in E^*} \phi_e^T \left( \sum_{f \in \binom{V}{2}} \frac{b^2}{\log n} \phi_f \phi_f^T \right)^\dagger \phi_e + o(1) \\
&\leq O(\log n/n^2) \frac{a^2}{b^2} \cdot \text{Tr} \left( \left( \sum_{e \in \binom{V}{2}} \phi_e \phi_e^T \right) \left( \sum_{f \in \binom{V}{2}} \phi_f \phi_f^T \right)^\dagger \right) + o(1) \\
&= O(\log n/n^2) \frac{a^2}{b^2} \text{rank} \left( \sum_{e \in \binom{V}{2}} \phi_e \phi_e^T \right) + o(1) \\
&\leq O\left(\frac{ka^2 \log n}{b^2 n^2}\right) + o(1).
\end{aligned}$$

Thus, whenever  $b^2 \geq a^2 \log n$ , this total variation distance is  $o(1)$ , contradicting Claim A.1.  $\square$

## B A hard instance for decomposing matrix-weighted graphs into large subgraphs without small nonzero eigenvalues

The proof of the following proposition appears in Appendix D (in “Proof of Proposition 7.4”).

**Proposition B.1.** *For a vector  $x \in \mathbb{R}^{nk}$ , we have*

$$x^T L x = \sum_{u \sim v} (x_u - x_v)^T \phi_{uv} \phi_{uv}^T (x_u - x_v).$$

By the above proposition, we have:

**Proposition B.2.** *For any  $nk$ -dimensional vector  $x$  such that  $x_u = x_v \in \mathbb{R}^k$  for any  $u, v \in V$ , we have  $x^T L x = 0$ . As a result, the null space of  $L$  has dimension at least  $k$ .*

First consider the following  $n$ -vertex  $2 \times 2$  matrix-weighted graph, call it  $R$ , whose vertices can be embedded into points that are uniformly distributed along a unit 2D-circle. Specifically, for each vertex  $u \in \{1..n\}$ , assign to it a coordinate

$$c_u \stackrel{\text{def}}{=} (\cos \frac{2u\pi}{n}, \sin \frac{2u\pi}{n})^T.$$

Then for every  $u \neq v$ , connect them by an edge weighted by (the outer product of) the segment connecting  $c_u, c_v$ . That is, add an edge  $(u, v)$  with weight  $(c_u - c_v)(c_u - c_v)^T$ . Then one can see that the Laplacian matrix (and thus also the normalized Laplacian matrix) of graph  $R$  has three zero eigenvalues — apart from the two trivial zero eigenvectors mentioned in Proposition B.2, one can also construct another zero eigenvector  $x$  where  $x_u$  equals  $c_u$  rotated by 90 degrees

counter-clockwise around the origin, namely  $x_u = (-\sin \frac{2u\pi}{n}, \cos \frac{2u\pi}{n})^T$ . Then we have for any  $u \neq v$  that  $x_u - x_v$  is orthogonal to  $\phi_{uv}$ , and hence

$$x^T L_R x = \sum_{\{u,v\}} (x_u - x_v)^T \phi_{uv} \phi_{uv}^T (x_u - x_v) = 0.$$

Now consider creating a noisy version of the graph  $R$ , call it  $\mathcal{R}$ , as follows. Between every pair of vertices  $u, v$ , we add a small stochastic noise on the weight of the edge connecting them. Specifically, let  $\tilde{\phi}_{uv}$ 's be drawn independently from the 2-dimensional Gaussian  $\mathcal{N}(0, \epsilon I)$ , for some sufficiently small  $\epsilon$  (say  $o(1/\text{poly}(n))$ ). Then we let the edge weight of  $(u, v)$  be

$$(\phi_{uv} + \tilde{\phi}_{uv})(\phi_{uv} + \tilde{\phi}_{uv})^T.$$

Now for the vector  $x$  with  $x_u = (-\sin \frac{2u\pi}{n}, \cos \frac{2u\pi}{n})^T$ , we have

$$\begin{aligned} \mathbb{E}_{\tilde{\phi}_{uv}} [x^T L_{\mathcal{R}} x] &= \sum_{\{u,v\}} \mathbb{E}_{\tilde{\phi}_{uv}} \left[ (x_u - x_v)^T (\phi_{uv} + \tilde{\phi}_{uv})(\phi_{uv} + \tilde{\phi}_{uv})^T (x_u - x_v) \right] \\ &= \sum_{\{u,v\}} \mathbb{E}_{\tilde{\phi}_{uv}} \left[ (x_u - x_v)^T \tilde{\phi}_{uv} \tilde{\phi}_{uv}^T (x_u - x_v) \right] = \Theta(\epsilon n^2). \end{aligned}$$

Also note that  $\|x\|^2 = n$ , and  $x$  is orthogonal to the two trivial zero eigenvectors of  $L_{\mathcal{R}}$  as  $\sum_u x_u = 0$ . Therefore  $L_{\mathcal{R}}$ 's smallest nonzero eigenvalue is at most  $o(1/\text{poly}(n))$ . One can verify that in  $\mathcal{R}$  the degree of each vertex  $u$  has a constant ratio between its largest and smallest eigenvalues, and as a result the normalized Laplacian  $N_{\mathcal{R}}$  also has a nonzero eigenvalue bounded by  $o(1/\text{poly}(n))$ . One can similarly show the same property for every large enough subgraph of  $\mathcal{R}$  (say, e.g., with  $\Omega(n^2)$  edges).

