Effective-Resistance-Reducing Flows and Asymmetric TSP

Nima Anari ∗ Shayan Oveis Gharan †

Abstract

Given a $k$-edge-connected graph $G = (V, E)$, a spanning tree $T \subseteq E$ is $\alpha$-thin w.r.t. $G$, if for any $S \subseteq V$, $|T(S, \overline{S})| \leq \alpha \cdot |E(S, \overline{S})|$. The thin tree conjecture asserts that for a sufficiently large $k$ (independent of size of $G$) every $k$-edge-connected graph has a $1/2$-thin tree. This conjecture is intimately related to designing approximation algorithms for Asymmetric TSP [AGM+10]. One can study thin trees by analyzing their generalizations known as spectrally thin trees. A spanning tree $T$ is $\alpha$-spectrally thin, if $L_T \preceq \alpha \cdot L_G$, where $L_T, L_G$ are the Laplacian matrices associated with $T, G$ respectively. Recently, Marcus, Spielman and Srivastava [MSS13] proved a sufficient condition for the existence of spectrally thin trees. On the other hand, it is known that many families of $k$-edge-connected graphs do not admit any spectrally thin trees.

In this paper, we introduce a procedure to “transform” graphs that do not admit any spectrally thin trees into those that may have spectrally thin trees. We prove the following statement and we conjecture that it implies improved upper-bounds on the integrality gap of LP relaxation of ATSP. For any $k = \Omega(\log(n))$-connected graph $G = (V, E)$, there is a set $F \subseteq E$ that induces an $\Omega(k)$-connected graph and a “transformation” $D$ of $G$ that preserves the structure of all cuts of $G$ such that the effective resistance of every edge in $F$ w.r.t. $D$ is at most $\tilde{O}(1/\sqrt{k})$. We show that a generalization of the recent breakthrough of [MSS13] together with the aforementioned statement implies an $\tilde{O}(\sqrt{\log(n)})$ upper-bound on the integrality gap of the LP relaxation of ATSP.

As simple applications of our techniques, we use [MSS13] to show that every $k$-edge-connected graph has an $\tilde{O}(1/k)$-thin edge cover and that every $K_h$-minor free graph has a $\tilde{O}(h/k)$-thin tree. The latter gives an upper-bound of $\tilde{O}(\text{polyloglog}(n) \cdot h)$ on the integrality gap of the LP relaxation of ATSP for $K_h$-minor free graphs.

∗ Computer Science Division, UC Berkeley. Email: anari@berkeley.edu
† Computer Science Division, UC Berkeley. Supported by a Miller fellowship. Email: oveisgharan@berkeley.edu
## Contents

1 Introduction .................................................. 1
   1.1 Spectrally Thin Trees ........................................ 2
   1.2 Our Contribution ............................................ 4
   1.3 Proof Overview ............................................. 7
   1.4 Organization ............................................... 11

2 Preliminaries .................................................. 12
   2.1 Notation .................................................... 12
   2.2 Facts from Linear Algebra ................................... 12
   2.3 Background in Graph Theory ................................. 15

3 From Spectral Thin Sets to Combinatorial Thin Sets ............... 16

4 Graph Partitioning ............................................ 19

5 The Dual of Hierarchical-CP .................................. 24

6 Upper-bounding the Dual of Hierarchical-CP ..................... 27
   6.1 Construction of Bag of Balls ............................... 32
   6.2 Construction of a Geometric Sequence of Families of Bags of Balls ................................. 35

7 Charging Arguments ............................................ 38
   7.1 Charging Argument for Compact Bags of Balls .............. 40
   7.2 Charging Argument for Assigned Bags of Balls ............ 43
      7.2.1 Ball Labeling .......................................... 43
      7.2.2 Order of Processing .................................... 47
      7.2.3 The Construction ....................................... 49
      7.2.4 Post Processing ........................................ 54

A Missing proofs of Section 1 .................................. 58
1 Introduction

In the Asymmetric Traveling Salesman Problem (ATSP) we are given a set $V$ of vertices and a non-negative cost function $c : V \times V \to \mathbb{R}_+$. The goal is to find the shortest tour that visits every vertex at least once.

If the cost function is symmetric, i.e., $c(u, v) = c(v, u)$ for all $u, v \in V$, then the problem is known as the Symmetric Traveling Salesman Problem (STSP). There is a $3/2$ approximation algorithm by Christofides [Chr76] for STSP.

There is a natural LP relaxation for ATSP proposed by Held and Karp [HK70],

$$
\begin{align*}
\min & \sum_{u,v \in V} c(u,v)x_{u,v} \\
\text{s.t.} & \sum_{u \in S, v \not\in S} x_{u,v} \geq 1 & \forall S \subseteq V, \\
& \sum_{v \in V} x_{u,v} = \sum_{v \in V} x_{v,u} = 1 & \forall u \in V, \\
& x_{u,v} \geq 0 & \forall u, v \in V.
\end{align*}
$$

(1)

It is conjectured that the integrality gap of the above LP relaxation is a constant, i.e., the optimum value of the above LP relaxation is within a constant factor of the length of the optimum ATSP tour. To this date, the best known lower-bound on the integrality gap of the above LP is $2$ [CGK06].

Despite many efforts, there is no known constant factor approximation algorithm for ATSP. Recently, Asadpour et al. [AGM+10] designed an $O(\log n / \log \log n)$ approximation algorithm for ATSP that breaks the $O(\log n)$ barrier from Frieze, Galbiati and Maffioli [FGM82] and subsequent improvements [Blä02, KLSS05, FS07]. The result of [AGM+10] also upper-bounds the integrality gap of the Held-Karp LP relaxation by $O(\log n / \log \log n)$. Later, the second author with Saberi [OS11] and subsequently Erickson and Sidiropoulos [ES14] designed constant factor approximation algorithms for ATSP on planar and bounded genus graphs.

The main ingredient of all of the above recent developments is the construction of a “thin” tree. Let $G = (V,E)$ be an unweighted undirected $k$-edge-connected graph with $n$ vertices. We allow $G$ to have an arbitrary number of parallel edges, so we think of $E$ as a multiset of edges. Roughly speaking, a spanning tree $T \subseteq E$ is $\alpha$-thin with respect to $G$ if it does not contain more than $\alpha$-fraction of the edges of any cut in $G$.

**Definition 1.1.** A spanning tree $T \subseteq E$ is $\alpha$-thin with respect to a (unweighted) graph $G = (V,E)$, if for each set $S \subseteq V$,

$$|T(S, \overline{S})| \leq \alpha \cdot |E(S, \overline{S})|,$$

where $T(S, \overline{S})$ and $E(S, \overline{S})$ are the set of edges of $T$ and $G$ in the cut $(S, \overline{S})$ respectively.

One can analogously define $\alpha$-thin edge covers, $\alpha$-thin paths, etc. Note that any spanning tree of an $\alpha$-thin connected subgraph of $G$ is an $\alpha$-thin spanning tree of $G$.

A key lemma in [AGM+10] shows that one can obtain an approximation algorithm for ATSP by finding a thin tree of small cost with respect to the graph defined by the fractional solution of the LP relaxation. In addition, proving the existence of a thin tree provides a bound on the integrality gap of the Held-Karp LP relaxation for ATSP.
Later, in [OS11] this connection is made more concrete. Namely, to break the \( \Theta(\frac{\log(n)}{\log\log(n)}) \)-barrier, it suffices to ignore the costs of the edges and construct a thin tree in every highly connected graph.

**Theorem 1.2.** For any sub-linear function \( f(\cdot) \), a poly-time construction of an \( f(\log(n))/\log(n) \) thin tree in any \( \Theta(\log(n)) \)-edge-connected graph gives an \( O(f(\log(n))) \) approximation algorithm for ATSP. In addition, even an existential proof gives an \( O(f(\log(n))) \) upper-bound on the integrality gap of the LP relaxation.

See Appendix A for the proof of the above theorem.

Independently of the above applications of thin trees, Goddyn formulated the thin tree conjecture because of the close connections to several long standing open problems regarding nowhere-zero flows.

**Conjecture 1.3** (Goddyn [God04]). There exists a function \( f(\alpha) \) such that, for any \( 0 < \alpha < 1 \), every \( f(\alpha) \)-edge-connected graph (of arbitrary size) has an \( \alpha \)-thin spanning tree.

Goddyn’s conjecture in the strongest form postulates that for a sufficiently large \( k \) that is independent of the size of \( G \), every \( k \)-edge-connected graph has a \( O(1/k) \)-thin tree. Goddyn proved that if the above conjecture holds for an arbitrary function \( f(\cdot) \) it implies a weaker version of the Jaeger’s conjecture on the existence of circular nowhere-zero flows [Jae84]. Very recently, Thomassen proved a weaker version of the Jaeger’s conjecture [Tho12, LTWZ13], but his proof has not yet shed any light on the resolution of the thin tree conjecture.

To this date, Conjecture 1.3 is only proved for planar and bounded genus graphs [OS11, ES14] and edge transitive graphs\(^1\) [HO14] for \( f(\alpha) = O(1/\alpha) \). We remark that if Goddyn’s thin tree conjecture holds for any arbitrary function \( f(\cdot) \), we get an upper-bound of \( O(\log^{1-\Omega(1)}(n)) \) on the integrality gap of the LP relaxation of ATSP.

### 1.1 Spectrally Thin Trees

As mentioned before, thin trees are the basis for the best known approximation algorithms for ATSP on planar, bounded genus, or general graphs. This follows from their intuitive definition and the fact that they eliminate the difficulty arising from the underlying asymmetry and the cost function. On the other hand, the major challenge in constructing thin trees or proving their existence is that we are not aware of any efficient algorithm for measuring or certifying the thinness of a given tree. In order to verify the thinness of a given tree, it seems that one has to look at exponentially many cuts.

One possibility to avoid this difficulty is to study a stronger definition of thinness, namely the spectral thinness. First, we define some notations. For a set \( S \subseteq V \) we use \( 1_S \) to denote the indicator vector of the set \( S \). For a vertex \( v \in V \), we abuse notation and write \( 1_v \in \mathbb{R}^V \), instead of \( 1_{\{v\}} \). For any edge \( e = \{u, v\} \in E \) we fix an arbitrary orientation, say \( u \to v \) and we use \( \chi_e := 1_u - 1_v \). The Laplacian of \( G \), \( L_G \) is defined as follows:

\[
L_G := \sum_{e \in E} \chi_e \chi_e^T.
\]

\(^1\)A graph \( G = (V, E) \) is edge transitive, if for any pair of edges \( e, f \in E \) there is an automorphism of \( G \) that maps \( e \) to \( f \).
We also use
\[ L_T := \sum_{e \in T} \chi_e \chi_e^\top, \]
to denote the Laplacian matrix of a set \( T \subseteq E \) of edges. We say \( T \) is \( \alpha \)-spectrally thin with respect to \( G \) if
\[ L_T \preceq \alpha \cdot L_G, \quad \text{i.e., for all } x \in \mathbb{R}^n, \quad x^\top L_T x \leq \alpha \cdot x^\top L_G x. \] (2)
Observe that if \( T \) is \( \alpha \)-spectrally thin, then it is also \( \alpha \)-(combinatorially) thin. To see that, note that for any set \( S \subseteq V, \ 1_S^\top L_T 1_S = |T(S, \overline{S})| \) and \( 1_S^\top L_G 1_S = |E(S, \overline{S})| \).

One can verify spectral thinness of \( T \) (in polynomial time) by finding the smallest \( \alpha \in \mathbb{R} \) such that
\[ L_G^{1/2} L_T L_G^{1/2} \preceq \alpha \cdot I, \]
i.e., by computing the largest eigenvalue of \( L_G^{1/2} L_T L_G^{1/2} \). Recall that \( L_G^\dagger \) is the pseudo-inverse of \( L_G \), and \( L_G^{1/2} \) is the square root of the pseudo-inverse of \( L_G; \ L_G^{1/2} \) is well-defined because \( L_G^\dagger \geq 0 \). So, unlike the combinatorial thinness, spectral thinness can be computed exactly in polynomial time.

The notion of spectral thinness is closely related to spectral sparsifiers of graphs, which has been studied extensively in the past few years [ST04, SS11, BSS14, FHHP]. Roughly speaking, a spectrally thin tree is a one-sided spectral sparsifier. A spectrally thin tree \( T \) would be a true spectral sparsifier if in addition to (2), it satisfies \( \alpha \cdot (1 - \epsilon) x^\top L_G x \leq L_T \) for some constant \( \epsilon \). Until the recent breakthrough of Batson, Spielman, and Srivastava, all constructions of spectral sparsifiers used at least \( \Omega(n \log(n)) \) edges of the graph [ST04, SS11, FHHP]. Because of this they are of no use for the particular application of ATSP. Batson, Spielman, and Srivastava [BSS14] managed to construct a spectral sparsifier that uses only \( O(n) \) edges of \( G \). But in their construction, they assign different weights to the edges of the sparsifier which again makes their contribution not helpful for ATSP.

Indeed, it was observed by several people that there is an underlying barrier for the construction of spectrally thin trees and unweighted spectral sparsifiers. Many families of \( k \)-edge-connected graphs do not admit spectrally thin trees (see [HO14, Thm 4.9]). Let us elaborate on this observation. The effective resistance of an edge \( e \) in \( G \), \( \text{Reff}_G(e) \), is the electrical resistance measured across the endpoints of \( e \) when the network represents an electrical circuit with each edge being a resistor of resistance 1 (and if \( G \) is weighted, the resistance is the inverse of the weight of \( e \)). Mathematically, the effective resistance can be computed using \( L_G^\dagger \),
\[ \text{Reff}_G(e) := \chi_e L_G^\dagger \chi_e. \]

It is not hard to see that the spectral thinness of any spanning tree \( T \) of \( G \) is at least the maximum effective resistance of the edges of \( T \) in \( G \).

**Lemma 1.4.** For any graph \( G = (V, E) \), the spectral thinness of any spanning tree \( T \subseteq E \) is at least \( \max_{e \in T} \text{Reff}_G(e) \).

**Proof.** Say the spectral thinness of \( T \) is \( \alpha \). By definition, for any edge \( e \in T, \)
\[ L_G^{1/2} \chi_e \chi_e^\top L_G^{1/2} = L_G^{1/2} L_E L_G^{1/2} \preceq L_G^{1/2} L_T L_G^{1/2} \preceq \alpha \cdot I. \]
See Lemma 2.4 for a proof of the first matrix inequality. So, for any vector \( x \),
\[ x^\top L_G^{1/2} \chi_e \chi_e^\top L_G^{1/2} x \leq \alpha \cdot x^\top x. \]
Letting \( x = L_G^{1/2} \chi_e \), we get \( \text{Reff}_G(e) \leq \alpha \).

In light of the above lemma, a necessary condition for \( G \) to have a spanning tree with spectral thinness bounded away from 1 is that every cut of \( G \) must have at least one edge with effective resistance bounded away from 1. In other words, any graph \( G \) with at least one cut where the effective resistance of every edge is very close to 1 has no spectrally thin tree (see Figure 1 for an example of a graph where the effective resistance of every edge in a cut is very close to 1).

In a very recent breakthrough, Marcus, Spielman and Srivastava [MSS13] proved the Kadison-Singer conjecture. As a byproduct of their result, they show that a stronger version of the above condition is sufficient for the existence of spectrally thin trees.

**Theorem 1.5 ([MSS13]).** Any connected graph \( G = (V, E) \) has a spanning tree with spectral thinness \( O(\max_{e \in E} \text{Reff}_G(e)) \).

See [HO14, Appendix E] for a detailed proof of the above theorem. It follows from the above theorem that every \( k \)-edge-connected edge-transitive graph has an \( O(1/k) \) spectrally thin tree. This is because in any edge-transitive graph, by symmetry, the effective resistances of all edges are equal.

It follows from the simple extension of [MSS13] by Akemann and Weaver [CA13] that even if a constant fraction of edges in every cut of \( G \) have small effective resistance, then \( G \) has a spectrally thin tree.

**Theorem 1.6.** Given a connected graph \( G = (V, E) \), suppose that there is a set \( E' \subseteq E \) such that for any set \( S \subseteq V \), \( |E'(S, S)| \geq \alpha \cdot |E(S, S)| \). Then \( G \) has an \( O(\frac{1}{\alpha} \cdot \sqrt{\max_{e \in E'} \text{Reff}_G(e)}) \) spectrally thin spanning tree.

See Appendix A for the proof of the theorem.

As a side remark, we note that the sum of effective resistances of all edges of any connected graph \( G \) is \( n - 1 \),

\[
\sum_{e \in E} \chi^\top_e L_G^\dagger \chi_e = \sum_{e \in E} \text{Tr}(L_G^{1/2} \chi_e \chi^\top_e L_G^{1/2}) = \text{Tr} \left( \sum_{e \in E} L_G^{1/2} \chi_e \chi^\top_e L_G^{1/2} \right) = \text{Tr}(L_G^{1/2} L_G L_G^{1/2}) = n - 1.
\]

If \( G \) is \( k \)-connected, at most a quarter of the edges have effective resistance more than \( 8/k \). Therefore, by an application of [MSS13] any \( k \)-connected graph \( G \) has an \( O(1/k) \)-spectral thin set of edges, \( F \subseteq E \) where \( |F| \geq \Omega(n) \) [HO14]. But unfortunately, the corresponding subgraph \( (V, F) \) may have \( \Omega(n/k) \) connected components. So, this does not give any improved bounds on approximability of ATSP.