## C Missing proofs from Section 3

*Proof of proposition 3.2.* By Proposition 3.1, with probability  $1 - 1/n$ , the effective resistance between  $\pi(1)$  and  $\pi(n/2 + 1)$  is at least  $1/48$ . Thus whenever the crossing edge has positive weight at least  $\frac{1}{4}$ , one can tell if the crossing edge is present or not by looking at a 1.0001-spectral sparsifier of the graph. On the other hand, if we do add the crossing edge, it has weight  $\geq \frac{1}{4}$  with probability at least 0.4, by properties of a standard Gaussian. Therefore if a linear sketch can compute a 1.0001-spectral sparsifier with probability 0.9, then it can decide if the crossing edge is present with probability

$$\geq (1 - 1/n)(0.5 + 0.5 \cdot 0.4)0.9 \geq 0.6.$$

□

*Proof of Proposition 3.3.* For the  $i^{\text{th}}$  row of the sketching matrix  $\Phi$ , if its support is on edges incident on vertex  $u$ , we say the  $i^{\text{th}}$  row belongs to  $u$ . Let  $U \subseteq V$  be the subset of vertices  $u$  such that the number of rows of  $\Phi$  that belong to  $u$  is at most  $\frac{1000N}{n}$ . By Markov's inequality we have  $|U| \geq 0.999n$ . Let  $F = \binom{U}{2}$  be the edge slots within  $U$ . Then  $|F| \geq 0.998 \binom{n}{2}$ .

Let  $(\Phi w)_{\pi,\text{yes}}$  be  $\Phi w$  conditioned on  $\pi$  and the presence of the crossing edge; similarly let  $(\Phi w)_{\pi,\text{no}}$  be  $\Phi w$  conditioned on  $\pi$  and the absence of the crossing edge. Let  $\phi_{uv} \in \mathbb{R}^N$  be the column of  $\Phi$  corresponding to edge slot  $(u, v)$ . By Theorem 4.6, for any  $\pi$  we have

$$d_{\text{TV}}((\Phi w)_{\pi,\text{yes}}, (\Phi w)_{\pi,\text{no}}) \leq O(1) \cdot \min \left\{ 1, \phi_{\pi(1)\pi(n/2+1)}^T \left( \sum_{\text{non-crossing } (u,v)} n^{4/5} \log^{-1} n \phi_{uv} \phi_{uv}^T \right)^\dagger \phi_{\pi(1)\pi(n/2+1)} \right\}. \quad (91)$$

Note that we have slightly abused the notation  $\dagger$  in the second term on the RHS of (91): in the case that  $\phi_{\pi(1)\pi(n/2+1)}$  is not in the range of the matrix inside the summation, the term should be understood as infinity.

By restricting the summation on the RHS to non-crossing edges in  $F$ , we have, by Fact 4.3, for any  $\pi$

$$\phi_{\pi(1)\pi(n/2+1)}^T \left( \sum_{\text{non-crossing } (u,v)} n^{4/5} \log^{-1} n \phi_{uv} \phi_{uv}^T \right)^\dagger \phi_{\pi(1)\pi(n/2+1)} \leq \phi_{\pi(1)\pi(n/2+1)}^T \left( \sum_{\text{non-crossing } (u,v) \in F} n^{4/5} \log^{-1} n \phi_{uv} \phi_{uv}^T \right)^\dagger \phi_{\pi(1)\pi(n/2+1)}. \quad (92)$$

Now consider a sketching matrix  $\Phi'' \in \mathbb{R}^{N \times \binom{n}{2}}$  that is obtained from  $\Phi$  by zeroing out the columns corresponding to edge slots *not* in  $F$ . Then we have by (92) that for any  $\pi$  such that  $(\pi(1), \pi(n/2 + 1)) \in F$  (thus the column of  $\Phi''$  corresponding to the crossing edge is *not* zeroed out)

$$d_{\text{TV}}((\Phi w)_{\pi,\text{yes}}, (\Phi w)_{\pi,\text{no}}) \leq d_{\text{TV}}((\Phi'' w)_{\pi,\text{yes}}, (\Phi'' w)_{\pi,\text{no}}).$$

Using the fact that  $\Pr_\pi[(\pi(1), \pi(n/2 + 1)) \in F] = |F|/\binom{n}{2} \geq 0.998$ , we have

$$\mathbb{E}_\pi [d_{\text{TV}}((\Phi'' w)_{\pi,\text{yes}}, (\Phi'' w)_{\pi,\text{no}})] \geq \mathbb{E}_\pi [d_{\text{TV}}((\Phi w)_{\pi,\text{yes}}, (\Phi w)_{\pi,\text{no}})] - 0.002,$$

and as a result, there exists an incidence sketch using  $\Phi''$  that can distinguish between the Yes and No distributions with probability 0.55. We then show that we can simulate such an incidence sketch that uses  $\Phi''$  by a signed sketch with a sketching matrix  $\Phi' \in \mathbb{R}^{k \times \binom{n}{2}}$  such that  $k \leq O(1) \cdot \max \left\{ 1, \frac{N \log n}{n} \right\}$ .

Let  $R_u \subseteq \{1, \dots, N\}$  denote the indices of the rows that belong to  $u$ . Then we know  $|R_u| \leq 1000N/n$  for all  $u \in U$ . It suffices to construct a  $\Phi'$  from which we can recover, for each vertex  $u \in U$ , the sketch  $\sum_{v \in U \setminus \{u\}} w_{uv} \cdot (\phi_{uv})_{R_u} \in \mathbb{R}^{|R_u|}$ . Now consider sketching matrices  $\Phi'_1, \dots, \Phi'_{10 \log n} \in \mathbb{R}^{\frac{1000N}{n} \times \binom{n}{2}}$ , which are initially set to be all-zero. We then assign values to these matrices as follows. For each  $i = 1, \dots, 10 \log n$ , we generate a subset of vertices  $S_i$  by including each vertex independently in  $S_i$  with probability  $1/2$ . By Chernoff bound, we know that with high probability, we have that for every vertex  $u$  and every incident edge  $(u, v)$ ,  $u \in S_i, v \in V - S_i$

holds for at least one  $i$ . Let  $i_{uv}$  be the smallest  $i$  such that  $u \in S_i, v \in V - S_i$ . Then for each vertex  $u$  and each incident edge  $(u, v) \in F$ , we let the column of  $\Phi'_{i_{uv}}$  corresponding to  $(u, v)$  be

$$\Phi'_{i_{uv}}(:, (u, v)) = \begin{cases} (\phi_{uv})_{R_u} & u \text{ is } (u, v)'s \text{ head} \\ -(\phi_{uv})_{R_u} & u \text{ is } (u, v)'s \text{ tail}, \end{cases}$$

where we augment  $(\phi_{uv})_{R_u}$  by 0's if  $|R_u| < \frac{1000N}{n}$ . Then for each vertex  $u \in S_i$ , we have

$$(\Phi'_i B^w)(:, u) = \sum_{v \in U : i_{uv} = i} w_{uv} \cdot (\phi_{uv})_{R_u},$$

and therefore

$$\sum_{i=1}^{10 \log n} (\Phi'_i B^w)(:, u) = \sum_{v \in U \setminus \{u\}} w_{uv} \cdot (\phi_{uv})_{R_u}.$$

This implies that we can construct a desired  $\Phi'$  by stacking up the  $\Phi'_i$ 's:

$$\Phi' = \begin{pmatrix} \Phi'_1 \\ \vdots \\ \Phi'_{10 \log n} \end{pmatrix} \in \mathbb{R}^{\frac{10000N \log n}{n} \times \binom{n}{2}}.$$

□

*Proof of Proposition 3.4.* Consider fixing a sketching matrix  $\Phi \in \mathbb{R}^{k \times \binom{n}{2}}$  and a permutation  $\pi : \{1..n\} \rightarrow \{1..n\}$ . We are interested in the total variation distance between  $(\Phi B^w)_{\pi, \text{yes}}$  and  $(\Phi B^w)_{\pi, \text{no}}$ . Let us also fix a vertex  $u \in \{1..n\}$  and consider the  $u$ -th column  $(\Phi B^w)(:, u) \in \mathbb{R}^k$  of the sketch obtained. We have (recall that  $\phi_e$  is the  $e$ -th column of  $\Phi$ )

$$(\Phi B^w)(:, u) = \sum_e B_{eu}^w \phi_e. \quad (93)$$

If  $e$  is not present in the graph, or if  $e$  is not incident on  $u$ , then the entry  $B_{eu}^w$  is zero. Otherwise ( $e$  present and incident on  $u$ ), we know that  $B_{eu}^w$  follows a univariate Gaussian distribution  $\mathcal{N}(\mu_{e,u}, \sigma_{e,u}^2)$ , where for non-crossing  $e$

$$\mu_{e,u} = \begin{cases} 8n^{2/5} & u \text{ is } e \text{'s head} \\ -8n^{2/5} & u \text{ is } e \text{'s tail} \end{cases} \quad \text{and} \quad \sigma_{e,u}^2 = n^{4/5} \log^{-1} n \quad (94)$$

and for the crossing edge  $e$

$$\mu_{e,u} = 0 \quad \text{and} \quad \sigma_{e,u}^2 = 1. \quad (95)$$

Therefore,  $B_{eu}^w \phi_e$  follows a  $k$ -dimensional Gaussian distribution:

$$B_{eu}^w \phi_e \sim \mathcal{N}(\mu_{e,u} \phi_e, \sigma_{e,u}^2 \phi_e \phi_e^T). \quad (96)$$

Also notice that the distributions of  $B_{eu}^w$ 's are independent for different edges. Thus by Fact 4.5

$$(\Phi B^w)(:, u) \sim \mathcal{N} \left( \sum_{e \sim u} \mu_{e,u} \phi_e, \sum_{e \sim u} \sigma_{e,u}^2 \phi_e \phi_e^T \right) \quad (97)$$

As a result, the sketch  $\Phi B^w$  can be seen as an  $nk$ -dimensional Gaussian.

We next consider the correlations between different columns of  $\Phi B^w$ . For two vertices  $u, v$ , if there is no edge between them, then the distributions of  $(\Phi B^w)(:, u)$  and  $(\Phi B^w)(:, v)$  are independent, so the correlation between them is zero. Otherwise (if the edge  $e = (u, v)$  is present), the only correlation between  $(\Phi B^w)(:, u)$  and  $(\Phi B^w)(:, v)$  is the one between  $B_{eu}^w \phi_e$  and  $B_{ev}^w \phi_e$ , who are negations of each other. Therefore the correlation between  $(\Phi B^w)(:, u)$  and  $(\Phi B^w)(:, v)$  is just the negative covariance of  $B_{eu}^w \phi_e$ :

$$\mathbb{E} \left[ ((\Phi B^w)(:, u) - \mathbb{E}[(\Phi B^w)(:, u)]) ((\Phi B^w)(:, v) - \mathbb{E}[(\Phi B^w)(:, v)])^T \right] = -\sigma_{e,u}^2 \phi_e \phi_e^T. \quad (98)$$

We are now ready to write down the covariance matrix of the sketch  $\Phi B^w$ . First consider the covariance matrix of  $(\Phi B^w)_{\pi, \text{no}}$ , call it  $\Sigma_{(\Phi B^w)_{\pi, \text{no}}} \in \mathbb{R}^{nk \times nk}$ . We can write  $\Sigma_{(\Phi B^w)_{\pi, \text{no}}}$  as an  $n \times n$  block matrix (with block size  $k \times k$ ) by

$$(\Sigma_{(\Phi B^w)_{\pi, \text{no}}})_{uv} = \begin{cases} \sum_{\text{non-crossing } e \sim u} n^{4/5} \log^{-1} n \phi_e \phi_e^T & u = v \\ -n^{4/5} \log^{-1} n \phi_e \phi_e^T & u \neq v \text{ and } e = (u, v) \text{ is non-crossing} \\ 0 & \text{otherwise.} \end{cases}$$

One can also verify that  $\Sigma_{(\Phi B^w)_{\pi, \text{no}}} = \sum_{\text{non-crossing}(u,v)} n^{4/5} \log^{-1} n b_{uv} b_{uv}^T$ , where  $b_{uv}$ 's are defined in (1). That is  $\Sigma_{(\Phi B^w)_{\pi, \text{no}}}$  is exactly the matrix  $L_\pi$ .