### 1.2 Our Contribution

In this paper we introduce a procedure to “transform” graphs that do not admit spectrally thin sets of edges into those that may have these sets. We then use [MSS13] to find spectrally thin sets of edges in the transformed graph. Finally, we run the inverse of our procedure to obtain (combinatorially) thin sets of edges in the original graph. From a high level perspective, our transformation symmetrizes the effective resistance of the edges, while keeping the cut structure of the graph intact. This allows us to bypass the the spectral thinness barrier described in Lemma 1.4.

For two matrices \( A, B \in \mathbb{R}^{n \times n} \), we write, \( A \preceq \Box B \), if for any \( S \subseteq V \),

\[
1^\top_S A 1_S \preceq 1^\top_S B 1_S.
\]
Figure 1: The top shows a $k$-connected planar graph that has no spectraly thin tree. Vertex 1 is connected to $k + 1$ vertices $2, n/k, 2n/k, \ldots, n$ and for each $2 \leq i \leq n$ there are $k$ parallel edges between $i, i + 1$. The effective resistance of each edge incident to 1 is $1 - O(k/n)$. The bottom shows a graph $G + D$ where the effective resistance of every black edge is $O(1/k)$. The red edges are edges in $D$ and have weight $k$ chosen such that $L_D \preceq L_G$.

Note that $A \preceq B$ implies $A \preceq L_G$, but the converse is not necessarily true.

In light of Theorem 1.6, if there is a (weighted) graph $D$ such that $L_D \preceq L_G$ and that for any edge $e \in E$, $\text{Reff}_{G+D}(e) \ll 1$, then, by Theorem 1.6, $D + G$ has a spectrally thin tree supported on the edges of $E$. But, since $L_D \preceq L_G$, any $\alpha$-spectrally thin tree of $D + G$ is a $2\alpha$-combinatorially thin tree of $G$.

Let us give a clarifying example. Consider the $k$-connected planar graph $G$ illustrated at the top of Figure 1. In this graph, all edges incident to the vertex 1 have effective resistance very close to 1. Now, let $D$ consist of the red edges (with weight $k$) connecting the neighbors of vertex 1 (as shown at the bottom). Observe that, $L_D \preceq L_G$. Now, the effective resistance of every black edge in $G + D$ is $O(1/k)$. So, by Theorem 1.6, $G + D$ has a spectrally thin tree supported on black edges, i.e., the edges of $G$. Such a tree is combinatorially thin with respect to $G$.

Unfortunately, such a graph $D$ may not necessarily exist for any $k$-connected graph $G$. Our first contribution is that there are $\Omega(\log(n))$-connected (planar) graphs $G$ such that for any graph $D$ where $L_D \preceq L_G$, the maximum effective resistance of the edges of $G$ in $D + G$ is $\Omega(1)$.

**Theorem 1.7.** The $k$ connected graph $G = (V, E)$ illustrated in Figure 2, satisfies the following: For any graph $D$ with $L_D \preceq L_G$,

\[
\max_{e \in E} \text{Reff}_{G+D}(e) \geq \frac{\log(n)^2}{8(\log(n) + k)^2}.
\]

The above impossibility theorem prevents a direct reduction of the spectral thinness to combinatorial thinness. Therefore, we take a detour. In our main theorem we show that there is a matrix $D$ that decreases the effective resistance of a set $F \subseteq E$ of edges such that $(V, F)$ is $\Omega(k)$-connected.

**Theorem 1.8 (Main).** For any $k$-connected graph $G = (V, E)$ where $k \geq 4\log(n)$, there is a matrix $0 < D \preceq L_G$ and a set $F \subseteq E$ such that the graph $(V, F)$ is $\Omega(k)$ connected and that for any edge...
Figure 2: A tight example for Theorem 1.7. The graph has $2^h + 1$ vertices labeled with \{0, 1, \ldots, 2^h\}. There are $k$ parallel edges connecting each pair of consecutive vertices. For any $1 \leq i \leq h$ and any $0 \leq j < 2^{h-i}$ there is an edge \{\(j \cdot 2^i, (j+1) \cdot 2^i\}\}.

e \in F, \chi_e^T D^{-1} \chi_e \leq \tilde{O}(1/\sqrt{k}).^2

The above theorem can be seen as a corollary of the thin tree conjecture. Suppose that every $k$-connected graph has a $C/k$-thin tree. Then, a $k$-connected graph $G = (V, E)$ has at least $\sqrt{k}$ disjoint $2C/k$ thin trees, say $T_1, \ldots, T_{\sqrt{k}}$. Now, let $D$ be the union of these $\sqrt{k}$ trees each with weight $\sqrt{k}/2C$. Then, $L_D \preceq L_G$. Now, let $F = \bigcup_{i=1}^{\sqrt{k}} T_i$. For any edge $e \in F$, $\Reff_{G+D}(e) \leq \frac{1}{1+\sqrt{k}/2C}$ because each edge of $F$ is parallel with an edge of weight $\sqrt{k}/2C$ in $G + D$.

We conjecture that any graph $G = (V, E)$ with a set $F \subseteq E$ such that $(V, F)$ is $k$-connected and $\max_{e \in F} \Reff_G(e) \ll 1$ has a $1/2$-spectrally thin tree. In fact, a stronger variant of the main theorem of [MSS13] together with our main theorem implies that the integrality of LP (1) is $O(\sqrt{\log(n)})$.

For $\gamma : E \to \mathbb{R}$, a $\gamma$-uniform random spanning tree distribution $\mu$, is a distribution on spanning trees of $G$, where for each spanning tree $T$,

$$\mathbb{P}[T] \propto \prod_{e \in T} \exp(\gamma_e).$$

**Conjecture 1.9.** Given are a graph $G = (V, E)$, a set $F = \{e_1, \ldots, e_m\} \subseteq E$, and a $\gamma$-uniform random spanning tree distribution, $\mu$, on the edges of $F$. For any $e_i \in F$, let $p_{e_i} = \mathbb{P}_{T \sim \mu}[e \in T]$. For some $\epsilon > 0$, if we assign a vector $x_i$ to each edge $e_i \in F$ such that

$$\sum_{i=1}^{m} x_i x_i^T \preceq I$$

and for all $1 \leq i \leq m$,

$$\frac{\|x_i\|^2}{p_{e_i}} \leq \epsilon,$$

then,

$$\mathbb{P}_{T \sim \mu} \left[ \left\| \sum_{e_i \in T} x_i x_i^T \right\| \leq O(1 + \epsilon) \cdot \max_{1 \leq i \leq m} p_{e_i} \right] > 0.$$

---

^2For functions $f(\cdot), g(\cdot)$ we write $g = \tilde{O}(f)$ if $g(n) \leq \text{polylog}(f(n)) \cdot f(n)$ for a sufficiently large $n$. 

6
It is instructive to compare the above conjecture with Theorem 3.1 that is the main theorem of [MSS13]. In the above conjecture we assume the given vectors are assigned to the edges of a random spanning tree spanning tree distribution (or more generally to the elements of a determinantal distribution), whereas Theorem 3.1 assumes a product distribution on the given random vectors. By our main theorem, any proof of the above conjecture implies that the integrality gap of the LP relaxation of ATSP is $\tilde{O}(\sqrt{\log(n)})$.

**Proposition 1.10.** If Conjecture 1.9 holds true, then any $4\log(n)$-connected graph has an $\tilde{O}(1/\sqrt{\log(n)})$-thin tree. Therefore, by Theorem 1.2, the integrality gap of the LP relaxation of ATSP is at most $\tilde{O}(\sqrt{\log(n)})$.

See Appendix A for the proof.

Although we do not resolve the thin tree conjecture, to show the strength of our techniques, we use our tools and [MSS13] to show that every $k$-connected graph has an $\tilde{O}(1/k)$ thin edge-cover and every $K_h$ minor-free graph has a $\tilde{O}(h/k)$-thin spanning tree.

We start with the former result. Recall that a set $F \subseteq E$ is an edge cover if any vertex $v \in V$ is incident to at least one edge of $F$.

**Theorem 1.11.** Every $k$-edge-connected graph $G = (V, E)$ has an $\tilde{O}(1/k)$-thin edge cover.

We remark that by Lemma 1.4 many families of graphs do not admit any spectrally thin edge cover. For example, the graph in the top of Figure 1 does not have any $1/2$ spectrally thin edge cover because all edges incident to 1 have effective resistance close to 1.

Given a graph $G = (V, E)$, we say $G$ is $h$-sparse, if the graph obtained by keeping only one copy of every parallel edge of $G$ has at most $h \cdot |V|$ edges. We say $G$ is minor $h$-sparse, if every minor of $G$ is $h$-sparse. For example, every planar graph is minor 2.5-sparse and every $K_h$-minor free graph is minor $O(h \sqrt{\log(h)})$-sparse [Tho01].

**Theorem 1.12.** Every minor $h$-sparse $k$-connected graph $G$ has an $\tilde{O}(h/k)$-thin spanning tree. Therefore, every $K_h$-minor free, $k$-connected graph has an $\tilde{O}(h/k)$-thin tree. Consequently, the integrality gap of the LP relaxation of ATSP on $K_h$-minor free graphs is at most $\tilde{O}(h \cdot \text{polyloglog}(n))$.

We emphasize that all of the above results are existential. To find a polynomial time algorithm for finding thin trees, firstly we need to have a constructive (in poly-time) variant of the proof of [MSS13]. Secondly, we need to find the promised matrix $D$ in the main theorem in polynomial time.

### 1.3 Proof Overview

As mentioned before, our main idea is to “symmetrize” the $L_2$ structure of the graph while preserving its $L_1$ structure. Formally, we write a convex program to minimize the effective resistance of the edges of the graph, while preserving the value of every cut up to a constant factor. This corresponds to a convex optimization problem, because for any fixed vector $x$ and $D \succ 0$, $x^TD^{-1}x$ is a convex function of $D$. See Lemma 2.3 for the proof. The problem of minimizing the sum of effective resistances of all pairs of vertices in a given graph is previously studied in [GBS08].

The following (exponentially sized) convex program finds the best matrix $D$ that minimizes the maximum effective resistance of the edges of $G$ while preserving the cut-structure of $G$. Note that the optimum matrix $D$ is not necessarily a Laplacian of a graph and indeed for our application of
Max-CP:

\[
\begin{align*}
\min & \quad \mathcal{E}, \\
\text{s.t.} & \quad \chi_e^T D^{-1} \chi_e \leq \mathcal{E} \quad \forall e \in E, \\
& \quad \frac{1}{2} L_G \preceq D \preceq L_G, \\
& \quad D \succ 0.
\end{align*}
\]

finding spectrally thin trees, any positive definite matrix \( D \) that approximates the cut structure of \( G \) is enough.

We can simplify the above program by dropping the constraint \( \frac{1}{2} L_G \preceq D \preceq L_G \), and instead averaging out the optimum \( D \) with the matrix \( L_G \). Note that if we replace the contract \( D \preceq L_G \) with \( D \preceq L_G \), i.e., if we require \( D \) to be upper-bounded by \( L_G \) in the \( L_2 \) sense, then the optimum \( D \) for any graph \( G \) is exactly \( L_G \) and the optimum value is the maximum effective resistance of edges of \( G \).

Unfortunately, as alluded to in Theorem 1.7, the optimum of the above program can be very close to \( 1 \) even if the input graph \( G \) is \( \log(n) \)-edge-connected. In the proof, we construct a feasible solution of the dual (of a relaxed variant) of the above convex program (see Section 5 for details).

Instead of the above strong program we may as well work with a relaxation that is still enough for finding spectrally thin trees. If the average effective resistance of edges across every cut of \( G \) is at most \( \mathcal{E} \), then we can let \( \alpha = 1/2 \) and \( F \) be the edges with effective resistance at most \( 2\mathcal{E} \) and use Theorem 1.6 to find a spectrally thin tree. The following convex program finds the best matrix \( D \) that minimizes the maximum of average effective resistance of edges across all cuts while preserving the cut structure of \( G \). Unfortunately, the same example shows that the optimum of the

Average-CP:

\[
\begin{align*}
\min & \quad \mathcal{E} \\
\text{s.t.} & \quad \mathbb{E}_{e \sim E(S, \overline{S})} \chi_e^T D^{-1} \chi_e \leq \mathcal{E} \quad \forall \emptyset \subset S \subset V, \\
& \quad D \preceq L_G, \\
& \quad D \succ 0.
\end{align*}
\]

new convex program is very close 1 for an \( \Omega(\log(n)) \)-edge-connected graph. In fact, in the proof of Theorem 1.7, we lower-bound the optimum of the above program.

In our main theorem 1.8, we decrease the maximum effective resistance of edges in a set \( F \subset E \) to a value bounded away from 1 such that \((V, F)\) is \( \Omega(k) \)-connected. The main difficulty in proving Theorem 1.8 is that the set \( F \) is unknown. In addition, the smallest effective resistance (or the \( k \)-th smallest effective resistance) of the edges of a cut of \( G \) is not a convex function of \( D \). So, we cannot write a single program that gives us the best matrix \( D \).

So, we take a detour. We select (almost) disjoint subsets of edges of \( G \), say \( F_1, F_2, \) etc. and we find a matrix \( D \) that minimizes the maximum of the average effective resistance of edges across all \( F_i \)'s. The question is how to choose \( F_i \)'s such that (i) There is a matrix \( D \) such that the maximum of the average effective resistance of edges with respect to \( D \) in each \( F_i \) is bounded away from 1, and (ii) Assuming that the average effective resistance of edges in each \( F_i \) is bounded away from 1,
there is a set $F \subset \bigcup_t F_t$ such that $(V,F)$ is $\Omega(k)$ connected.

Note that (i) is nontrivial because for the graph illustrated in Figure 2 there are disjoint subsets of edges such that for any matrix $0 < D \preceq L_G$, the maximum of average effective resistance of edges in the chosen sets is very close to 1. (ii) is non-trivial because $F_t$’s are disjoint.

First, we observe if $F_t$’s are the degree cuts of the vertices, then the maximum average effective resistance is bounded away from 1. This implies Theorem 1.8 for expander graphs (see Subsection 2.3 for the definition of expander graphs). If $G$ is an $\varepsilon$-expander, we can let $F$ be the edges with effective resistance at most $2/\varepsilon$ of the maximum average effective resistance of degree cuts w.r.t. $D$. Since $G$ is $\varepsilon$-expander, $F$ is $k/2$-connected. One might think that, in an expander graph, the effective resistance of every edge is very small [CRR+96], so the above argument is pointless. On the contrary, the importance of this argument is that it extends to all minors of $G$, that is, if a minor of $G$ is an expander, we can use the above argument to decrease the effective resistance of a set $F$ of the edges of that minor.

Now, suppose $G$ is a union of two expanders $G_1, G_2$ of size $n/2$ connected by $k$ edges. Even if the average effective resistance of all vertices is small, it can be the case that all edges separating $G_1, G_2$ have effective resistance close to 1. So, we also need to minimize the average effective resistance of edges of the cut separating $G_1, G_2$. Our high-level idea is as follows: First, we construct a hierarchical structure on the edges of $G$ that we call a hierarchical decomposition of $G$. Roughly speaking, we recursively decompose $G$ into expander graphs and we contract each expander into a single vertex. Then, we simultaneously minimize the average effective resistance of the degree cuts of each of the expanders in the hierarchy using our main convex program.

For a graph $\hat{G} = (V, E)$, a hierarchical decomposition, $T$, is a tree where each leaf corresponds to a unique vertex of $G$. We use the terminology node to refer to vertices of $T$. For each node $t \in T$ let $V(t) \subseteq V$ be the set of vertices of $G$ that are mapped to the leaves of the subtree of $t$, $E(t)$ be the set of edges between the vertices of $V(t)$ and $P(t) := E(V(t),\overline{V(t)})$ be the set of edges that leave $V(t)$. Say $t^*$ is the father of a node $t$, we define $O(t) := E(V(t),V(t^*) - V(t))$ as the set of edges between $V(t)$ and the rest of the vertices of $G$ in the subtree of $t^*$. We abuse notation and use $T$ to also denote the set of nodes of $T$.

Let us give a clarifying example. Say $G$ is the “bad” graph of Figure 2. In Figure 8 we give a hierarchical decomposition of $G$. For each node $t_i$, $V(t_i) = \{0,1,\ldots,i\}$. For each $1 \leq i \leq 2^h$, the set $O(i)$ is the set of edges from vertex $i$ to all vertices $j$ with $j < i$. In addition, since $t_{i+1}$ has exactly two children, $O(i) = O(t_i)$. Finally, $P(i)$ is all edges incident to vertex $i$ and $P(t_i)$ is the set of edges $E(\{0,1,\ldots,i-1\},\{i,i+1,\ldots,2^h\})$.

For an integer $k > 1$, $0 < \lambda < 1$ and $T \subseteq T$, we say $T$ is a $T(k,\lambda,T)$ hierarchical decomposition of $G$ if
1. For each node $t \in T$ the induced graph $G[V(t)]$ is $k$-edge connected.
2. For any node $t \in T$ that is not the root, $|O(t)| \geq k$. This property is a consequence of 1.
3. For any node $t \in T$, $|O(t)| \geq \lambda \cdot |P(t)|$.