If the crossing edge is present, using the fact that its weight follows a standard Gaussian distribution, the covariance matrix can be written as

$$\Sigma_{(\Phi B^w)_{\pi, \text{yes}}} = \Sigma_{(\Phi B^w)_{\pi, \text{no}}} + b_{\pi(1)\pi(n/2+1)} b_{\pi(1)\pi(n/2+1)}^T \quad (99)$$

Notice that, since the crossing edge's weight has zero mean,  $(\Phi B^w)_{\pi, \text{yes}}$  and  $(\Phi B^w)_{\pi, \text{no}}$  have the same mean. Therefore, we can invoke Theorem 4.6 and obtain an upper bound (up to a constant factor) on the total variation distance between the two distributions as

$$O(1) \cdot \min \left\{ 1, b_{\pi(1)\pi(n/2+1)}^T \Sigma_{(\Phi B^w)_{\pi, \text{no}}}^\dagger b_{\pi(1)\pi(n/2+1)} \right\}, \quad (100)$$

as desired.  $\square$

*Proof of Claim 3.7.* Let  $n' = |U| \geq d_{\min} \geq \frac{|E_\phi|}{8n}$  be the number of vertices in  $I$ . Notice that by conditioning on an  $f \in I$  being the crossing edge, we have  $\pi(1), \pi(n/2 + 1) \in U$  both. Let  $U_i$  denote the vertices in  $U$  that are in the  $i^{\text{th}}$  block of the cycle. We have with high probability over  $\pi$  that each  $|U_i| \in [\frac{n'}{2n^{4/5}}, \frac{2n'}{n^{4/5}}]$ . By Lemma 3.6, with probability at least  $1 - 1/n^4$  over  $\pi$ , each  $I[U_i \cup U_{i+1}]$  is a  $\frac{1}{n'^{(1)}}$ -expander with minimum degree  $\geq \frac{|E_\phi|}{16n^{9/5}}$ . Now consider constructing a graph  $\tilde{I}$  supported on  $U$  by, for each  $0 \leq i < \ell$ , adding a clique of weight  $\frac{|E_\phi|}{n'n}$  on  $U_i \cup U_{i+1}$  (whose vertex degrees are all  $\Theta(\frac{|E_\phi|}{n^{9/5}})$ ). One can show that in the graph  $\tilde{I}$ , the effective resistance

between  $\pi(1), \pi(n/2 + 1)$  is maximized when we set  $n'$  to be its minimum  $\Theta(\frac{|E_\phi|}{n})$ , in which case the effective resistance is  $\Theta(n^{4/5}) \cdot \left(\frac{n^{9/5}}{|E_\phi|}\right)^2$ . Then by the expander property, the effective resistance between  $\pi(1), \pi(n/2 + 1)$  in  $H_\pi$  is at most  $\Theta(n^{4/5+o(1)}) \cdot \left(\frac{n^{9/5}}{|E_\phi|}\right)^2$ . Finally, since  $\mathcal{H}_\pi$  is  $n^{4/5} \log^{-1} n H_\pi$ , we have that the effective resistance between  $\pi(1), \pi(n/2 + 1)$  in  $\mathcal{H}_\pi$  is at most  $\Theta(n^{o(1)}) \cdot \left(\frac{n^{9/5}}{|E_\phi|}\right)^2$ , which is bounded by  $u_f$ .  $\square$

*Proof of Proposition 3.12.* Suppose w.l.o.g.  $\lambda_1, \dots, \lambda_\ell$  are all eigenvalues between  $(0, \zeta]$ . Then

$$\begin{aligned} & \sum_{(u,v) \in E} \left( D^{\dagger/2} b_{uv} \right)^T \left( \sum_{i=1}^{\ell} \lambda_i f_i f_i^T \right)^\dagger D^{\dagger/2} b_{uv} \\ &= \sum_{(u,v) \in E} \text{Tr} \left( \left( \sum_{i=1}^{\ell} \lambda_i f_i f_i^T \right)^\dagger D^{\dagger/2} b_{uv} b_{uv}^T D^{\dagger/2} \right) \\ &= \text{Tr} \left( \left( \sum_{i=1}^{\ell} \lambda_i f_i f_i^t \right)^\dagger \left( \sum_{(u,v) \in E} D^{\dagger/2} b_{uv} b_{uv}^T D^{\dagger/2} \right) \right) \\ &= \text{Tr} \left( \left( \sum_{i=1}^{\ell} \lambda_i f_i f_i^t \right)^\dagger N_G \right) \\ &= \text{Tr} \left( \left( \sum_{i=1}^{\ell} \lambda_i f_i f_i^t \right)^\dagger \left( \sum_{i=1}^n \lambda_i f_i f_i^T \right) \right) \\ &= \text{Tr} \left( \left( \sum_{i=1}^{\ell} \lambda_i f_i f_i^t \right)^\dagger \left( \sum_{i=1}^{\ell} \lambda_i f_i f_i^T \right) \right) = \ell, \end{aligned}$$

as desired.  $\square$

## D Missing proofs from Section 7

*Proof of Proposition 7.4.* The claim follows by

$$\begin{aligned} x^T L x &= x^T \left( \sum_{u \sim v} b_{uv} b_{uv}^T \right) x \\ &= \sum_{u \sim v} \langle x, b_{uv} \rangle^2 \\ &= \sum_{u \sim v} (\langle x_u, \phi_{uv} \rangle - \langle x_v, \phi_{uv} \rangle)^2 \\ &= \sum_{u \sim v} \langle x_u - x_v, \phi_{uv} \rangle^2. \end{aligned}$$

$\square$

*Proof of Proposition 7.6.* The claim follows by

$$\begin{aligned}
x^T Nx &= x^T \left( \sum_{u \sim v} (D^{\dagger/2} b_{uv})(D^{\dagger/2} b_{uv})^T \right) x \\
&= \sum_{u \sim v} \langle x, D^{\dagger/2} b_{uv} \rangle^2 \\
&= \sum_{u \sim v} \left( \langle x_u, D_u^{\dagger/2} \phi_{uv} \rangle - \langle x_v, D_v^{\dagger/2} \phi_{uv} \rangle \right)^2.
\end{aligned}$$

□

*Proof of Proposition 7.7.* The lower bound of 0 follows from that  $N$  is positive semi-definite. We then prove the upper bound by showing that for any  $x \in \mathbb{R}^{nk}$ , its Rayleigh quotient  $R(x) = \frac{x^T Nx}{x^T x}$  is at most 2. It is then equivalent to show that  $x^T Nx \leq 2 \|x\|^2$ :

$$\begin{aligned}
x^T Nx &= \sum_{u \sim v} \left( \langle x_u, D_u^{\dagger/2} \phi_{uv} \rangle - \langle x_v, D_v^{\dagger/2} \phi_{uv} \rangle \right)^2 \\
&\leq \sum_{u \sim v} \left( 2 \langle x_u, D_u^{\dagger/2} \phi_{uv} \rangle^2 + 2 \langle x_v, D_v^{\dagger/2} \phi_{uv} \rangle^2 \right) \\
&= 2 \sum_u \sum_{v \sim u} \langle x_u, D_u^{\dagger/2} \phi_{uv} \rangle^2 \\
&= 2 \sum_u \sum_{v \sim u} x_u^T D_u^{\dagger/2} \phi_{uv} \phi_{uv}^T D_u^{\dagger/2} x_u \\
&= 2 \sum_u x_u^T D_u^{\dagger/2} \left( \sum_{v \sim u} \phi_{uv} \phi_{uv}^T \right) D_u^{\dagger/2} x_u \\
&= 2 \sum_u x_u^T D_u^{\dagger/2} D_u D_u^{\dagger/2} x_u \\
&\leq 2 \sum_u x_u^T x_u = 2 \sum_u \|x_u\|^2 = 2 \|x\|^2.
\end{aligned}$$

Here the last inequality follows from that  $D_u^{\dagger/2} D_u D_u^{\dagger/2} = \Pi_{D_u} \preceq I$ . □

## E Missing proofs from Section 10

*Proof of Lemma 10.2.* Define  $Y = D^{1/2} X$ . Then

$$\min_{X \in \mathcal{X}} \frac{\det(X^T L X)}{\det(X^T D X)} = \min_{D^{\dagger/2} Y \in \mathcal{X}} \frac{\det(Y^T D^{\dagger/2} L D^{\dagger/2} Y)}{\det(Y^T Y)}. \quad (101)$$

Let the eigenvalues of  $Y^T Y$  be  $\nu_1, \dots, \nu_\ell$  and let  $g_1, \dots, g_\ell \in \mathbb{R}^\ell$  be a corresponding set of orthonormal eigenvectors. We also know that  $Y Y^T$  has eigenvalues  $\nu_1, \dots, \nu_\ell, 0, \dots, 0$ . Let  $h_1, \dots, h_{nk} \in \mathbb{R}^{nk}$  be a corresponding set of orthonormal eigenvectors of  $Y Y^T$ . Then

$$\det(Y^T Y) = \prod_{i=1}^\ell \nu_i$$

and

$$\begin{aligned}
\det \left( Y^T D^{\dagger/2} L D^{\dagger/2} Y \right) &= \det_+ \left( L^{1/2} D^{\dagger/2} Y Y^T D^{\dagger/2} L^{1/2} \right) \\
&= \det_+ \left( L^{1/2} D^{\dagger/2} \left( \sum_{i=1}^{\ell} \nu_i h_i h_i^T \right) D^{\dagger/2} L^{1/2} \right) \\
&= \det_+ \left( L^{1/2} D^{\dagger/2} H V H^T D^{\dagger/2} L^{1/2} \right) \\
&= \det \left( V^{1/2} H^T D^{\dagger/2} L D^{\dagger/2} H V^{1/2} \right) \\
&= \left( \prod_{i=1}^{\ell} \nu_i \right) \det \left( H^T D^{\dagger/2} L D^{\dagger/2} H \right),
\end{aligned}$$

where we define  $H = (h_1 \ \dots \ h_\ell) \in \mathbb{R}^{nk \times \ell}$  and  $V = \text{diag}(\nu_1, \dots, \nu_\ell) \in \mathbb{R}^{\ell \times \ell}$ . The claim in the lemma then follows by noting that  $\det(H^T D^{\dagger/2} L D^{\dagger/2} H)$  is minimized (over all  $H$ 's with orthonormal columns in the range of  $N$ ) when  $H$ 's columns are bottom nonzero eigenvectors of  $N = D^{\dagger/2} L D^{\dagger/2}$ , a result of Cauchy interlacing. Therefore (101) equals  $\det_\ell(N)$ .  $\square$

*Proof of Lemma 10.6.* Consider the following process for transitioning  $G$  to  $G^s$ , where we use  $t$  to denote the current scaling, which is all one in the beginning.

1. Initially, let  $t_e \leftarrow 1$  for all  $e \in E$ .
2. While  $t \neq s$ :
  - (a) Let  $F \leftarrow \{e : t_e > s_e\}$ , and let  $\eta \leftarrow \min_{e \in F} t_e / s_e$ .
  - (b) For each  $e \in F$ , let  $t_e \leftarrow t_e / \eta$ .