We also write $T(k, \lambda, \cdot)$ to denote the decomposition where $T$ is all nodes of $T$. For example, the tree of Figure 8 is a $T(k, 1/2, \{1, 2, \ldots, 2^h\})$-hierarchical decomposition of the graph illustrated in Figure 2. Condition 1 holds because there are $k$ parallel edges between any pairs of vertices $i, i+1$, so $|O(i)| = |O(t_i)| = |E(\{0, \ldots, i-1\}, \{i\})| \geq k$.

It is easy to see that for any node $i$, $|O(i)| \geq d(i)/2 = |P(i)|/2$.

Given a hierarchical decomposition $T(k, \lambda, \cdot)$ of $G$, the sets $O(t)$ for nodes $t \in T$ are the $F_i$'s that we talked about above. First, observe that they are almost disjoint because each edge is contained in at most two of these sets. In our main technical theorem, we show that there is a matrix $D$ that reduces the average effective resistance of edges in all sets $O(t)$ to $\tilde{O}(1/k)$.

In Section 4, we construct hierarchical decompositions of $G$ by finding and contracting induced expanders such that, assuming the average effective resistance of the sets $O(t)$ is small, there is an $\Omega(k)$-connected set $F \subset E$ with edges of small effective resistance.

The following convex program finds a matrix $0 \prec D \preceq L_G$ that minimizes the maximum of the average effective resistance of edges in $O(t)$ for all $t \in T$.

**Hierarchical-CP**($T(k, \lambda, T)$):

$$
\begin{align*}
\min\ E \\
\text{s.t. } \sum_{e \sim O(t)} x_e^T D^{-1} x_e \leq \mathcal{E} & \quad \forall t \in T, \\
D \preceq L_G, \\
D \succ 0.
\end{align*}
$$

**Theorem 1.13** (Main Technical). For any $k$-edge-connected graph $G$, and any hierarchical decomposition $T(k, \lambda, \cdot)$ of $G$, there is a positive definite matrix $D$ such that $D \preceq L_G$ and for any $t \in T$,

$$
\sum_{e \sim O(t)} x_e^T D^{-1} x_e \leq \frac{f_1(k, \lambda)}{k}.
$$

Here $f_1(k, \lambda)$ is a poly-logarithmic function of $k, \lambda$.

We emphasize that the above theorem does not give a polynomial time algorithm for finding the matrix $D$. This is because a separation oracle for Hierarchical-CP must test exponentially many constraints to verify $D \preceq L_G$. There is a weaker method, analogous to the dual of the ARV SDP relaxation of the uniform sparsest cut problem [ARV09], to test the inequality $D \preceq L_G$.

For a Laplacian matrix $D$, if we can route a flow with capacity 1 on each edge of $G$ such that $D$ is the Laplacian of the demand graph of the flow, then we have $D \preceq L_G$. Having this in mind, we can write a weaker convex program to minimize the average effective resistance of edges of sets $O(t)$. For any pair of vertices $u, v \in V$, let $P_{u,v}$ be the set of all simple paths from $u$ to $v$ in $G$. 
For each path $P$, let $y_P \geq 0$ be the flow that we route along the path $P$. Then $\sum_{P \in P_{u,v}} y_P$ is the routed demand from $u$ to $v$.

\[
\min \quad \mathcal{E} \\
\text{s.t.} \quad \sum_{e \sim O(t)} \chi^T D^{-1} \chi_e \leq \mathcal{E} \quad \forall t \in T \\
D = \frac{J}{2n^2} + \frac{1}{2} L_G + \sum_{u,v \in V} \sum_{P \in P_{u,v}} y_P \cdot \left( L_{\{u,v\}} - \sum_{e \in P} L_e \right), \\
y_P \geq 0 \quad \forall P \\
D > 0.
\]

In the above we use $J$ to denote the all 1 matrix. The term $J/2n^2$ in the RHS of second equation is used to make sure that $D$ is not orthogonal to the all 1 vector. It is easy to see that for any feasible matrix $D$ in the above program, $D \preceq L_G$. Roughly speaking, the main difference between the dual of the above program and the dual of Hierarchical-CP is that the dual of the above program corresponds to maximizing a function over all $L^2$ metrics of vertices of $G$ while the dual of Hierarchical-CP corresponds to maximizing the same function over all cut metrics of $G$. Although we do not analyze the above program, we conjecture that for any hierarchical decomposition $T(k, \lambda, T)$ of a $k$-connected graph $G$, the optimum of the above program is $O(\text{polylog}(k, \lambda)/k)$. This would make our main technical theorem algorithmic.

Note that unlike our main theorem, the main technical theorem does not depend on the size of $G$. In Section 5 we write the dual of Hierarchical-CP. Then, the main body of the paper is dedicated to the analysis of the dual. Another way for approaching Theorem 1.13 is to explicitly exhibit a matrix $D$ or a flow-routing that reduces the average effective resistance of edges in $O(t)$ for all $t \in T$. The difficulty of this approach is that any loss in the routing should not depend on the size of $G$. This is particularly hard because the height of the given hierarchical decomposition can be infinitely large compared to $k$.

At the heart of our dual proof in Section 7, we use an inductive argument with no loss in $n$. Roughly speaking, our goal is to prove the following geometric question: Suppose we have an $L_1$ embedding of $G$, and we are given a geometric sequence of families of $L_1$ balls $B_1, B_2, \ldots$, such that centers of all balls in all families are vertices of $G$, balls in each $B_i$ are disjoint and their radii are exactly equal to $\delta_i$, where $\delta_1, \delta_2, \ldots$ form a geometric sequence where the scale factor is $\text{poly}(k)$. We prove that under some technical conditions on $B_1, B_2, \ldots$, we can construct a set of label-disjoint balls such that the sum of the radii of balls in our set is a constant factor of the sum of the radii of balls in the given geometric sequence, where by label-disjoint balls we mean that we can label each ball with an $\Omega(k)$-connected subgraph of $G$ such that any two intersecting balls are labeled with edge-disjoint subgraphs of $G$.

1.4 Organization

The rest of the paper is organized as follows: We start with an overview of linear algebraic tools and graph theoretic tools that we use in the paper. In Section 3 we prove Theorems 3.2, 1.11 and 1.12. Then in Section 4 we prove the main theorem 1.8 assuming the main technical theorem 1.13. The rest of the paper is dedicated to the proof of Theorem 1.13. In Section 5 we characterize the dual of Hierarchical-CP and we prove Theorem 1.7, then in sections 6 and 7 we upper-bound the dual of Hierarchical-CP.
2 Preliminaries

2.1 Notation

Unless otherwise specified, we assume that $G = (V, E)$ is a $k$-connected graph with $n$ vertices and $m$ edges. Note that conventionally the term “$k$ connectivity” refers to $k$ vertex connectivity, here we abuse the terminology we use $k$-connectivity to refer to $k$-edge connectivity. All graphs that we are working with are unweighted but they may have arbitrary number of parallel edges. We assume that they don’t have any loops.

For disjoint sets $A, B \subseteq V$ we write $E(A, B) := \{\{u, v\} : u \in A, v \in B\}$.

For a set $S$ of elements we write, $\mathbb{E}_{e \sim S}[,]$ to denote the expectation under the uniform distribution over the elements of $S$.

We think of a permutation of a set $S$ as a bijection mapping the elements of $S$ to $1, 2, \ldots, |S|$. For a vector $x \in \mathbb{R}^d$, we write

$$
\|x\| = \sqrt{\sum_{i=1}^{d} x_i^2},
$$

$$
\|x\|_1 = \sum_{i=1}^{d} |x_i|.
$$

For $x \in \mathbb{R}^d$ and $r \in \mathbb{R}$, an $L_1$ ball is the set of points at distance $r$ from $x$,

$$
B_1(x, r) := \{y \in \mathbb{R}^d : \|x - y\|_1 < r\}.
$$

Unless otherwise specified, any ball that we consider in this paper is a $L_1$ ball. We also allow for a portion of a ball. For a point $x \in \mathbb{R}^d$ and $0 \leq r_1 < r_2$ we write,

$$
B(x, r_1, r_2) := \{y \in \mathbb{R}^d : r_1 < \|x - y\|_1 < r_2\}
$$

Observe that $B(x, r) = B_1(x, 0, r)$. We say a point $y \in \mathbb{R}^d$ is inside a ball $B = B(x, r_1, r_2)$ if

$$
r_1 < \|x - y\|_1 < r_2,
$$

and we say it is outside of $B$ otherwise. We also say a ball $B_1$ is inside a ball $B_2$ if any point $x \in B_1$ is also in $B_2$.

For an integer $k \geq 1$, we use $[k]$ to denote the set $\{1, \ldots, k\}$.

2.2 Facts from Linear Algebra

We use $I$ to denote the identity matrix and $J$ to denote the all 1 matrix. For a matrix $A \in \mathbb{R}^{m \times n}$ we write

$$
\text{Tr}(A) := \sum_{i=1}^{\min\{m, n\}} A_{i,i}.
$$
A matrix $U \in \mathbb{R}^{n \times n}$ is called unitary if $UU^\top = U^\top U = I$. A unitary matrix is a non-singular matrix where all singular values are 1. It follows by definition that unitary operators preserve $L_2$ norms of vectors, i.e., for any vector $x \in \mathbb{R}^n$,

$$\|Ux\| = \sqrt{(Ux)^\top Ux} = \sqrt{x^\top U^\top Ux} = \sqrt{x^\top x} = \|x\|.$$ 

For two matrices $A, B$ of the same dimension we define the matrix inner product $A \bullet B := \text{Tr}(AB)$. For any matrix $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$,

$$\text{Tr}(AB) = \text{Tr}(BA).$$

**Lemma 2.1.** If $A, B$ are positive semidefinite matrices of the same dimension, then

$$\text{Tr}(AB) \geq 0.$$ 

**Proof.**

$$\text{Tr}(AB) = \text{Tr}(AB^{1/2}B^{1/2}) = \text{Tr}(B^{1/2}AB^{1/2}) \geq 0.$$ 

**Fact 2.2 (Schur Complement [BV06, Section A.5]).** For any symmetric PD matrix $A \in \mathbb{R}^{n \times n}$ a (column) vector $x \in \mathbb{R}^n$ where and $c \geq 0$, we have $x^\top A^{-1}x \leq c$ if and only if

$$\begin{bmatrix} c & x^\top \\ x & A \end{bmatrix} \succeq 0.$$ 

**Lemma 2.3.** For any two symmetric $n \times n$ matrices $A, B \succ 0$,

$$\left(\frac{1}{2}A + \frac{1}{2}B\right)^{-1} \geq \frac{1}{2}A^{-1} + \frac{1}{2}B^{-1}.$$ 

**Proof.** For any vector $x \in \mathbb{R}^n$,

$$\frac{1}{2} \begin{bmatrix} x^\top A^{-1}x & x^\top \\ x & A \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x^\top B^{-1}x & x^\top \\ x & B \end{bmatrix} = \left[ \frac{1}{2}x^\top A^{-1}x + \frac{1}{2}x^\top B^{-1}x \\ x \\ \frac{1}{2}A + \frac{1}{2}B \right].$$

By Schur complement both of the matrices in the LHS of above equality are PSD. Therefore, by convexity of PSD matrices, the matrix in RHS is also PSD. By another application of Schur complement to the matrix in RHS we obtain the lemma.

**Lemma 2.4.** Suppose $A, B \in \mathbb{R}^{n \times n}$ are symmetric and that $A \preceq B$. Then,

i) for any symmetric matrix $D \in \mathbb{R}^{n \times n}$,

$$DAD \preceq DBD$$

ii) if $A, B$ are positive definite matrices, then

$$B^{-1} \preceq A^{-1}.$$
Proof. To see the part (i) observe that for any vector \( x \),
\[
x^\top DADx = (Dx)^\top A(Dx) \leq (Dx)^\top B(Dx) = x^\top DBDx.
\]
The inequality follows from \( A \preceq B \). Next we prove part (ii). By (i),
\[
B^{-1/2}AB^{-1/2} \preceq B^{-1/2}BB^{-1/2} = I.
\]
So,
\[
B^{1/2}A^{-1}B^{1/2} = (B^{-1/2}AB^{-1/2})^{-1} \succeq I
\]
Now, (ii) follows by another application of (i),
\[
A^{-1} = B^{-1/2}B^{1/2}A^{-1}B^{1/2}B^{-1/2} \succeq B^{-1/2}IB^{-1/2} = B^{-1}.
\]

\(\square\)

Definition 2.5 (Matrix Norms). The trace norm (or nuclear norm) of a matrix \( A \in \mathbb{R}^{m \times n} \) is defined as follows:
\[
\|A\|_* := \text{Tr}((A^\top A)^{1/2}) = \sum_{i=1}^{\min\{m,n\}} \sigma_i,
\]
where \( \sigma_i \)'s are the singular values of \( A \). The Frobenius norm of \( A \) is defined as follows:
\[
\|A\|_F := \sqrt{\sum_{1 \leq i \leq m, 1 \leq j \leq n} A_{i,j}^2} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}.
\]

Lemma 2.6. For any matrix \( A \in \mathbb{R}^{n \times m} \) such that \( n \geq m \),
\[
\|A\|_* = \max_{\text{unitary } U} \text{Tr}(UA),
\]
where the maximum is over all unitary matrices \( U \in \mathbb{R}^{n \times n} \). In particular, \( \text{Tr}(A) \leq \|A\|_* \).

Proof. First we “square” the matrix \( A \) by adding a block of zeros to get a \( n \times n \) matrix. It is easy to see that this modification preserves the singular values of \( A \) and the right side of the above equation remains invariant. By singular value decomposition we can write
\[
A = VSW^\top
\]
where \( V, W \in \mathbb{R}^{n \times n} \) are unitary matrices and \( S \in \mathbb{R}^{n \times n} \) is the diagonal matrix of the singular values of \( A \). Now, let \( U := WV^\top \). Then,
\[
UA = WV^\top VSW^\top = WSW^\top = S.
\]
Therefore, \( \|A\|_* \leq \max_{\text{unitary } U} \text{Tr}(UA) \).

It remains to prove the other side of the equation. By von Neumann’s inequality [Mir75], for any unitary matrix \( U \) we can write
\[
\text{Tr}(UA) \leq \sum_{i=1}^{m} \sigma_i = \|A\|_* ,
\]
where \( \sigma_1, \ldots, \sigma_m \) are the singular values of \( A \). \(\square\)
Theorem 2.7 (Hoffman-Wielandt Inequality). Let $A, B \in \mathbb{R}^{n \times n}$ with singular values $\sigma_1 \leq \sigma_2 \leq \ldots \sigma_n$ and $\sigma'_1 \leq \sigma'_2 \leq \ldots \leq \sigma'_n$. Then,
\[
\sum_{i=1}^{n} (\sigma_i - \sigma'_i)^2 \leq \|A - B\|_F^2.
\]

2.3 Background in Graph Theory

For a graph $G = (V, E)$, and a set $S \subseteq V$, we define
\[
\phi_G(S) := \frac{\partial G(S)}{d_G(S)}
\]
where $\partial G(S) := |E(S, V - S)|$, is the number of edges that leaves $S$, and $d_G(S)$ is the sum of the degree (in $G$) of vertices of $S$. Note that, by definition, $d_G(v) = \partial G(v)$ for any vertex. If the graph is clear in the context we drop the subscript $G$.

The expansion of $G$ is defined as follows:
\[
\phi(G) := \min_{S \subseteq V} \min \left\{ \frac{\partial G(S)}{d_G(S)} \right\} = \min \max \{\phi_G(S), \phi_G(V - S)\},
\]
We say a graph $G$ is $\epsilon$-expander, if $\phi(G) \geq \epsilon$. Recall that in an expander graph, $\phi(G) \geq \Omega(1)$.

An (unweighted) graph $G = (V, E)$ is $k$-connected if and only if for any pair of vertices $u, v \in V$, there are at least $k$ edge disjoint paths between $u, v$ in $G$. Equivalently, $G$ is $k$-connected if for any sets $S \subset V$, $\partial(S) \geq k$.

There is a well-known theorem by Nash-Williams that gives an almost (up to a factor of 2) necessary and sufficient condition for $k$ connectivity.

Theorem 2.8 ([NW61]). For any $k$ connected graph, $G = (V, E)$, there at least $k/2$ disjoint spanning tree in $G$.

Note that any union of $k/2$ edge disjoint spanning trees is a $k/2$ edge connected graph. So, the above theorem does not give a necessary and sufficient condition for $k$ connectivity. A cycle gives a tight example for the loss of 2 in the above theorem.

Throughout the paper we may use a natural decomposition of a graph $G$ (that is not $k$ connected) into $k$ connected subgraphs.

Definition 2.9. For a graph $G = (V, E)$ a natural decomposition into $k$-connected subgraphs is defined as follows: Start with a partition $S_1 = V, S_2 = S_3 = \ldots = \emptyset$. While there is a non-empty set $S_i$ such that $G[S_i]$ is not $k$ connected, find an induced cut $S_{i,1}, S_{i,2}$ where $S_{i,1} \cup S_{i,2} = S_i$ of size less than $k$, replace $S_i$ with $S_{i,1}$ and add $S_{i,2}$ as a new set in the partition.

The following fact follows directly from the above definition.