Note that this process terminates in finite time as in each loop we make  $t_e = s_e$  for at least one extra  $e$ .

**Claim E.1.** At the beginning of every while loop, we have for any  $e$  such that  $s_e < t_e$

$$R^t(e) \leq \frac{2\gamma k}{n}.$$

*Proof.* We prove this by induction. In the beginning of the process, by  $\gamma$ -regularity, we have  $R^t(e) \leq \frac{2\gamma k}{n}$  for all  $e \in E$ . Then for the induction step, note that in each while loop, we downscale the weights of *all* edges with  $s_e < t_e$  by a same amount. From the point view of leverage scores, this is equivalent to increasing the weights of all other edges by a same multiple. As a result, the leverage scores of *all* edges with  $s_e < t_e$  can only decrease (by Fact 4.3), and thus we have  $R^t(e) \leq \frac{2\gamma k}{n}$  for these edges throughout.  $\square$

**Claim E.2.** Consider fixing an iteration of the while loop and letting  $t$  and  $t'$  be the scaling before and after the execution of Line 2b respectively. Then we have

$$\det_+(D^{t'}) \geq \left( 1 - \frac{2\gamma k}{n} \right)^{2|F| \log \eta} \det_+(D^t).$$

*Proof.* For an  $x \in [1, \ln \eta]$ , define  $t^x : E \rightarrow [0, 1]$  by

$$t_e^x = \begin{cases} e^{-x} \cdot t_e & e \in F \\ t_e & \text{o.w.} \end{cases}$$

Then we have  $t' = t^{\ln \eta}$ . Consider differentiating  $\ln \det_+(D^{t^x})$ :

$$\begin{aligned} \frac{d \ln \det_+(D^{t^x})}{dx} &= \text{Tr} \left( (D^{t^x})^\dagger \frac{d D^{t^x}}{dx} \right) \\ &= -2 \text{Tr} \left( (D^{t^x})^\dagger (D^{t^x})^F \right) \\ &= -2 \text{Tr} \left( (D^{t^x})^\dagger \left( \sum_{e \in F} \left( e_{u \leftarrow v}^{t^x} (e_{u \leftarrow v}^{t^x})^T + e_{v \leftarrow u}^{t^x} (e_{v \leftarrow u}^{t^x})^T \right) \right) \right) \\ &= -\sum_{e \in F} 2R^{t^x}(e) \geq -2|F| \cdot \frac{2\gamma k}{n}. \end{aligned}$$

Therefore

$$\ln \det_+(D^{t'}) - \ln \det_+(D^t) = \int_0^{\ln \eta} \frac{d \ln \det_+(D^{t^x})}{dx} dx \geq -|F| \cdot \frac{4\gamma k}{n} \cdot \ln \eta.$$

Thus we have

$$\frac{\det_+(D^{t'})}{\det_+(D^t)} \geq e^{-\frac{2\gamma k}{n} \cdot 2|F| \cdot \ln \eta} \geq \left(1 - \frac{2\gamma k}{n}\right)^{2|F| \cdot \ln \eta} \geq \left(1 - \frac{2\gamma k}{n}\right)^{2|F| \cdot \log \eta}$$

as desired.  $\square$

The lemma is then a direct consequence of the above claim.  $\square$

*Proof of Lemma 10.7.* Let  $s'$  be the scaling obtained at the end of the  $t^{\text{th}}$  iteration of the outermost while loop. Let  $t_1$  be the total number of iterations executed so far by the first inner while loop (at Lines 15-16). Then by matrix determinant lemma

$$\begin{aligned} \frac{\det_+(D^{s'})}{\det_+(D^{s^0})} &\leq \left(1 - \frac{3\gamma_2 \cdot k}{8n}\right)^{2t_1} \left(1 + \frac{3\gamma_1 \cdot k}{2n}\right)^{2(t+t_1)} \\ &\leq \left(1 - \frac{6\gamma_1 \cdot k}{n}\right)^{2t_1} \left(1 + \frac{3\gamma_1 \cdot k}{2n}\right)^{2(t+t_1)} \\ &\leq \left(1 - \frac{4\gamma_1 \cdot k}{n}\right)^{2t_1} \left(1 + \frac{3\gamma_1 \cdot k}{2n}\right)^{2t} \end{aligned}$$

On the other hand, by Lemma 10.6,

$$\frac{\det_+(D^{s'})}{\det_+(D^{s^0})} \geq \left(1 - \frac{\gamma_1 k}{n}\right)^{2(t+t_1)}.$$

These two together imply that  $t_1 \leq 2t$ .  $\square$

## F Missing proofs from Section 11

*Proof of Proposition 11.18.* Multiplying the RHS out gives

$$\begin{pmatrix} L_{FF} & L_{FF}L_{FF}^\dagger L_{FC} \\ L_{CF}L_{FF}^\dagger L_{FF} & L_{CF}L_{FF}^\dagger L_{FF}L_{FC}^\dagger L_{FC} + L_{CC} - L_{CF}L_{FF}^\dagger L_{FC} \end{pmatrix}. \quad (102)$$

Then it suffices to show that  $L_{CF}L_{FF}^\dagger L_{FF} = L_{CF}$ , which will imply that (102) is equal to  $L$ . Notice that  $L_{CF}$  contains the edges between  $C$  and  $F$ . Moreover, we have  $L_{FF} = L_{G[F]} + \sum_{(u,v) \in E \cap (F \times C)} (e_{u \leftarrow v})_F (e_{u \leftarrow v})_F^T$ , and thus all rows of  $L_{CF}$  are in the range of  $L_{FF}$ , which gives that  $L_{CF}L_{FF}^\dagger L_{FF} = L_{CF}$ .  $\square$

*Proof of Claim 11.25.* It suffices to prove that  $D_G^{\dagger/2} L_{G^2} D_G^{\dagger/2} \preceq 2D_G^{\dagger/2} L_G D_G^{\dagger/2}$ . Notice that

$$D_G^{\dagger/2} L_G D_G^{\dagger/2} = I - D_G^{\dagger/2} A_G D_G^{\dagger/2}$$

and

$$D_G^{\dagger/2} L_{G^2} D_G^{\dagger/2} = I - D_G^{\dagger/2} A_G D_G^\dagger A_G D_G^{\dagger/2} = I - \left( D_G^{\dagger/2} A_G D_G^{\dagger/2} \right)^2.$$

Let  $\lambda_1, \dots, \lambda_n$  be the eigenvalues of  $D_G^{\dagger/2} L_G D_G^{\dagger/2}$ , and let  $f_1, \dots, f_n$  be a set of orthonormal eigenvectors. Then we have

$$\begin{aligned} D_G^{\dagger/2} L_G D_G^{\dagger/2} &= \sum_{i=1}^n \lambda_i f_i f_i^T \\ D_G^{\dagger/2} L_{G^2} D_G^{\dagger/2} &= \sum_{i=1}^n (1 - (1 - \lambda_i)^2) f_i f_i^T = \sum_{i=1}^n (2 - \lambda_i) \lambda_i f_i f_i^T. \end{aligned}$$

Since  $\lambda_i \in [0, 2]$ , we have our desired result.  $\square$

*Proof of Proposition 11.26.* Notice that by definition

$$\begin{aligned} \sum_{u \in V} \text{SC}_n(L, V \setminus \{u\}) &= \sum_{u \in V} \left( L_G - L_G(:, u) L_{uu}^\dagger L_G(:, u)^T \right) \\ &= nL_G - L_G D^\dagger L_G \\ &= n(D - A) - (D - A) D^\dagger (D - A) \\ &= nD - nA - (D - 2A + AD^\dagger A) \\ &= (n - 2)(D - A) + D - AD^\dagger A = (n - 2)L_G + L_{G^2} \end{aligned}$$

as desired.  $\square$

*Proof of Claim 11.27.* It suffices to prove that

$$\mathbb{E}_{v_{i+1}} [(60) \mid V_i] = L_{H[V - V_i]}.$$

Let us calculate the LHS term by term.

$$\begin{aligned} & \mathbb{E}_{v_{i+1}} \left[ L_{H[V-V_i]}(:, v_{i+1}) \left( D_{v_{i+1}}^{H[V-V_i]} \right)^\dagger L_{H[V-V_i]}(:, v_{i+1})^T \mid V_i \right] \\ &= \frac{1}{t-i} L_{H[V-V_i]} D_{H[V-V_i]}^\dagger L_{H[V-V_i]} \\ &= \frac{1}{t-i} \left( D_{H[V-V_i]} - 2A_{H[V-V_i]} + A_{H[V-V_i]} D_{H[V-V_i]}^\dagger A_{H[V-V_i]} \right) \end{aligned}$$

$$\mathbb{E}_{v_{i+1}} \left[ \left( 1 + \frac{2}{t-2-i} \right) L_{H[V-V_{i+1}]} \mid V_i \right] = \frac{t-2-i}{t-i} L_{H[V-V_i]} + \frac{2}{t-i} L_{H[V-V_i]} = L_{H[V-V_i]}$$

$$\begin{aligned} & \mathbb{E}_{v_{i+1}} \left[ \frac{1}{t-i} (L_{H[V-V_i]^2} - 2L_{H[V-V_i]}) \mid V_i \right] \\ &= \frac{1}{t-i} (L_{H[V-V_i]^2} - 2L_{H[V-V_i]}) \\ &= \frac{1}{t-i} \left( D_{H[V-V_i]} - A_{H[V-V_i]} D_{H[V-V_i]}^\dagger A_{H[V-V_i]} \right) - \frac{2}{t-i} L_{H[V-V_i]}. \end{aligned}$$

One can verify that these three add up to  $L_{H[V-V_i]}$ .  $\square$

## References

- [AGM12a] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Analyzing graph structure via linear measurements. In Yuval Rabani, editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 459–467. SIAM, 2012.
- [AGM12b] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Graph sketches: sparsification, spanners, and subgraphs. In *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2012, Scottsdale, AZ, USA, May 20-24, 2012*, pages 5–14, 2012.
- [AM85] Noga Alon and Vitali D Milman.  $\lambda_1$ , isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.
- [BK15] András A. Benczúr and David R. Karger. Randomized approximation schemes for cuts and flows in capacitated graphs. *SIAM J. Comput.*, 44(2):290–319, 2015.
- [CCL<sup>+</sup>15] Dehua Cheng, Yu Cheng, Yan Liu, Richard Peng, and Shang-Hua Teng. Efficient sampling for gaussian graphical models via spectral sparsification. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pages 364–390, 2015.
- [CF14] Graham Cormode and Donatella Firmani. A unifying framework for  $\ell_0$ -sampling algorithms. *Distributed Parallel Databases*, 32(3):315–335, 2014.