Lemma 2.10. For any natural decomposition of a graph $G = (V, E)$ into $k$-connected subgraphs $S_1, \ldots, S_j$ and any $I \subseteq [j]$,
\[
\sum_{i_1, i_2 \in I, i_1 < i_2} |E(S_{i_1}, S_{i_2})| \leq 2(k - 1)(|I| - 1)
\]
Consequently,
\[ \sum_{i=1}^{j} \partial(S) = 2 \sum_{i_1, i_2 \in [j]: i_1 < i_2} |E(S_{i_1}, S_{i_2})| \leq 2(k - 1)(j - 1). \]

**Proof.** Let \( S = \cup_{i \in I} S_i \). A natural decomposition of the induced subgraph, \( G[S] \) into \( k \)-connected subgraphs gives exactly all set \( S_i \) where \( i \in I \). This decomposition partitions \( G[S] \) exactly \( |I| - 1 \) times and each time adds at most \( k - 1 \) new edges between the sets in the partition. ✷

Given a graph \( G = (V, E) \), and a set \( S \subseteq V \) we write \( G/S \) to denote the graph where the set \( S \) is contracted, i.e., we remove all vertices \( v \in S \) and add a new vertex \( u \) instead, for any vertex \( w \notin S \), the number of (parallel) edges between \( u \) and \( w \) is equal to \( |E(S, \{u\})| \). The following fact will be used in the paper.

**Fact 2.11.** For any \( k \)-connected graph \( G = (V, E) \) and any set \( S \subseteq V \), \( G/S \) is \( k \)-connected.

## 3 From Spectral Thin Sets to Combinatorial Thin Sets

In this section we prove simple applications of our main technical theorem using [MSS13]. In particular, we use the following main theorem of [MSS13] to prove Theorem 3.2. Then, we prove Theorems 1.11 and 1.12 as simple corollaries.

**Theorem 3.1** (Marcus, Spielman and Srivasatava [MSS13]). If \( \epsilon > 0 \) and \( x_1, \ldots, x_m \) are independent random vectors in \( \mathbb{R}^d \), with finite support such that

\[ \sum_{i=1}^{m} \mathbb{E} x_i x_i^\top \leq I \]

and for all \( i \),

\[ \mathbb{E} \|x_i\|^2 \leq \epsilon, \]

then

\[ \mathbb{P} \left[ \left\| \sum_{i=1}^{m} v_i v_i^\top \right\| \leq (1 + \sqrt{\epsilon})^2 \right] > 0. \]

We remark that the above version is slightly different from [MSS13, Thm 1.4] where they ask that \( \sum_{i=1}^{m} \mathbb{E} x_i x_i^\top = I \). If we have the above condition we can simply add many deterministic small norm vectors to make the sum equal to \( I \). Applying the original [MSS13, Thm 1.4] this modification only strengthens the conclusion of the theorem.

**Theorem 3.2.** Given a \( k \) connected graph \( G = (V, E) \), a hierarchical decomposition \( T(k, \lambda, T) \) of \( G \), and a positive definite matrix \( D \) such that \( D \preceq \square L_G \) and for any \( t \in T \),

\[ \mathbb{E}_{e \sim \mathcal{O}(t)} \chi_e^\top D^{-1} \chi_e \leq \epsilon. \]

Then, there is a \( 8/k + 4\epsilon \)-thin set \( F \subseteq E \) such that for any \( t \in T \), \( F \cap \mathcal{O}(t) \neq \emptyset \).
Theorem 3.1

Let $x_t$ be a realization of $x_t$ for each node $t \in T$. For each edge $e \in O(t)$ we let $x_t = \sqrt{k/2}A^{-1/2}x_e$ with probability $1/|O(t)|$. Now, we simply verify the conditions of Theorem 3.1. It follows that for each $t \in T(G)$,

$$\mathbb{E}\|x_t\|^2 = \sum_{e \in O(t)} \frac{k}{2|O(t)|} \chi_e A^{-1/2} \chi_e \leq \frac{k}{2|O(t)|} \sum_{e \in O(t)} \chi_e D^{-1/2} \chi_e \leq \frac{k}{2} \cdot \epsilon.$$ 

The first inequality uses part (ii) of Lemma 2.4. The last inequality follows by the assumption of the theorem.

On the other hand, using $k \leq |O(t)|$,

$$\sum_{t \in T} \mathbb{E}x_t x_t^\top = \sum_{t \in T} \sum_{e \in O(t)} \frac{k}{2|O(t)|} A^{-1/2} \chi_e \chi_e^\top A^{-1/2} \leq \sum_{e \in E} A^{-1/2} \chi_e \chi_e^\top A^{-1/2} = A^{-1/2} L_G A^{-1/2} \preceq A^{-1/2} AA^{-1/2} = I,$n

where the first inequality uses that each edge is in at most two sets $O(t)$, $O(t')$ and the last (matrix) inequality uses part (i) of Lemma 2.4 and that $L_G \preceq A$.

By Theorem 3.1 there is a realization of $x_t$ vectors, i.e., a subset $F$ of edges of $G$ such that for any $t \in T$, $F \cap O(t) \neq \emptyset$ and

$$\sum_{e \in F} k A^{-1/2} \chi_e \chi_e^\top A^{-1/2} \leq \left(1 + \sqrt{k \cdot \epsilon / 2}\right)^2 \cdot I. \quad (3)$$

Now, we upper-bound the thinness of $F$. Let $S \subset V$ be any non-empty subset of vertices and let $y = A^{1/2} 1_S$. Then,

$$y^\top \left( \sum_{e \in F} \frac{k}{2} \cdot A^{-1/2} \chi_e \chi_e^\top A^{-1/2} \right) y = \frac{k}{2} \cdot 1_S^\top A^{1/2} \left( \sum_{e \in F} A^{-1/2} \chi_e \chi_e^\top A^{-1/2} \right) A^{1/2} 1_S = \frac{k}{2} \cdot 1_S^\top L_F 1_S = \frac{k}{2} \cdot |F(S, \overline{S})|.$$ 

On the other hand,

$$y^\top I y = 1_S^\top A 1_S = 1_S^\top L_G 1_S + 1_S^\top D 1_S \leq 21_S^\top L_G 1_S.$$

where the inequality uses the assumption of the theorem. Therefore, by (3),

$$|F(S, \overline{S})| \leq \frac{2\left(1 + \sqrt{k \cdot \epsilon / 2}\right)^2}{k/2} \leq \frac{8}{k} + 4\epsilon.$$ 

Theorem 1.11 follows simply from the above theorem.

Proof of Theorem 1.11. First we construct a hierarchical decomposition of $G$, $T = T(k, 1, \ldots)$. We simply map each vertex $v$ of $G$ to a unique leaf node, $t_v$ of $T$. Then, we add the root node $t^*$ and simply connect it to all of the leafs. In other words, the height of the tree is 1.
It follows that $G[V(t^*)] = G$ and is $k$-connected. Furthermore, for each leaf $t_v$ of $T$, $O(t_v) = \mathcal{P}(t_v)$. Therefore, $T$ is a $T(k, 1, \cdot)$ hierarchical decomposition. Let $D \subseteq L_G$ be the optimum solution of \textsc{Hierarchical-CP}$(T)$. By \textbf{Theorem 1.13}, for any vertex $v$,

$$\mathbb{E}_{e \sim O(t)} \chi_e D^{-1} \chi_e \leq \frac{f_1(k)}{k}.$$ 

Therefore, by \textbf{Theorem 3.2}, there is a $\tilde{O}(1/k)$-thin set $F \subseteq E$ such that for any vertex $v$, $F \cap O(t_v) \neq \emptyset$, i.e., each vertex $v$ is incident to an edge of $F$. Therefore, $F$ is an edge cover. 

\textbf{Proof of Theorem 1.12.} First, we construct a hierarchical decomposition of $G$, $T = T(k/2h, 1/2h, T)$, then we use Theorems 1.13 and 3.2 to prove the theorem. The details of the construction is given in Algorithm 1. First of all, observe that the algorithm terminates in exactly $n - 1$ iterations of the loop, because in any non-leaf node of $T$ has exactly two children, so $|W|$ decreases by 1 in each iteration. Second, we show that $T$ is a $T(k/2h, 1/2h, T)$ hierarchical decomposition. First, we show that for any non-leaf node $t$ of $T$, $G[V(t)]$ is $k/2h$ connected. We prove this by induction. Say, $t_1, t_2$ are the two children of $t$, and by induction, $G[V(t_1)]$ and $G[V(t_2)]$ are $k/2h$-connected. But the selection of $t_2$ there are at least $k/2h$ parallel edges between $t_1, t_2$, so $G[V(t)]$ is $k/2h$-connected. Now, let $t_1 \in T$, and $t^*$ is the father of $t_1$, and $t_2$ is the other child of $t^*$. Since $t_1$ has at most $2h$ neighbors, and $E(t_1, t_2)$ has the largest number of parallel edges,

$$|O(t_1)| = |E(t_1, t_2)| \geq \frac{|\mathcal{P}(t_1)|}{2h}.$$ 

Now, let $D$ be the optimum of \textsc{Hierarchical-CP}$(T)$. By \textbf{Theorem 1.13}, for any node $t \in T$, $\mathbb{E}_{e \sim O(t)} \chi_e D^{-1} \chi_e \leq \frac{f_1(k/2h, 1/2h)}{k/2h}$. Therefore, by \textbf{Theorem 3.2}, there is a $\tilde{O}(h/k)$-thin set $F \subseteq E$ such that for any $t \in T$, $F \cap O(t) \neq \emptyset$. But, since each non-leaf node $t \in T$ has exactly two children and exactly one of them is in $T$, $F$ has at least one edge between the two children of each node $t \in T$. So, $(V,F)$ is connected. 

\begin{algorithm}
\textbf{Algorithm 1 Construction of a hierarchical decomposition for minor $h$-sparse graphs.}
\hspace*{0.02in} \textbf{Input:} A $k$-connected minor $h$-sparse graph $G$.
\hspace*{0.02in} \textbf{Output:} A hierarchical decomposition, $T = T(k/2h, 1/2h, T)$.
\begin{enumerate}
    \item For each vertex $v \in V$, add a unique leaf node of $T$ and map $v$ to it. Let $W$ be the set of these leaf nodes and $T = \emptyset$. We keep the invariant that $W$ is the nodes of $T$ that do not have a father yet, but their sub-tree is fixed, i.e., $V(t)$ is well-defined for any $t \in W$.
    \item while $|W| > 1$ do
    \item \hspace*{0.02in} Add a new node $t^*$ to $W$.
    \item \hspace*{0.02in} Let $G_{t^*}$ be the graph where for each node $t \in W$, $V(t)$ is contracted to a single vertex; identify each $t \in W$ with the corresponding contracted vertex.
    \item \hspace*{0.02in} Let $t_1$ be a vertex with at most $2h$ neighbors in $G_{t^*}$. $t_1$ exists because $G_{t^*}$ is $h$-sparse.
    \item \hspace*{0.02in} Let $t_2$ be a neighbor of $t_1$ \{t_1, t_2\} has the largest number of parallel edges among all neighbors of $t_1$. Note that $t_1, t_2$ are not necessarily vertices of $G$, so parallel edges between them do not correspond to parallel edges of $G$.
    \item \hspace*{0.02in} Make $t^*$ the father of $t_1, t_2$, remove $t_1, t_2$ from $W$. Add $t_1$ to $T$. So, $V(t^*) = V(t_1) \cup V(t_2)$.
    \item end while
\end{enumerate}
\textbf{return} $T$.
\end{algorithm}
4 Graph Partitioning

In this section we prove our main theorem 1.8 assume the main technical theorem 1.13. Before that, we need to prove a technical lemma. We say an induced subgraph $H$ of an unweighted graph $G = (V, E)$ is $\epsilon$-dense if for any $v \in V(H)$,
\[ d_H(v) \geq \epsilon \cdot d_G(v), \]
where we use $V(H)$ to denote the vertex set of $H$.

**Proposition 4.1.** Given a $k$-connected graph $G$. If any $k/20$-edge connected $1/4$-dense subgraph $H$ of $G$ satisfies $\phi(H) < \phi^*$, then for any induced subgraph $H$ of $G$,
\[ \log_2(|V(H)|) \geq \frac{3/10 - \phi_G(V(H))}{\phi^*}. \]

The above proposition says that every $k$-connected graph $G = (V, E)$ (with $n$ vertices) has an induced $\Omega(k)$-connected, $\Omega(1)$-dense subgraph $H$ that is a $\Theta(1/\log(n))$ expander. For example, for every edge $\{u, v\} \in E$, the induced graph on these vertices is a 1/2-expander. But, if there is only one edge between $u, v$ in $G$, then this induced graph is 1-connected and $O(1/k)$-dense. A tight example (up to constants) for the above proposition is the Hypercube $\{0, 1\}^k$, which is a $k$-connected, $1/k$-expander.

**Proof.** We prove the claim by induction on the number of vertices of $H$. Fix a subgraph $H = (U, F)$. Without loss of generality, assume that $\phi_G(U) < 3/10$. We consider two cases.

**Case 1: There is a vertex $v$ such that $d_H(v) \leq 7d_G(v)/20$.** We show that $\phi_G(U)$ decreases when we remove $v$ from $U$.
\[ \phi_G(U) = \frac{\partial(U - \{v\}) + d_G(v) - 2d_H(v)}{d_G(U - \{v\}) + d_G(v)} \geq \frac{\partial(U - \{v\}) + 6d_G(v)/20}{d_G(U - \{v\}) + d_G(v)} \geq \phi_G(U - \{v\}) \]

The last inequality uses that $\phi_G(U) < 3/10$. Note that if this case does not happen then $H$ is $1/4$-dense in $G$.

**Case 2: for some $S \subseteq U$, $\max\{\phi_H(S), \phi_H(U - S) < \phi^*\}$.** Let $T := U - S$. Observe that if $\phi_G(S) \leq \phi_G(U)$ or $\phi_G(T) \leq \phi_G(U)$, then we are done by induction. So assume that none of the two conditions hold. We show that $\phi_G(S), \phi_G(T) \leq \phi_G(U) + 2\phi^*$.

First, it follows from
\[ \phi_G(U) = \frac{\partial_G(S) + \partial_G(T) - \partial_H(T)}{d(S) + d(T)} \]
and $\phi_G(S) > \phi_G(U)$ that
\[ \phi_G(U) > \frac{\partial_G(T) - \partial_H(T)}{d(T)} \geq \phi_G(T) - \frac{\partial_H(T)}{d(T)} = \phi_G(T) - \phi_H(T). \]

Therefore, $\phi_G(T) \leq \phi_G(U) + \phi^*$. Similarly, we can show $\phi_G(S) \leq \phi_G(U) + \phi^*$. So, by induction,
\[ \log_2(|U|) = \log_2(|S| + |T|) \geq 1 + \log_2(\min\{|S|, |T|\}) \geq 1 + \frac{3/10 - \phi_G(U) - \phi^*}{\phi^*} = \frac{3/10 - \phi_G(U)}{\phi^*}. \]
Case 3: for some $S \subseteq U$, $\partial_H(S) < k/20$ but $\phi_H(S) > \phi^*$. We show that this case does not happen. Note that, since Case 1 does not happen, $H$ is $1/4$-dense in $G$, so for each $v \in U$,

$$d_H(v) \geq d_G(v)/4 \geq k/4,$$

where we used the $k$-connectivity of $G$.

We start with a natural decomposition of the induced graph $G[S]$ into $k/20$ connected subgraphs as defined in Definition 2.9. Let $S_1, \ldots, S_j$ be the final partition. We show that for each $1 \leq i \leq j$, $\partial_H(S_i) \geq k/10$. This already gives a contradiction, because by Lemma 2.10

$$\frac{k}{20} + (j-1) \frac{k}{20} > \partial_H(S) + \sum_{i=1}^{j} \partial_{G[S]}(S_i) = \sum_{i=1}^{j} \partial_H(S_i) \geq j \cdot \frac{k}{10}. \tag{5}$$

It remains to show that $\partial_H(S_i) \geq k/10$. For the sake of the contradiction suppose that $\partial_H(S_i) < k/10$. First, observe that $S_i$ can not be a singleton, because the induced degree of each vertex of $H$ is at least $k/4 > k/10$. We reach to a contradiction showing that $G[S_i]$ is a $1/4$-dense, $k/20$-edge connected induced subgraph of $G$ with expansion $\phi(G[S_i]) \geq \phi^*$. By construction of our natural decomposition, $G[S_i]$ is $k/20$-connected. Next, we show $G[S_i]$ is dense. For every vertex $v \in S_i$,

$$d_{G[S_i]}(v) \geq d_H(v) - \partial_H(S_i) \geq d_H(v) - k/10 \geq \frac{7d_G(v)}{20} - \frac{d_G(v)}{10} \geq d_G(v)/4.$$

where the third inequality uses that case 1 does not happen. Therefore $G[S_i]$ is $1/4$-dense.

Finally, we show that $G[S_i]$ is an expander. This is because for any set $T \subseteq S_i$,

$$\phi_{G[S_i]}(T) \geq \frac{\partial_{G[S_i]}(T)}{d_H(T)} \geq \frac{k/20}{d_H(T)} \geq \frac{\partial_H(S)}{d_H(S)} = \frac{\partial_H(S)}{d_H(S)} \geq \phi^*.$$

Therefore, $G[S_i]$ is a $k/20$-connected, $1/4$-dense and $\phi^*$-expander which is a contradiction. So, $\partial_H(S_i) \geq k/10$, which gives a contradiction by (5) \hfill \square

Our proof strategy of Theorem 1.8 is a generalization of the proof of Theorem 1.13. Ideally, we would like to build a hierarchical decomposition by repeatedly contracting $k$-connected subgraphs of $G$ which induce $\Omega(1)$-expanders. If $G$ and its minors have induced $k$-connected $\Omega(1)$-expander subgraphs, then we can simply use Theorem 1.13 to decrease the average effective resistance of the degree cuts of each contracted expander to $\widetilde{O}(1/k)$. Then, it is easy to see that the edges of small effective resistance induce an $\Omega(k)$-connected subgraph.

Unfortunately, a $k$-connected graph does not necessarily induce a $k$-connected $\Omega(1)$-expander. By the above theorem, if $k = \Theta(\log(n))$, then we can always find a $\Omega(k)$-connected subgraph that is a $1/k$-expander. These graphs are the building blocks of our hierarchical decomposition. But because we can only find weak expanders we have to apply Hierarchical-CP twice to two hierarchical decompositions of $G$, $\mathcal{T}_1, \mathcal{T}_2$. This is the reason for the $\sqrt{k}$ loss in the effective resistance of edges of $F$ in the statement of Theorem 1.8 compared to Theorem 1.12.

We start by an inductive construction of $\mathcal{T} = \mathcal{T}(k/20, 1/4, \ldots)$. The details of our construction is given in Algorithm 2. Note that in step 5 we crucially use the fact that $k \geq 10 \log(n)/3$. First of all, observe that the algorithm always terminates in at most $n - 1$ iterations of the loop, because in each iteration $|W|$ decreases by at least 1.
It is not hard to see that \(T_1\) is indeed a \(T(k/20, 1/4, .)\) hierarchical decomposition of \(G\). Fix a node \(t\). Recall that \(V(t) \subseteq V\) is the set of vertices mapped to the father of \(t\). We show that \(G[V(t)]\) is \(k/20\) connected and that \(|O(t)| \geq \frac{1}{4} \cdot |P(t)|\). Suppose, by induction, that for any node \(t'\) in \(G_t\), \(G[V(t')]\) is \(k/20\) connected. Then, since \(H_t\) is \(k/20\) connected, \(G[V(t)]\) is \(k/20\)-connected. Now, suppose \(t^*\) is the father of \(t\). Then, \(O(t)\) is simply the edges adjacent to \(t\) in \(H_{t^*}\). Since \(H_{t^*}\) is \(1/4\)-dense with respect to \(G_{t^*}\), \(|O(t)| \geq \frac{1}{4} \cdot |P(t)|\).

**Algorithm 2** Construction of a hierarchical decomposition for a \(10 \log(n)/3\)-connected graph.

**Input:** A \(k\) connected graph \(G\) where \(k \geq 10 \log(n)/3\).

**Output:** A hierarchical decomposition, \(T = T(k/20, 1/4, .)\), of \(G\).

1. For each vertex \(v \in V\) add a unique (leaf) node of \(T\) and map \(v\) to it. Let \(W\) be the set of these leaf nodes. \(\triangleright\) Throughout the algorithm, we keep the invariant that \(T\) is the nodes of \(T\) that do not have a father yet, but their corresponding sub-tree is fixed, i.e., \(V(t)\) is well-defined for any \(t \in W\).
2. **while** \(|W| > 1\) **do**
   3. Add a new node \(t^*\) to \(W\).
   4. Let \(G_{t^*}\) be the graph where for each node \(t \in W\), \(V(t)\) is contracted to a single vertex, and identify \(t\) with the corresponding contracted vertex. \(\triangleright\) \(G_{t^*}\) is \(k\) connected by Fact 2.11.
   5. Let \(H_{t^*} = (U_{t^*}, E_{t^*})\) be a \(k/20\) connected, \(1/4\)-dense subgraph of \(G_{t^*}\) such that \(\phi(H_{t^*}) \geq 1/k\). \(\triangleright\) By Proposition 4.1, \(H_{t^*}\) always exists.
   6. Let \(W = W - U_{t^*}\), and make \(t^*\) the father of all nodes of \(U_{t^*}\). \(\triangleright\) So, \(V(t^*) = \cup_{t \in U_{t^*}} V(t)\).
7. **end while**

return \(T\).

Now, let \(D_1\) be the optimum solution of Hierarchical-CP(\(T_1\)) and let \(\epsilon_1 = \frac{\ell_1(k/20, 1/4)}{k/20}\). Let

\[
F_1 := \{e \in E : \chi^T_1D^{-1}_1 \chi_e \leq \sqrt{k} \cdot \epsilon_1\}
\]

Think of the edges of \(F_1\) as the good edges. Note that if \(E = F_1\), then we are already done. By Theorem 1.13, for any \(t' \in U_t\), \(E_{t' \cap O(t')} \chi^T_1D^{-1}_1 \chi_e \leq \epsilon_1\). So, by the Markov inequality,

\[
|O(t') - F_1| \leq \frac{1}{\sqrt{k}} \cdot |O(t')|.
\]

(6)

In other words, only \(1/\sqrt{k}\) fraction of the edges of \(G\) are not good.

Now, fix a node \(t \in T_1\) that is not a leaf. Consider the graph \(H'_t = (U_t, F_1 \cap E_t)\). We say \(t\) is good if \(H'_t\) is \(k/20\) connected. If all (non-leaf) nodes of \(T_1\) are good, then \(V, F_1\) is \(k/20\)-connected, so we simply declare \(F = F_1\) and we are done. Otherwise, we construct \(T_2\) and we use another application of Hierarchical-CP to find another set of good edges \(F_2\) such that for any \(t\), \((U_t, (F_1 \cup F_2) \cap E_t)\) is \(\Omega(k)\) connected. Next, we construct \(T_2\).

We start letting \(T_2 = T_1\). For any bad (non-leaf) node \(t \in T_1\), we find a natural decomposition of \(H'_t\) into \(k/80\)-connected sets \(S_{t,1}, \ldots, S_{t,j(t)}\) as defined in Definition 2.9. Note that \(\cup_{i=1}^{j(t)} S_{t,i} = U_t\). Without loss of generality we assume that \(d_{H_t}(S_{t,j(t)}) = \max_{1 \leq i \leq j(t)} d_{H_t}(S_{t,i})\). Now, for any \(S_{t,i}\) we add a new node \(s_{t,i}\) to \(T_2\) and we make all nodes of \(S_{t,i}\) children of \(s_{t,i}\) and we make \(t\) the father of \(s_{t,i}\). Therefore, in \(T_2\), \(t\) has exactly \(j\) children. We also add all nodes \(s_{t,1}, \ldots, s_{t,j(t)-1}\) to \(T\). If \(d_{H_t}(S_{t,j(t)}) \leq \frac{1}{k}d_{H_t}(U_t)\), then we also add \(s_{t,j(t)}\) to \(T\). See Figure 4 for an example. Note that \(T\) does not have any nodes of \(T_1\).
Therefore, for any $\phi$ also in $\mathbb{H}$, it shows the tree $T$ when the new nodes $s_{t,1}, s_{t,2}, s_{t,3}$ corresponding to the sets $S_{t,1}, S_{t,2}, S_{t,3}$ are added.

Before, analyzing the optimum solution of Hierarchical-CP($T_2$), we prove two properties of $T_2$ in the following two claims.

**Claim 4.2.** $T_2$ is a $T(k/80, 1/4k, T)$ hierarchical decomposition of $G$.

**Proof.** First, since $T_1$ is a $(k/20, 1/4, \ldots)$ hierarchical decomposition, for any node $t \in T_2$ that is also in $T_1$, $G[V(t)]$ is $k/20$ connected. For any new node $s_{t,i} \in T_2$, since $s_{t,i}$ is a $k/80$ connected subgraph of $H'_t$, $G[V(s_{t,i})]$ is $k/80$ connected. On the other hand, by construction, $\phi(H_t) \geq 1/k$. Therefore, for any $s_{t,i} \in T$,

$$\frac{|O(s_{t,i})|}{|P(s_{t,i})|} = \frac{\partial H_t(S_{t,i})}{\partial G_t(S_{t,i})} = \frac{\partial H_t(S_{t,i})}{\partial G_t(S_{t,i})} \cdot \frac{d_{H_t}(S_{t,i})}{d_{G_t}(S_{t,i})} \geq \phi_H(S_{t,i}) \cdot \max_{t' \in s_{t,i}} \frac{d_{H_t}(t')}{d_{G_t}(t')} \geq \frac{1}{k} \cdot \frac{1}{4}.$$

The last inequality uses that $d_{H_t}(S_{t,i}) \leq \frac{1}{2}d_{H_t}(U_i)$ for any $s_{t,i} \in T$ and that $H_t$ is 1/4-dense with respect to $G_t$. \hfill $\square$

For a node $t \in T_1$ and a set of indices $I \subseteq [j(t)]$ let $U_I = \cup_{i \in I} S_{t,i}$.

**Claim 4.3.** For any (non-leaf) node $t \in T_1$, and $I \subseteq [j(t)]$ such that $d_{H_t}(U_I) \leq \frac{1}{2}d_{H_t}(U_t)$,

$$\frac{\partial H_t(U_I)}{\sum_{i \in I} \partial H_t(S_{t,i})} \geq \frac{1}{2(\sqrt{k} + 1)}.$$

In other words, if we contract all sets $S_{t,i}$ in $H_t$ we obtain a $\frac{1}{2(\sqrt{k} + 1)}$ expander.

**Proof.** First, since $\phi(H_t) \geq 1/k$,

$$1/k \leq \phi(H_t) \leq \phi_H(U_I) = \frac{\partial H_t(U_I)}{d_{H_t}(U_I)}.$$ 

So, we just need to show that $d_{H_t}(U_I) \geq \Omega(\sqrt{k}) \cdot \sum_{i \in I} \partial H_t(S_{t,i})$.

First, since $S_{t,1}, \ldots, S_{t,j(t)}$ is a natural decomposition of $H'_t$ into $k/80$-connected subgraphs, by Lemma 2.10,

$$\sum_{i_1, i_2 \in I, i_1 < i_2} |F_1(S_{t,i_1}, S_{t,i_2})| \leq (|I| - 1)(k/80) \leq \frac{1}{4} \sum_{i \in I} \partial H_t(S_{t,i}), \quad (7)$$

where the second inequality uses that $H_t$ is $k/20$-connected.

Figure 4: A bad node $t$ and its children, $t_1, t_2, \ldots$, in $T_1$ are illustrated in left. The right diagram shows the tree $T_2$ when the new nodes $s_{t,1}, s_{t,2}, s_{t,3}$ corresponding to the sets $S_{t,1}, S_{t,2}, S_{t,3}$ are added.
Second, by (6), for any $1 \leq i \leq j(t)$,
\[
\partial H_t(S_{t,i}) - \partial H_t(U_I) = \sum_{t' \in S_{t,i}} |O(t') - F_1| \leq \frac{1}{\sqrt{k}} \cdot \sum_{t' \in S_{t,i}} |O(t')| = \frac{1}{\sqrt{k}} \cdot d_{H_t}(S_{t,i}). \tag{8}
\]

Putting above two equations together,
\[
\sum_{i \in I} \partial H_t(S_{t,i}) = \partial H_t(U_I) + 2 \sum_{i_1,i_2 \in I \mid i_1 < i_2} |F_1(S_{t,i_1}, S_{t,i_2})| + \sum_{i \in I} \partial H_t(S_{t,i}) - \partial H_t(U_I) \\
\leq 2\partial H_t(U_I) + 2 \sum_{i \in I} \partial H_t(S_{t,i}) - \partial H_t(U_I) \\
\leq 2\partial H_t(U_I) + \frac{2}{\sqrt{k}} \sum_{t' \in S_{t,i}} d_{H_t}(S_{t,i}) = 2\partial H_t(U_I) + \frac{2}{\sqrt{k}} d_{H_t}(U_I).
\]

The first inequality uses (7) and the second inequality uses (8). Dividing both sides of the above inequality by $\partial H_t(U_I)$ we get,
\[
\frac{\sum_{i \in I} \partial H_t(S_{t,i})}{\partial H_t(U_I)} \leq 2 + \frac{2}{\sqrt{k}} \cdot \frac{1}{\phi_{H_t}(U_I)} \leq 2 + \frac{2}{\sqrt{k}} \cdot \frac{1}{\phi(H_t)} \leq 2(\sqrt{k} + 1).
\]

We are almost done! Let $D_2$ be the optimum solution of Hierarchical-CP$(T_2)$ and let $\epsilon_2 = \frac{\ell_i(k/80.1/4k)}{k/80}$. Define,
\[
F_2 := \{ e \in E : \chi^*_e D^{-1}_2 \chi_e \leq 4(\sqrt{k} + 1) \cdot \epsilon_2 \}
\]

For any non-leaf node $t \in T_1$, let $H''_t = (U_t, (F_1 \cup F_2) \cap E_t)$. We claim that $H''_t$ is $k/80$-connected, and hence $(V, F_1 \cup F_2)$ is $k/80$-connected. Since each set $S_{t,i}$ is a $k/80$-connected subgraph of $H'_t$ and $H'_t$ is a subgraph of $H''_t$, all we need to show is that for any set of indices $I \subseteq [j(t)]$, $\partial H''_t(U_I) \geq k/80$

Without loss of generality, assume $d_{H_t}(U_I) \leq \frac{1}{2}d_{H_t}(U_t)$. Since $s_{t,j(t)} \in T$ only if $d_{H_t}(S_{t,j(t)}) \leq \frac{1}{2}d_{H_t}(U_I)$, for all $i \in I$, $s_{t,i} \in T$. Therefore, by Theorem 1.13, $i \in I. iEe \sim O(s_{t,i})\chi_e^{-1}D_2^{-1}\chi_e \leq \epsilon_2$. So, by the Markov inequality,
\[
|O(s_{t,i}) - F_2| \leq \frac{1}{4(\sqrt{k} + 1)} \cdot |O(s_{t,i})|.
\]

So,
\[
\partial H_t(U_I) - \partial H''_t(U_I) = \sum_{i \in I} |O(s_{t,i}) - F_2| \leq \frac{1}{4(\sqrt{k} + 1)} \cdot \sum_{i \in I} |O(s_{t,i})| = \frac{1}{4(\sqrt{k} + 1)} \cdot \sum_{i \in I} \partial H_t(S_{t,i}).
\]

But, by Claim 4.3, the RHS of the above inequality is at most $\partial H_t(U_I)/2$. Therefore, since $H_t$ is $k/20$-connected,
\[
\partial H''_t(U_I) \geq \frac{1}{2} \partial H_t(U_I) \geq k/40.
\]

So, $H''_t$ is $k/80$-connected.

Letting $D = \frac{1}{2}D_1 + \frac{1}{2}D_2$ and $F = F_1 \cup F_2$ we get that $(V, F)$ is a $k/80$-connected, $D \leq L_G$, and for any edge $e \in F$,
\[
\chi^*_e D^{-1} \chi_e \leq 2\sqrt{k} \cdot \epsilon_1 + 8(\sqrt{k} + 1) \cdot \epsilon_2 = \tilde{O}(1/\sqrt{k}).
\]

This completes the proof of Theorem 1.8.
5 The Dual of Hierarchical-CP

In the rest of this document we upper-bound the value of Hierarchical-CP.

Lemma 5.1. The optimum of the above semidefinite program is equal to

$$\sup_{U, X, \pi: E \to |E|} \frac{\sum_{t \in T} \left| \mathcal{O}(t) \right| (UX\chi_e)\pi(e)}{\sum_{e \in E} \|X\chi_e\|_1}$$

where the supremum is over all unitary matrices $U \in \mathbb{R}^{2n \times 2n}$, cut metrics $X: V \to \mathbb{R}^{2n}$ and all permutations $\pi$ of the edges of $G$.

Proof. First, we show Hierarchical-CP satisfies the Slater’s condition, i.e., that CP has a non-empty interior. It is easy to see that $D = \frac{1}{2}L_G + \frac{1}{3n^2}J$ is a positive definite matrix that satisfies all constraints strictly. In particular, since $G$ is connected, for any set $S$, $1_S^T L_G 1_S \geq 1$, so $\frac{1}{3n^2} 1_S^T J 1_S \leq \frac{1}{3} < \frac{1}{2} 1_S^T L_G 1_S$.

Therefore, $1_S^T D 1_S < 1_S^T L_G 1_S$ for all $S$. Hence, the Slater’s condition is satisfied, and the strong duality is satisfied and the primal optimum is equal to the Lagrangian dual (see [BV06, Section 5.2.3] for more information).