- [CKM<sup>+</sup>14] Michael B. Cohen, Rasmus Kyng, Gary L. Miller, Jakub W. Pachocki, Richard Peng, Anup B. Rao, and Shen Chen Xu. Solving SDD linear systems in nearly  $m \log^{1/2} n$  time. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 343–352, 2014.
- [CLM<sup>+</sup>15] Michael B. Cohen, Yin Tat Lee, Cameron Musco, Christopher Musco, Richard Peng, and Aaron Sidford. Uniform sampling for matrix approximation. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*, pages 181–190, 2015.
- [CSWZ16] Jiecao Chen, He Sun, David P. Woodruff, and Qin Zhang. Communication-optimal distributed clustering. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 3720–3728, 2016.
- [DKP<sup>+</sup>17] David Durfee, Rasmus Kyng, John Peebles, Anup B. Rao, and Sushant Sachdeva. Sampling random spanning trees faster than matrix multiplication. In Hamed Hatami, Pierre McKenzie, and Valerie King, editors, *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 730–742. ACM, 2017.
- [DMR18] Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional gaussians. *arXiv preprint arXiv:1810.08693*, 2018.
- [FGL<sup>+</sup>21] Sebastian Forster, Gramoz Goranci, Yang P. Liu, Richard Peng, Xiaorui Sun, and Mingquan Ye. Minor sparsifiers and the distributed laplacian paradigm. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 989–999. IEEE, 2021.
- [FKN21] Arnold Filtser, Michael Kapralov, and Navid Nouri. Graph spanners by sketching in dynamic streams and the simultaneous communication model. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1894–1913, 2021.
- [GR99] Oded Goldreich and Dana Ron. A sublinear bipartiteness tester for bounded degree graphs. *Comb.*, 19(3):335–373, 1999.
- [Har98] David A Harville. Matrix algebra from a statistician’s perspective, 1998.
- [Ind06] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006.
- [JKPS17] Gorav Jindal, Pavel Kolev, Richard Peng, and Saurabh Sawlani. Density independent algorithms for sparsifying k-step random walks. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2017, August 16-18, 2017, Berkeley, CA, USA*, pages 14:1–14:17, 2017.

- [JL84] William B Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space 26. *Contemporary mathematics*, 26, 1984.
- [JS21] Arun Jambulapati and Aaron Sidford. Ultrasparse ultrasparsifiers and faster laplacian system solvers. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 540–559, 2021.
- [JST11] Hossein Jowhari, Mert Saglam, and Gábor Tardos. Tight bounds for  $\ell_p$  samplers, finding duplicates in streams, and related problems. In *Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2011, June 12-16, 2011, Athens, Greece*, pages 49–58, 2011.
- [JW21] Rajesh Jayaram and David P. Woodruff. Perfect  $\ell_p$  sampling in a data stream. *SIAM J. Comput.*, 50(2):382–439, 2021.
- [KKTY21] Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. Towards tight bounds for spectral sparsification of hypergraphs. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 598–611. ACM, 2021.
- [KLM<sup>+</sup>14] Michael Kapralov, Yin Tat Lee, Cameron Musco, Christopher Musco, and Aaron Sidford. Single pass spectral sparsification in dynamic streams. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 561–570, 2014.
- [KLM<sup>+</sup>17] Michael Kapralov, Yin Tat Lee, Cameron Musco, Christopher Musco, and Aaron Sidford. Single pass spectral sparsification in dynamic streams. *SIAM J. Comput.*, 46(1):456–477, 2017.
- [KLOS14] Jonathan A. Kelner, Yin Tat Lee, Lorenzo Orecchia, and Aaron Sidford. An almost-linear-time algorithm for approximate max flow in undirected graphs, and its multi-commodity generalizations. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 217–226, 2014.
- [KLP<sup>+</sup>16] Rasmus Kyng, Yin Tat Lee, Richard Peng, Sushant Sachdeva, and Daniel A Spielman. Sparsified cholesky and multigrid solvers for connection laplacians. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 842–850, 2016.
- [KMM<sup>+</sup>20] Michael Kapralov, Aida Mousavifar, Cameron Musco, Christopher Musco, Navid Nouri, Aaron Sidford, and Jakab Tardos. Fast and space efficient spectral sparsification in dynamic streams. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1814–1833, 2020.

- [Kou14] Ioannis Koutis. Simple parallel and distributed algorithms for spectral graph sparsification. In Guy E. Blelloch and Peter Sanders, editors, *26th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA ’14, Prague, Czech Republic - June 23 - 25, 2014*, pages 61–66. ACM, 2014.
- [KS16] Rasmus Kyng and Sushant Sachdeva. Approximate gaussian elimination for laplacians - fast, sparse, and simple. In Irit Dinur, editor, *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 573–582. IEEE Computer Society, 2016.
- [KVV04] Ravi Kannan, Santosh S. Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *J. ACM*, 51(3):497–515, 2004.
- [LGT14] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6):37:1–37:30, 2014.
- [LMP13] Mu Li, Gary L. Miller, and Richard Peng. Iterative row sampling. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 127–136, 2013.
- [LPYZ20] Huan Li, Stacy Patterson, Yuhao Yi, and Zhongzhi Zhang. Maximizing the number of spanning trees in a connected graph. *IEEE Trans. Inf. Theory*, 66(2):1248–1260, 2020.
- [LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh S. Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1131–1140, 2012.
- [LS18] Huan Li and Aaron Schild. Spectral subspace sparsification. In Mikkel Thorup, editor, *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 385–396. IEEE Computer Society, 2018.
- [LZ18] Huan Li and Zhongzhi Zhang. Kirchhoff index as a measure of edge centrality in weighted networks: Nearly linear time algorithms. In Artur Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 2377–2396. SIAM, 2018.
- [MU17] Michael Mitzenmacher and Eli Upfal. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.
- [Nis92] Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992.

- [Pen16] Richard Peng. Approximate undirected maximum flows in  $O(m\text{polylog}(n))$  time. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1862–1867, 2016.
- [RSW18] Aviad Rubinstei, Tselil Schramm, and S. Matthew Weinberg. Computing exact minimum cuts without knowing the graph. In *9th Innovations in Theoretical Computer Science Conference, ITCS 2018, January 11-14, 2018, Cambridge, MA, USA*, pages 39:1–39:16, 2018.
- [She13] Jonah Sherman. Nearly maximum flows in nearly linear time. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 263–269, 2013.
- [SS11] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011.
- [ST04] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 81–90, 2004.
- [ST14] Daniel A. Spielman and Shang-Hua Teng. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM J. Matrix Anal. Appl.*, 35(3):835–885, 2014.
- [Tro11] Joel Tropp. Freedman’s inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, 2011.
- [Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.
- [Yao77] Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity (extended abstract). In *18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977*, pages 222–227. IEEE Computer Society, 1977.