For every $t \in T$ we associate a Lagrange multiplier $\lambda_t$ corresponding to the first constraints, and for every set $S$ we associate a non-negative Lagrange multiplier $y_S$ corresponding to the second constraints of the Hierarchical-CP. The Lagrange function is defined as follows:

$$g(\lambda, y) = \inf_{D \succeq 0} \mathcal{E} - \sum_{t \in T} \frac{\lambda_t}{|\mathcal{O}(t)|} \left( \sum_{e \in \mathcal{O}(t)} \chi_e^T D^{-1} \chi_e - \mathcal{E} \right) + \sum_{S \subset V} y_S (1_S^T D 1_S - 1_S^T L_G 1_S)$$

First of all, differentiating $g(\lambda, y)$ w.r.t. $\mathcal{E}$ we obtain that

$$\sum_{t \in T} \lambda_t = 1. \quad (10)$$

Let

$$A := \sum_{t \in T} \frac{\lambda_t}{|\mathcal{O}(t)|} \left( \sum_{e \in \mathcal{O}(t)} \chi_e \chi_e^T \right) \quad \text{and} \quad Z := \sum_{\emptyset \subset S \subset V} y_S 1_S 1_S^T.$$ 

Note that by definition $A$ and $Z$ are symmetric PSD matrices. The Lagrange dual function simplifies to,

$$g(A, Z) = \inf_{D \succ 0} A \cdot D^{-1} + Z \cdot D - Z \cdot L_G,$$

subject to $\sum_t \lambda_t = 1$. Now, we solve the above function for $D$. First, we assume that $A$ and $Z$ is non-singular. This is without loss of generality by the continuity of $g(.)$ and because the assumption $\sum_t \lambda_t = 1$ can be satisfied by adding arbitrarily small perturbations. Differentiating with respect to $D$ we obtain

$$D^{-1} AD^{-1} = Z.$$
Since, $A,D$ are non-singular there is a unique solution to the above equation,

$$D = Z^{-1/2}(Z^{1/2}AZ^{1/2})^{1/2}Z^{-1/2}$$

We refer interested readers to [SLB74] to solve the above matrix equation. Using

$$D^{-1} = Z^{1/2}(Z^{1/2}AZ^{1/2})^{-1/2}Z^{1/2},$$

we have

$$A \bullet D^{-1} + Z \bullet D = \text{Tr}(AZ^{1/2}(Z^{1/2}AZ^{1/2})^{-1/2}Z^{1/2}) + \text{Tr}(Z^{1/2}(Z^{1/2}AZ^{1/2})^{1/2}Z^{-1/2})$$

$$= 2 \text{Tr}((Z^{1/2}AZ^{1/2})^{1/2}).$$

Therefore,

$$g(A,Z) = 2 \text{Tr}((Z^{1/2}AZ^{1/2})^{1/2}) - Z \bullet L_G$$

Let $\mathcal{E}^*$ be the optimum value of Hierarchical-CP. By the strong duality,

$$\mathcal{E}^* = \sup_{\lambda, y \geq 0} g(A,Z) = \sup_{\lambda, y \geq 0} 2 \text{Tr}((Z^{1/2}AZ^{1/2})^{1/2}) - Z \bullet L_G.$$ 

It remains to characterize values of $\lambda, y$ that maximizes the above function. Let $B \in \mathbb{R}^{n \times |E|}$ where for each edge $e \in \mathcal{O}(t)$ there is a column $\sqrt{\lambda_t} \chi_e$ in $B$. Observe that $A = BB^\top$. Furthermore eigenvalues of $Z^{1/2}AZ^{1/2} = Z^{1/2}BB^\top Z^{1/2}$ are the same as the eigenvalues of $B^\top Z B$. Therefore,

$$\mathcal{E}^* = \sup_{\lambda, y \geq 0} \frac{2 \text{Tr}((B^\top ZB)^{1/2}) - Z \bullet L_G}{Z \bullet L_G} \leq \sup_{\lambda, y \geq 0} \text{Tr}((B^\top ZB)^{1/2})^2.$$ 

To see the second inequality note that $Z \bullet L_G$ scales linearly with $y$ and $\text{Tr}((B^\top ZB)^{1/2})$ scales with $\sqrt{\gamma}$. We are almost done! Write $Z = X^\top X$ where $X \in \mathbb{R}^{2n \times n}$ and each row of $X$ corresponds to a vector $y_S 1_S$ for a set $S \subseteq V$. Observe that $X$ defines a cut metric on the vertices of $G$. Let $X_v$ be the coordinate of $X$ in that metric, i.e., $X_v$ is the column $v$ of $X$. By the definition of the nuclear norm, $\text{Tr}((B^\top ZB)^{1/2})^2 = \|XB\|_*^2$, so

$$\mathcal{E}^* = \sup_{X, \lambda} \frac{\|XB\|_*^2}{\sum_{\{u,v\} \in E} \|X_u - X_v\|_2^2}$$

$$= \sup_{X, U, \pi: E \rightarrow |E|}, \lambda \left( \frac{\sum_{t \in T} \sum_{e \in \mathcal{O}(t)} \sqrt{\lambda_t}/|\mathcal{O}(t)| (U X \chi_e)_{\pi(e)}^2}{\sum_{\{u,v\} \in E} \|X_u - X_v\|_1} \right)^2$$

(11)

The supremum in the RHS is over all cut matrices $X$, unitary matrices $U$ and all permutations $\pi: E \rightarrow |E|$. In the last equality we use Lemma 2.6 and that $X$ is a cut metric. Note that, since $XB$ has rank $n - 1$, only $n - 1$ entries of the diagonal of $UXB$ is non-zero.

Finally, using the Cauchy-Schwarz inequality we can write,

$$\mathcal{E}^* = \sup_{X, U, \pi: E \rightarrow |E|} \left( \frac{\sum_{t \in T} \lambda_t}{\sum_{t \in T} |\mathcal{O}(t)|} \left( \frac{1}{|\mathcal{O}(t)|} \left( \sum_{e \in \mathcal{O}(t)} (U X \chi_e)_{\pi(e)}^2 \right)^2 \right) \right)$$
Theorem 1.7

Theorem 1.7

$S$ is at least $\Omega(\ldots)$ for an example of Lemma 5.1.

The other side of the above equality follows by letting

$$\lambda_t := \frac{|O(t)|^{-1} \left( \sum_{e \in O(t)} (UX_e)_{\pi(e)} \right)^2}{\left( \sum_{t \in T} \sum_{e \in O(t)} |O(t)|^{-1} (UX_e)_{\pi(e)} \right)^2},$$

which satisfies $\sum_t \lambda_t = 1$.

**Proof of Theorem 1.7.** We prove a stronger result. We show that for any integer $h > 2$, the optimum of Average-CP for the graph illustrated in Figure 2 is at least $\Omega\left(\frac{h^2}{(h+k)^2}\right)$. In words, we show that for any matrix $0 < D \leq L_G$ there is a cut $(S, \bar{S})$ such that

$$\mathbb{E}_{e \sim E(S, \bar{S})} UX_e D^{-1} UX_e \geq \frac{h^2}{8(h+k)^2}.$$ 

Let $n = 2^h$ (so $G$ has $n + 1$ vertices). Consider the mapping $X : V \rightarrow \mathbb{R}^{2^n}$ where $X_i := 1_{[i]}$ for $i \geq 1$, i.e., $X_i$ is 1 in the coordinates 1 to $i$ and it is zero otherwise, and let $X_0 = 0$. It follows that

$$\sum_{\{i,j\} \in E} \|X_i - X_j\|_1 = n \cdot k + n \cdot h.$$ 

Let $U \in \mathbb{R}^{2^n \times 2^n}$. The first $n - 1$ rows of $U$ are defined as follows: For each $0 < i \leq h$ and each even integer $0 \leq j < 2^{h-i}$ there is a row $r$ such that

$$U_{r,c} = \begin{cases} 
2^{-((i+1)/2)} & \text{if } j \cdot 2^i \leq c < (j + 1) \cdot 2^i, \\
-2^{-((i+1)/2)} & \text{if } (j + 1) \cdot 2^i \leq c < (j + 2) \cdot 2^i, \\
0 & \text{otherwise.} 
\end{cases}$$

Now, for any $0 \leq i < h$ and odd $0 \leq j < 2^{h-i}$, we let $\pi(\{j \cdot 2^i, (j + 1) \cdot 2^i\}) = r$. (Note that if $i = 1$ then $G$ has $k$ copies of the edges $\{j, j + 1\}$. In this case we simply map an arbitrary one of those edges to $r$). Let $E'$ be the set of $2^h - 1$ edges that we just defined their map under $\pi(\cdot)$. The rest of the rows of $U$ are defined to be orthogonal to the first $2^h - 1$ rows. Also, the value $\pi(\cdot)$ for the rest of edges of $G$ is defined arbitrarily. By definition, $U$ is unitary and $\pi(\cdot)$ is a permutation. See Table 1 for an example of $U$ for $n = 4$.

| $2^{-1}$ | $2^{-1}$ | $-2^{-1}$ | $-2^{-1}$ | $0 \ldots 0$ |
| $2^{-1/2}$ | $-2^{-1/2}$ | $0$ | $0$ | $0 \ldots 0$ |
| $0$ | $0$ | $2^{-1/2}$ | $-2^{-1/2}$ | $0 \ldots 0$ |
| $\ldots$ | | | | |

Table 1: An example of a unitary matrix $U$ for $n = 4$ in proof of Theorem 1.7.

By the above construction for each edge $\{j \cdot 2^i, (j + 1) \cdot 2^i\} \in E'$,

$$(UX_j \cdot 2^i - UX_{(j+1) \cdot 2^i})_{\pi(\{j \cdot 2^i, (j + 1) \cdot 2^i\})} = \frac{2^i}{2^{-(i+1)/2}} = 2^{(i-1)/2}. \quad (12)$$

By Lemma 5.1 (in particular (11)), the optimum of Average-CP is at least,

$$\max_{\lambda} \frac{\left( \sum_{\{i,j\} \in E} \sqrt{\gamma_{i,j}} (UX_i - UX_j) \right)^2}{\sum_{\{i,j\} \in E} \|X_i - X_j\|_1} \geq \max_{\lambda} \frac{\left( \sum_{\{j \cdot 2^i, (j+1) \cdot 2^i\} \in E'} \sqrt{\gamma_{j \cdot 2^i, (j+1) \cdot 2^i}} \cdot 2^{(i-1)/2} \right)^2}{n \cdot k + n \cdot h}.$$
In the above \( \gamma_{ij} := \sum_{S_i \in S_j} \lambda(S_i, S_j) \) where \( \lambda(S_i, S_j) \) is a probability distribution on all cuts of \( G \) and the maximum is taken over all choices of \( \lambda \).

Now, let \( \lambda(S_i, S_j) = 1/n \) for every cut \( \{0, 1, \ldots, \ell\} \) for all \( 0 \leq \ell \leq n-1 \). Then, for any edge \( \{j \cdot 2^i, (j+1) \cdot 2^i\} \),

\[
\gamma_{ij} = \sum_{j \cdot 2^i \leq \ell < (j+1) \cdot 2^i} \frac{1}{n \cdot |E(\{0, \ldots, \ell\}, \{\ell+1, \ldots, n+1\})|} \geq \frac{2^i}{n(h+k)}
\]

Therefore, the optimum of Average-CP is at least,

\[
\left( \sum_{i=0}^{h-1} \sum_{0 \leq j < 2^{h-i}} \sqrt{\frac{2^i}{n(h+k)}} \cdot 2^{(i-1)/2} \right)^2 \geq \left( \sum_{i=0}^{h-1} \frac{n \cdot 2^{-i-1} \cdot \sqrt{\frac{2^i}{n(h+k)}} \cdot 2^{(i-1)/2}}{n(h+k)} \right)^2
\]

\[
= \frac{2^{-3/2} \cdot h \cdot \sqrt{n}}{n(h+k)} = \frac{h^2}{8(h+k)^2}
\]

\( \square \)

### 6 Upper-bounding the Dual of Hierarchical-CP

In the rest of the paper we prove the following theorem which together with Lemma 5.1 proves Theorem 1.13.

**Theorem 6.1.** For any \( k \)-connected graph \( G \) and any hierarchical decomposition \( T = T(k, \lambda, T) \) of \( G \), any cut metric \( X : V \rightarrow \mathbb{R}^{2^n} \), any unitary matrix \( U \in \mathbb{R}^{2^n \times 2^n} \) and a any permutation \( \pi : E \rightarrow [||E||] \),

\[
\frac{\sum_{t \in T} \frac{1}{|O(t)|} \left( \sum_{e \in O(t)} (UX_e)^{\pi(e)} \right)^2}{\sum_{e \in E} \|X_e\|_1} \leq \frac{f_1(k, \lambda)}{k}.
\]

Note that since \( X \) is a cut metric for any pair of vertices \( u, v \in V \), \( \|X_e\|^2 = \|X_e\|_1 \), so, throughout the proof, we use either of the two norms.

Before going into the details of the proof of the above theorem, we upper-bound the above ratio for a hierarchical decomposition of the graph illustrated in Figure 2. Let \( T \) be the the tree shown in Figure 3 and \( T = \{1, 2, \ldots, 2^h\} \). Let \( U \), \( X \) and \( \pi(.) \) be the unitary matrix, the \( L_1 \) mapping and the permutation constructed in the proof of Theorem 6.1. Let us estimate \( \sum_{e \in O(t)} (UX_e)^{\pi(e)} \) for nodes \( 2^i \in T \); the rest of the terms can be estimated similarly. For node \( 2^i \), \( O(2^i) \) has \( k \) copies of the edge \( \{2^{i-1}, 2^i\} \) and for each \( 1 \leq j \leq i \), it has an edge \( \{2^{i-j}, 2^i\} \). By (12), for each edge \( e = \{2^{i-j}, 2^i\} \), \( (UX_e)^{\pi(e)} = 2^{(j-1)/2} \). Therefore, for any node \( 2^i \), \( \sum_{e \in O(2^i)} (UX_e)^{\pi(e)} \) is a geometric sum and we can approximate it with the largest projection on \( U \), i.e., \( \max_{e \in O(2^i)} (UX_e)^{\pi(e)} \). Therefore,

\[
\frac{\sum_{t \in T} \frac{1}{|O(t)|} \left( \sum_{e \in O(t)} (UX_e)^{\pi(e)} \right)^2}{\sum_{t \in T} \frac{1}{|O(t)|} \max_{e \in O(t)} (UX_e)^{\pi(e)}^2} \leq \frac{\sum_{t \in T} \frac{1}{|O(t)|} \max_{e \in O(t)} (UX_e)^{\pi(e)}^2}{\frac{1}{k} \sum_{e \in E} \|X_e\|^2}.
\]

27
In the second inequality we use the crucial fact that each edge $e$ is contained in $O(t)$ for at most two nodes of $T$ and that $|O(t)| \geq k$ for all $t$.

Generalizing the above argument, it is easy to see that if all nodes of $T$ satisfy the following inequality, then we are done.

$$\left(\mathbb{E}_{e \sim O(t)} (UX\chi_{e})_{\pi(e)}\right)^2 \lesssim \frac{1}{k} \cdot \mathbb{E}_{e \sim O(t)} \|X\chi_{e}\|_2^2.$$ 

Say a node of $T$ is bad if it does not satisfy the above inequality and let $T'$ be the set of bad nodes. All we need to do, is to upper-bound

$$\sum_{t \in T'} \frac{1}{|O(t)|} \left(\sum_{e \in O(t)} (UX\chi_{e})_{\pi(e)}\right)^2$$

by the sum of the $L_1$ length of all edges of $G$. It turns out that, for each bad node, edges of $O(t)$ must have “similar” projection length on the unitary matrix $U$.

Our main tool to upper-bound the above quantity is by covering the space with (disjoint) balls. Say we have an $L_1$ ball $B(X_u, r)$ around a vertex $u$ such that there is a vertex $v$ where $X_v \notin B(X_u, r)$. It follows that the sum of the $L_1$ length of edges in this ball is at least $k \cdot r$. This is because by $k$-edge connectivity of $G$ there are at least $k$ edge disjoint paths from $X_v$ to the outside of the ball. By triangle inequality the sum of the $L_1$ length of edges along each of these paths is at least $r$.

In the first step of our proof, we show that if many edges of $G$ have “similar” projection on the unitary operator $U$, then we can find many disjoint $L_1$ balls in the space each having a vertex of $G$ as a center.

**Lemma 6.2.** Given $F \subseteq E$ a mapping $Y : V \to \mathbb{R}^n$ and a permutation $\pi : E \to |E|$ such that,

$$\Upsilon := \left(\mathbb{E}_{e \sim F} (Y\chi_{e})_{\pi(e)}\right)^2 \geq \alpha \cdot \mathbb{E}_{e \sim F} \|Y\chi_{e}\|_2^2,$$

for $\alpha > 0$. Then, for any $0 < \epsilon < 1/2$ there are $b$ disjoint $L_2^2$ balls $B_1, \ldots, B_b$ with radius $r$ such that the center of each ball is an endpoint of an edge in $F$, $b \geq \alpha|F|/C_1(\epsilon)$ and

$$r \cdot b \geq \frac{\alpha^\epsilon \cdot \Upsilon \cdot |F|}{C_1(\epsilon)}.$$ 

where $C_1(.)$ is a polynomial function of $1/\epsilon$.

Before getting into the proof of the above lemma, as a simple application, we use the above lemma to upper-bound Max-CP by $O(\log^2(n)/k^{1-\epsilon})$.

**Definition 6.3.** For $c > 1$, we say a set $F \subseteq E$ of edges is $c$-similar if for any two edges $e, f \in F$,

$$\frac{(UX\chi_{e})_{\pi(e)}^2}{(UX\chi_{f})_{\pi(f)}^2} < c$$ and $$\frac{\|X\chi_{e}\|_2^2}{\|X\chi_{f}\|_2^2} < c.$$ 

Given $U, X, \pi$, by a simple bucketing argument we can show that there is a 2-similar set $F$ of the edges such that

$$\sum_{e \in F} (UX\chi_{e})_{\pi(e)}^2 \gtrsim \frac{1}{\log^2(n)} \sum_{e \in E} (UX\chi_{e})_{\pi(e)}^2.$$ 

28
Say for all \( e \in F \), \( (UX\chi_e)^2 \sim c_1 \) and \( \|X\chi_e\|^2 \sim c_2 \). Letting \( Y = UX \) (and noting that since \( X \) is a \( L_1 \) metric, \( \|Y_u - Y_v\| = \|X_u - X_v\| \) for all \( u, v \in V \)), for \( \varepsilon > 0 \) and \( \alpha = (E(Y\chi_e)^2)/E\|Y\chi_e\|^2 \), we get at least \( b \) disjoint balls of radius \( r \) such that

\[
b \cdot r \gtrsim \alpha^\varepsilon \cdot \left( E(Y\chi_e)^2 \right)^2 \geq \frac{\alpha^\varepsilon}{\log^2(n)} \sum_{e \in E}(UX\chi_e)^2.
\]

where we used the fact that \( F \) is 2-similar. Therefore,

\[
\frac{\sum_{e \in E}(UX\chi_e)^2}{\sum_{e \in E}\|X\chi_e\|^2} \leq \frac{\log^2(n) \sum_{e \in E}(UX\chi_e)^2}{\sum_{e \in E}\|X\chi_e\|^2} \leq \frac{\log^2(n) \sum_{e \in E}(UX\chi_e)^2}{\|X\chi_e\|^2} \leq \log^2(n) \cdot \alpha.
\]

On the other hand, since the sum of the \( L_1 \) length of the edges in each ball is at least \( r \cdot k \) we get,

\[
\frac{\sum_{e \in E}(UX\chi_e)^2}{\sum_{e \in E}\|X\chi_e\|^2} \leq \frac{\log^2(n) \sum_{e \in E}(UX\chi_e)^2}{k \cdot b \cdot r} \leq \frac{\log^2(n)}{k\alpha^\varepsilon}.
\]

So, \( \text{Max-CP} \leq \log^2(n) \cdot \max \{ \alpha_1, \frac{1}{\alpha_k} \} \leq \frac{\log^2(n)}{k^{1/(1+\varepsilon)}} \).

**Proof of Lemma 6.2.** For a radius \( r > 0 \) run the following greedy algorithm. Scan the endpoints of the edges in an arbitrary order, for each point \( Y_u \), if the \( L_2 \) ball \( B(Y_u, r) \) doesn’t touch the balls we already included, add \( B(Y_u, r) \). Suppose we manage to draw \( b \) balls. We say the algorithm succeeds if both of the lemma’s conclusion are satisfied. In the rest of the proof we show that this algorithm always succeeds for some value of \( r \).

Perhaps after relabeling the coordinates of vectors \( Y_v \), we assume that \( 1 \leq \pi(e) \leq |F| \) for all \( e \in F \). Let \( A \in \mathbb{R}^{m \times |F|} \) where for each \( 1 \leq i \leq |F| \), the \( i \)-th column of \( A \) is \( Y\chi_{\pi^{-1}(i)} \). Note that by definition of \( A \) \( \text{Tr}(A) = \sum_{e \in F}(Y\chi_e)^2 \). Let \( \sigma_1, \ldots, \sigma_{|F|} \) be the singular values of \( A \).

In the next claim we show that if the above algorithm finds \( b \) balls for a value of \( r \) that implies an upper bound on the singular values of \( A \).

**Claim 6.4.** Given \( r > 0 \) suppose that the greedy algorithm finds \( b \) disjoint balls of radius \( r \). Then

\[
r \geq \frac{1}{8|F|} \sum_{i=b+1}^{|F|} \sigma_i^2.
\]

**Proof.** We construct a low rank matrix \( C \in \mathbb{R}^{n \times |F|} \) then we use Theorem 2.7 to prove the claim. Let \( Y_{w_1}, \ldots, Y_{w_b} \) be the centers of the chosen balls. Then, for any endpoint \( v \) of an edge in \( F \), let \( c(v) \) be the closest center to \( Y_v \), i.e.,

\[
c(v) := \arg\min_{w_i} \|Y_{w_i} - Y_v\|^2_2
\]

We construct a matrix \( C \) of the same dimension as \( A \); the \( i \)-th column of \( C \) is defined as follows: say the \( i \)-th column of \( A \) is \( Y_{u_i} - Y_{v_i} \) for \( \{u_i, v_i\} \in E \), we let \( i \)-th column of \( C \) be \( Y_{c(u_i)} - Y_{c(v_i)} \). By definition, \( \text{rank}(C) \leq b \).

29
First, notice that

\[ \|A - C\|_F^2 = \sum_{i=1}^{|F|} \left\| (Y_{u_i} - Y_{c(u_i)}) - (Y_{c(u_i)} - Y_{c(v_i)}) \right\|_2^2 \]

\[ \leq \sum_{i=1}^{|F|} \left( \|Y_{u_i} - Y_{c(u_i)}\|_2^2 + \|Y_{v_i} - Y_{c(v_i)}\|_2^2 \right) \]

\[ \leq \sum_{i=1}^{|F|} 2 \|Y_{u_i} - Y_{c(u_i)}\|_2^2 + 2 \|Y_{v_i} - Y_{c(v_i)}\|_2^2 \leq 8r \cdot |F|, \]

where the first inequality follows by the triangle inequality and the last inequality follows by definition of greedy algorithm, i.e., \( \|Y_{u_i} - Y_{c(v_i)}\|_2^2 \leq 2r \) for all points. Therefore, by Theorem 2.7,

\[ 4r \cdot |F| \geq \|A - C\|_F^2 \geq \sum_{i=b+1}^{|F|} \sigma_i^2, \]

where the second inequality uses the fact that \( \text{rank}(C) \leq b \).

All we need to show is that there is a value of \( b \geq \alpha/C_1(\epsilon) \) such that \( \frac{b}{8|F|} \sum_{i=b+1}^{|F|} \sigma_i^2 \geq \frac{\alpha^\epsilon \cdot \Upsilon}{C_2(\epsilon)|F|^{1+\epsilon}}. \)

**Claim 6.5.** For any \( 0 < \epsilon < 1 \) and \( b_0 := \alpha/C_1(\epsilon) \), there is an integer \( b \geq b_0 \) such that

\[ \frac{b}{8|F|} \sum_{i=b+1}^{|F|} \sigma_i^2 \geq \frac{\alpha^\epsilon \cdot \Upsilon}{C_2(\epsilon)|F|^{1+\epsilon}}. \]

**Proof.** Let \( z := \max_{b \geq b_0} \frac{b}{8|F|} \sum_{i=b+1}^{|F|} \sigma_i^2. \) We will lower-bound \( z \). First, by definition of \( z \), for all \( b \geq b_0 \),

\[ \frac{z}{|F|^{1-\epsilon} \cdot b^{1-\epsilon}} \geq \frac{b^\epsilon}{8|F|^{1+\epsilon}} \sum_{i=b+1}^{|F|} \sigma_i^2. \] \hspace{1cm} (14)

Note that the above inequality do not bound the first \( b_0 \) eigenvalues of \( A \). We can, however, use the Frobenius norm of \( A \) to bound those eigenvalues.

\[ \frac{1}{|F|} \sum_{i=1}^{|F|} \sigma_i^2 = \frac{\|A\|_F^2}{|F|} = \mathbb{E}_{e \sim F} \|Y_{e} \chi_{e}\|_2^2 \leq \frac{1}{\alpha} \left( \mathbb{E}_{e \sim F} (Y_{e} \chi_{e})_{\pi(e)}^2 \right) = \Upsilon / \alpha. \] \hspace{1cm} (15)

where the first inequality uses (13).

Let \( \beta > 0 \) be a parameter that we fix later. Summing up (14) for all \( b_0 \leq b \leq |F| \) and \( \beta \) times (15), we get

\[ \sum_{i=1}^{|F|} \left( \beta + \int_{x=b_0}^i (x-1)^\epsilon dx \right) \cdot \frac{\sigma_i^2}{|F|} \leq \beta \cdot \frac{\Upsilon}{\alpha} + \frac{z}{|F|^{1-\epsilon}} \int_{x=b_0}^{|F|} \frac{dx}{(x-1)^{1-\epsilon}}. \]
Lemma 2.6. It is sufficient to run the greedy algorithm.

On the other hand, by Lemma 6.2, where for \( x \in \mathbb{R}, [x]_+ = \max\{x, 0\} \).

Therefore, by Cauchy-Schwarz

\[
\pi \leq \left( \frac{1}{|F|} \cdot \sum_{i=1}^{|F|} \sigma_i \right)^2 \leq \left( \sum_{i=1}^{|F|} \left( \beta + \frac{\sum_i ((i-1)^{1+\epsilon} - (b_0-1)^{1+\epsilon}) + \epsilon}{16|F|^{1+\epsilon}} \right) \sigma_i^2 \right) \cdot \left( \sum_{i=1}^{|F|} \frac{1}{\beta + \sum_j ((j-1)^{1+\epsilon} - (b_0-1)^{1+\epsilon}) + \epsilon} \right) \leq \left( \frac{\beta \cdot \pi}{\alpha} + \frac{z}{\epsilon} \right) \cdot \frac{16(3+1/\epsilon)}{\beta^{1+\epsilon} |F|^{1+\epsilon}}.
\]

To see the last inequality we need to do algebraic manipulations. The first term in the RHS follows from (16). We obtain the last inequality asserting \( b_0 \leq 1 + \beta^{1/(1+\epsilon)} |F|^{\epsilon/(1+\epsilon)} \). In particular,

\[
\sum_{i=1}^{|F|} \beta + \frac{\sum_i ((i-1)^{1+\epsilon} - (b_0-1)^{1+\epsilon}) + \epsilon}{16|F|^{1+\epsilon}} \leq \frac{b_0 - 1}{\beta |F|} + \sum_{i=1}^{\infty} \sum_{j=(b_0-1)i^{1/(1+\epsilon)}+1}^{1} \frac{16}{i \cdot \beta \cdot |F|} \leq \frac{b_0 - 1}{\beta \cdot |F|} \leq \frac{16(3+1/\epsilon)}{\beta^{1+\epsilon} |F|^{1+\epsilon}}.
\]

where in second inequality we used

\[
(i + 1)^{1+\epsilon} - i^{1+\epsilon} \leq i^{1+\epsilon} ((1 + 1/i)^{1+\epsilon} - 1) \leq i^{1+\epsilon}.
\]

Now, the claim follows directly from (18). Letting \( \beta = \frac{\alpha^{1+\epsilon}|F|}{(96+36/\epsilon)^{1+\epsilon}} \), we obtain,

\[
z \geq \frac{\epsilon \cdot \beta^{1+\epsilon} \cdot |F|^{1+\epsilon} \cdot \pi}{16(3+1/\epsilon)} - \frac{\epsilon \cdot \beta \cdot \pi}{\alpha} \geq \frac{\alpha \epsilon \cdot \pi \cdot |F|}{(96/\epsilon + 32/\epsilon^2)^{1+\epsilon}}.
\]

The lemma follows by letting \( C_1(\epsilon) = (96/\epsilon + 32/\epsilon^2)^{1+\epsilon} \).

Observe that the above claim implies Lemma 6.2. It is sufficient to run the greedy algorithm with the infimum of values of \( r \) such that the greedy algorithm return at most \( b \) balls.

In the rest of the proof we fix a cut metric \( X \) on vertices of \( G \), a unitary matrix \( U \) and a permutation \( \pi : E \to [|E|] \) (we may assume that this triplet maximize the ratio in (9)).
6.1 Construction of Bag of Balls

A bag of balls, Bag(δ, b) is a set of b disjoint L₁ balls of radius δ such that the center of each ball is a point X_v for some v ∈ V. For β > 0, a β-compact bag of balls Bag(δ, Δ, b) is a bag of balls with centers X_{u_1}, ..., X_{u_b} such that diam({X_{u_1}, ..., X_{u_b}}) ≤ Δ and

\[ β \cdot Δ ≤ b \cdot δ. \]

We will construct β-compact bag of balls for a large constant β. For a hierarchical decomposition \( T \) and \( β > 0 \), a bag of balls Bag(δ, b) is β-assigned to a node \( t \) ∈ \( T \), if

\[ β \cdot |O(t)| ≤ b, \]

and for each ball \( B(X_u, δ) \in Bag(δ, b), u ∈ V(t) \) and there is an edge \( \{u, v\} ∈ O(t) \) such that \( \|X_u - X_v\|_1 < δ \). We keep the convention of writing Bag_t(δ, b_t) for a bag of balls assigned to a node \( t \). We will construct β-assigned bag of balls for \( β \gg 1/k \), where \( k \) is the connectivity of \( G \).

A homogeneous family of bags of balls is a set FBag(δ) = {Bag(δ, b)} of bags of all balls such that all bags in all bags of FBag(δ) are disjoint. We abuse notation and write a ball \( B ∈ FBag(δ) \) if there is a Bag(δ, b) ∈ FBag(δ) such that \( B ∈ Bag(δ, b) \). Note that two distinct bags in FBag(δ) may have distinct number of balls.

A homogeneous family of β-compact bags of balls FBag(δ, Δ) is a homogeneous family of β-compact bags of balls Bag(δ, Δ, b). For a set \( T ∈ T \), a homogeneous family of β-assigned bags of balls FBag(δ, S) is a homogeneous family of β-assigned bags of balls, Bag(δ, b), where each bag is assigned to a unique node of \( T \).

To upper-bound the value of the dual we need to find a sequence of homogeneous families of bags of balls with a geometrically decaying radii. In the next lemma, we use Lemma 6.2 to construct a homogeneous family of bags of balls for a given 4-similar subsets of edges, \( \cup_{t ∈ T} O'(t) \) where \( T ⊆ T \). Later, in Subsection 6.2 we will construct a sequence of homogeneous families of bags of balls.

**Lemma 6.6.** Given a hierarchical decomposition \( T \) of \( G \), a set \( T ⊆ T \), \( 0 ≤ ε < 1/3 \) and a set \( O'(t) ⊆ O(t) \) for each \( t ∈ T \) such that for each \( e ∈ O'(t), c_1 ≤ (UX χ_e)^2 < 4c_1, c_2 ≤ \|UX χ_e\|^2 < 4c_2 \), and

\[ c_1 \cdot \mathbb{P}_{e ∈ O(t)}[e ∈ O'(t)] ≥ α \cdot c_2. \]  

For any \( β > 1 \) and \( 0 ≤ ε < 1/2 \), if \( α^ε ≤ \frac{1}{400βC_1(ε)} \), then one of the following holds:

1. For \( α-c_1 ≤ δ, Δ ≤ c_1/α^2 \), there is a homogeneous family of β-compact bags of balls FBag(δ, Δ) = \{Bag(δ, Δ, b)\} such that

\[ \sum_{t ∈ T} \frac{1}{|O(t)|} \left( \sum_{e ∈ O'(t)} (UX χ_e)^2 \right) ≤ \frac{16C_1(ε)}{α^ε} \sum_{Bag(δ, b) ∈ FBag(δ)} δ \cdot b. \]

2. For \( c_1 ≤ ρ ≤ c_1/α^2 \) there is a set of nodes \( S ⊆ T \) and a homogeneous family of \( α^{1+2ε} \)-assigned bags of balls FBag(ρ, S), such that

\[ \sum_{t ∈ T} \frac{1}{|O(t)|} \left( \sum_{e ∈ O'(t)} (UX χ_e)^2 \right) ≤ \frac{400βC_1(ε)}{α^ε} \sum_{Bag(ρ, b) ∈ FBag(ρ, S)} ρ \cdot b. \]
Applying \( L \) to \( O \) and \( F \), we use the notation \( N := \left| \bigcup_{t \in T} O(t) \right| \) and \( N' := \left| \bigcup_{t \in T} O'(t) \right| \). Since each edge is in at most two sets \( O(t), O'(t) \), \( 2N \geq 2 \sum_{t \in T} |O(t)| \) and \( 2N' \geq \sum_{t \in T} |O'(t)| \). Note that \( N \geq N' \) by definition.

First, observe that,

\[
\frac{c_1}{c_2} \geq \alpha \cdot \max_{t \in T} \frac{1}{\mathbb{P}_{e \sim O(t)} [e \in O'(t)]} = \alpha \cdot \max_{t \in T} \frac{|O(t)|}{|O'(t)|} \geq \alpha \cdot \frac{\sum t \in T |O(t)|}{\sum_{t \in T} |O'(t)|} \geq \frac{\alpha N}{2N'},
\]

Let \( F := \bigcup_{t \in T} O'(t) \). Observe that \( N' = |F| \). Then,

\[
\left( \mathbb{E}_{e \sim F} (UX \chi e_{\pi(e)}) \right)^2 \geq c_1 = \left( \frac{c_1}{c_2} \right) \cdot c_2 \geq \left( \frac{\alpha N}{8N'} \right) \cdot \mathbb{E}_{e \sim F} \|X \chi e\|_2^2.
\]

Let \( \tilde{\alpha} = \frac{\alpha N}{8N'} \). Let \( Y_v := UX_v \) for all \( v \in V \). Since \( U \) is unitary, for each pair \( u, v \)

\[
\|Y_u - Y_v\|_2^2 = \|X_u - X_v\|_2^2 = \|X_u - X_v\|_1.
\]

Applying Lemma 6.2 to \( Y \) and \( F \), we obtain a family \( A \) of

\[
b \geq \frac{\tilde{\alpha} N'}{C_1(\epsilon)} \geq \frac{\alpha N}{8C_1(\epsilon)},
\]

disjoint \( L_2^2 \) balls with radius \( \delta \) such that

\[
\delta \cdot b \geq \frac{\tilde{\alpha} \cdot N' \cdot c_1}{C_1(\epsilon)} \geq \frac{c_1 \alpha^\epsilon N^\epsilon N'^{1-\epsilon}}{2C_1(\epsilon)}.
\]

Note that these balls also correspond to \( L_2^2 \) balls in the \( X \) embedding. Since \( X \) is a cut metric, we can treat them as \( L_1 \) balls of the same radius \( \delta \) in the \( X \) embedding while preserving the disjointness property. Without loss of generality, (perhaps by decreasing \( \delta \)) we assume \( \delta \cdot b = \frac{c_1 \alpha^\epsilon N^\epsilon N'^{1-\epsilon}}{2C_1(\epsilon)} \). Since \( \frac{\alpha N}{8C_1(\epsilon)} \leq b \leq 2N' \),

\[
\alpha \cdot c_1 \leq \frac{\alpha N}{8C_1(\epsilon)} \leq \delta \leq \frac{4c_1}{\alpha^{1-\epsilon}} \leq \frac{c_1}{\alpha}.
\]

It follows from (22) that

\[
\sum_{t \in T} \frac{1}{|O(t)|} \left( \sum_{e \in O(t)} (UX \chi)_{\pi(e)} \right)^2 \leq \frac{c_1 \alpha^\epsilon N^\epsilon N'^{1-\epsilon}}{2C_1(\epsilon)} \leq \frac{16C_1(\epsilon) N'}{\alpha^\epsilon N^\epsilon N'^{1-\epsilon}} \leq \frac{16C_1(\epsilon) N'}{\alpha^\epsilon}.
\]

In the above we used \( |O'(t)| \leq |O(t)| \) for all \( t \) and in particular \( N' \leq N \).

Let \( \rho = \delta \vee 4c_2 \). Let

\[
V'(t) = \{ u \in V(t) : \{ u, v \} \in O'(t) \}.
\]

be the endpoints of edges of \( O'(t) \) that are in \( V(t) \). Also, let \( V' = \bigcup_{t \in T} V'(t) \). Let \( B \) be a maximal family of disjoint \( L_1 \) balls of radius \( \rho \) on the points in \( V' \). To construct \( B \), we scan the points in \( V' \) in an arbitrary order; for each point \( X_u \) if the ball \( B(X_u, \rho) \) does not touch any of the balls already added to \( B \) we add \( B \) to \( B \).

Depending on the size of \( B \) we conclude with (1) or (2).
Case 1. $|B| < \frac{b \delta}{12 \beta \rho}$. For each ball $B = B(X_u, \rho) \in \mathcal{B}$ let

\[ f(B) := \{ B(X_v, \delta) \in A : \|X_u - X_v\|_1 = \min_{B'(X_{u'}, \rho) \in B} \|X_{u'} - X_v\|_1 \} , \]

be the balls of $A$ that are closer to $B$ than any other ball of $\mathcal{B}$. We break ties arbitrarily, making sure that $f(B) \cap f(B') = \emptyset$ for any two distinct balls of $\mathcal{B}$.

First, we show that for any $B(X_u, \rho) \in \mathcal{B}$, the diameter of the centers of balls of $f(B)$ is at most $6\rho$. Fix a ball $B = B(X_v, \delta) \in \mathcal{B}$. For any ball $B_1(X_{v_1}, \delta) \in f(B)$ we show that $\|X_u - X_{v_1}\|_1 \leq 3\rho$. Since for all $e \in F$, $\|X_{X_e}\|_1 \leq 4c_2$, there is a vertex $u_1 \in V'$ such that $\|X_{v_1} - X_{u_1}\| \leq 4c_2$. Furthermore, by construction of $\mathcal{B}$, there is a ball $B'(X_{u_2}, \rho) \in \mathcal{B}$ such that $\|X_{u_1} - X_{u_2}\|_1 \leq 2\rho$. Putting these together,

\[ \|X_{v_1} - X_u\|_1 \leq \|X_{v_1} - X_{u_2}\|_1 \leq \|X_{v_1} - X_{u_1}\|_1 + \|X_{u_1} - X_{u_2}\|_1 \leq 4c_2 + 2\rho \leq 3\rho. \]

So, diameter of centers of balls of $f(B)$ is at most $6\rho$.

Letting $\Delta = 6\rho$, for each $B \in \mathcal{B}$ if $|f(B)| \geq 6\beta \cdot \rho/\delta$, then the balls in $f(B)$ form a $\beta$-compact bag of balls, as

\[ \beta \cdot \Delta = 6\beta \cdot \rho \leq \delta \cdot |f(B)|. \]

In that case, we add $f(B)$ as a new bag of balls to $\text{FBag}(\delta, \Delta)$. Observe that all balls in $\text{FBag}(\delta, \Delta)$ are disjoint, so $\text{FBag}(\delta, \Delta)$ is a homogeneous family $\beta$-compact bag of balls. It remains to verify conclusion (1). First, by (23),

\[ \alpha \cdot c_1 \leq \delta \leq \Delta \leq 6\rho \leq \frac{24c_1}{\alpha} \leq \frac{c_1}{\alpha^2}. \]

On the other hand, it is easy to see that the number of balls in $\text{FBag}(\delta, \Delta)$ is at least $b/2$. This is because,

\[ \sum_{B \in \mathcal{B}} b \geq b - \sum_{B \in \mathcal{B}} \mathbb{I} \left[ |f(B)| < \frac{6\beta \cdot \rho}{\delta} \right] \cdot |f(B)| \geq b - |\mathcal{B}| \cdot \frac{6\beta \cdot \rho}{\delta} \geq b/2. \]

The last inequality uses the assumption $|\mathcal{B}| \leq \frac{b \delta}{12 \beta \rho}$. So, conclusion (1) follows from (24).

Case 2. $|\mathcal{B}| \geq \frac{b \delta}{12 \beta \rho}$. For any node $t \in T$ let $\text{Bag}_t$ be the set of balls in $\mathcal{B}$ such that their centers are in $V'(t)$. If the center of a ball $B$ in $\mathcal{B}$ belongs to multiple $V'(t)$'s we include $B$ in exactly one of those sets arbitrarily. Now, for each $t \in T$, if

\[ \frac{|\text{Bag}_t|}{|\mathcal{B}|} \geq \frac{|\mathcal{O}(t)|}{4N}, \]

then we assign $\text{Bag}_t$ to $t$ and we add $t$ to $S$. Next, we verify that this is a valid assignment. First, since for any ball $B(X_u, \rho) \in \text{Bag}_t$, $u \in V'(t)$, there is an edge $\{u, v\} \in \mathcal{O}'(t)$ such that $\|X_u - X_v\|_1 \leq 4c_2 \leq \rho$. If $\delta \geq 4c_2$, then

\[ |\text{Bag}_t| \geq \frac{|\mathcal{B}| \cdot |\mathcal{O}(t)|}{4N} \geq \frac{|\mathcal{O}(t)| \cdot b \cdot \delta}{48\beta \cdot \delta \cdot N} \geq \frac{\alpha \cdot |\mathcal{O}(t)|}{384\beta \cdot C_1(\epsilon)} \geq \alpha^{1+\epsilon} \cdot |\mathcal{O}(t)|, \]

34
where second to last inequality uses (21) and the last inequality uses \( \alpha^\varepsilon \leq \frac{1}{400^\beta C_1(\varepsilon)} \). Otherwise, if \( 4c_2 \geq \delta \),

\[
|\text{Bag}_t| \geq \frac{|B| \cdot |O(t)|}{4N} \geq \frac{b \cdot \delta \cdot |O(t)|}{48\beta \cdot \rho \cdot N} \geq \frac{\alpha^\varepsilon |O(t)|}{192 C_1(\varepsilon)\beta} \cdot \frac{c_1}{c_2} \cdot \left( \frac{N'}{N} \right)^{1-\varepsilon} \\
\geq \frac{\alpha^{1+\varepsilon} |O(t)|}{192\beta C_1(\varepsilon)} \cdot \left( \frac{N'}{N} \right)^{\varepsilon} \geq \alpha^{1+2\varepsilon} \cdot |O(t)|. \tag{26}
\]

The first inequality uses (25), the second inequality follows by (22), the third inequality uses (20), and the last inequality uses the assumption that \( \alpha^\varepsilon \geq \frac{1}{400^\beta C_1(\varepsilon)} \). Finally, it remains to upper-bound \( \sum_{t} \frac{1}{|O(t)|} (\sum_{e \in O(t)} (UX_Xe)_{\pi(e)})^2 \). First, we show that \( \sum_{t \in S} |\text{Bag}_t| \geq |B|/2 \). This is because by (25),

\[
\sum_{t \in T-S} |\text{Bag}_t| \leq \sum_{t \in T} \frac{|O(t)| \cdot |B|}{4N} \leq |B|/2.
\]

The conclusion (2) follows by (24) and the assumption that \( |B| \geq \frac{\beta \delta}{12^3 \rho} \).

### 6.2 Construction of a Geometric Sequence of Families of Bags of Balls

Next, we construct a geometric sequence of bags of balls, that is a sequence of families of bags of balls where the size of balls in bags decrease geometrically (as a polynomial function of \( k \)). In the following lemma we construct two geometric sequence of bags of balls, namely a geometric sequence of compact bags of balls and an geometric sequence of assigned bags of balls.

A \( \lambda \)-geometric sequence of homogeneous families of bags of balls is a sequence \( \text{FBag}_1(\delta_1), \text{FBag}_2(\delta_2), \) etc., such that for each \( i \geq 1 \),

\[
\delta_i \cdot \lambda > \delta_{i+1}.
\]

A \( \lambda \)-geometric sequence of homogeneous families of compact bags of balls is defined respectively as a sequence \( \text{FBag}_1(\delta_1, \Delta_1), \text{FBag}(\delta_2, \Delta_2), \) etc., such that for each \( i \geq 1 \),

\[
\delta_i \cdot \lambda_i > \Delta_{i+1}.
\]

In this part we prove the following lemma.

**Lemma 6.7.** Given a hierarchical decomposition \( \mathcal{T} \) of a \( G \) and a set \( T \subseteq \mathcal{T} \) such that for any \( t \in T \),

\[
\left( \mathbb{E}_{e \sim \mathcal{O}(t)} (UX_Xe)_{\pi(e)} \right)^2 \geq \alpha \cdot \mathbb{E}_{e \sim \mathcal{O}(t)} \|X_Xe\|_1. \tag{27}
\]

For \( \bar{\alpha} = \alpha/C_2(\alpha) \), any \( \beta > 1, 0 < \varepsilon < 1/2 \) and \( \lambda < 1 \) if \( \alpha < \lambda \), and \( \bar{\alpha}^\varepsilon \leq \frac{1}{400^\beta C_1(\varepsilon)} \), then one of the following holds:

1. There is a \( \lambda \)-geometric sequence of homogeneous families of \( \beta \)-compact bags of balls \( \text{FBag}_1(\delta_1, \Delta_1), \text{FBag}_2(\delta_2, \Delta_2), \ldots, \) such that

\[
\frac{\bar{\alpha}^\varepsilon}{\beta C_1(\varepsilon) C_2(\alpha) \log^2 \bar{\alpha} \cdot |\log \lambda|} \cdot \sum_{t \in T} |\mathcal{O}(t)| \left( \mathbb{E}_{e \sim \mathcal{O}(t)} (UX_Xe)_{\pi(e)} \right)^2 \leq \sum_{i \geq 1} \sum_{\text{Bag}(\delta_i, b) \in \text{FBag}_i(\delta_i, \Delta_i)} \delta_i \cdot b. \tag{28}
\]
2. There is a $\lambda$-geometric sequence of homogeneous families of $\hat{\alpha}^{1+2\epsilon}$-assigned bags of balls $\text{FBag}_1(p_1, S_1), \text{FBag}_2(p_2, S_2)$, etc., such that $S_1, S_2, \text{etc. are disjoint subsets of } T$ and

$$\frac{\hat{\alpha}^\epsilon}{\beta C_1(\epsilon) C_2(\alpha) \log^2 \hat{\alpha} \cdot \log \lambda} \cdot \sum_{t \in T} |\mathcal{O}(t)| (\mathbb{E}_{e \sim \mathcal{O}(t)}(UX\chi_e)_{\pi(e)})^2 \lesssim \sum_{i \geq 1} \sum_{t \in S_i} \rho_i \cdot b_t. \quad (29)$$

Before proving the above lemma we first prove a technical lemma.

**Lemma 6.8.** For a hierarchical decomposition $\mathcal{T}$ of $G$, and any node $t \in \mathcal{T}$, $\alpha < 1/200$, if

$$(\mathbb{E}_{e \sim \mathcal{O}(t)}(UX\chi_e)_{\pi(e)})^2 \geq \alpha \cdot \mathbb{E}_{e \sim \mathcal{O}(t)} \|X\chi_e\|^2_2,$$

then there is a 2-similar set $\mathcal{O}'(t) \subset \mathcal{O}(t)$ as defined in Definition 6.3, and $C_2(\alpha) = O(\log(\alpha)^4)$ such that,

$$\left( \sum_{e \in \mathcal{O}'(t)} (UX\chi_e)_{\pi(e)} \right)^2 \geq \frac{1}{C_2(\alpha)} \cdot \left( \sum_{e \in \mathcal{O}(t)} (UX\chi_e)_{\pi(e)} \right)^2 \quad (30)$$

$$\mathbb{P}_{e \sim \mathcal{O}(t)}[e \in \mathcal{O}'(t)] \cdot \min_{e \in \mathcal{O}'(t)} (UX\chi_e)_{\pi(f)}^2 \geq \frac{\alpha}{C_2(\alpha)} \cdot \min_{e \in \mathcal{O}'(t)} \|X\chi_e\|^2_2. \quad (31)$$

**Proof.** We fix $t$ throughout the proof and use $\mathcal{O}$ instead of $\mathcal{O}(t)$ for brevity. Throughout the proof all probabilities are measured under the uniform distribution on $\mathcal{O}(t)$. We also use $a_e := (UX\chi_e)_{\pi(e)}$ and $b_e := \|X\chi_e\|$. Let $\mu := \mathbb{E}_{e \sim \mathcal{O}}[a_e]$. Define,

$$\mathcal{O}_i := \{e \in \mathcal{O}(t) : c^i \leq a_e/\mu < c^{i+1}\},$$

where $c = \sqrt{2}$. We also use $\mathcal{O}_{\geq j} = \bigcup_{i=j}^{\infty} \mathcal{O}_i$. Also, for any $i$ let $a_i = \min_{e \in \mathcal{O}(t)} a_e$.

Next, we show that there exists $-4 \leq i < 2(2 + \log \alpha^{-1})$ such that $\mathbb{E}_{e \in \mathcal{O}} a_e \geq \mu / \log(1/\alpha)$. First, observe that,

$$\sum_{i = -\infty}^{-5} a_i \cdot \mathbb{P}[e \in \mathcal{O}_i] \leq \sum_{i = -\infty}^{-5} c^{-4} \cdot \mu \cdot \mathbb{P}[e \in \mathcal{O}_i] \leq \mu/4.$$

Let $q \asymp \log(1/\alpha)$ chosen such that $c^q = 4/\alpha$. Then,

$$\frac{4 \mu}{\alpha} \cdot \sum_{i = q}^{\infty} a_i \mathbb{P}[e \in \mathcal{O}_i] \leq \sum_{i = q}^{\infty} c^2 \cdot \mathbb{P}[e \in \mathcal{O}_i] \leq \mathbb{E}_{e \sim \mathcal{O}_{\geq q}}[b_e^2] \cdot \mathbb{P}[e \in \mathcal{O}_{\geq q}] \leq \frac{\mu^2}{\alpha}. \quad (32)$$

The second inequality uses $a_e \leq b_e$ and the lemma’s assumption. Putting above two equations together, we obtain

$$\max_{-4 \leq i \leq q} \mathbb{P}[e \in \mathcal{O}_i] a_i \geq \frac{1}{4 + q} \sum_{i = -4}^{q} \mathbb{P}[e \in \mathcal{O}_i] a_i \geq \frac{\mu}{c(4 + q)} \quad (33)$$

where in the last inequality we used that for any $e \in \mathcal{O}_i$, $a_i \geq a_e/c$. Let $i$ be the maximizer of the LHS of above equation. It remains to choose a subset of $\mathcal{O}_i$ such that $b_e^2/b_f^2 < 2$ for all $e, f$ in that subset.
For an integer \( j \geq 0 \), we define 
\[
\mathcal{O}_{i,j} := \{ e \in \mathcal{O}_i : c^j \leq b_e / a_i < c^{j+1} \}.
\]
Note that any set \( \mathcal{O}_{i,j} \) is 2-similar. Let \( \mathcal{O}_{i, \geq q} = \bigcup_{j=q}^{\infty} \mathcal{O}_{i,j} \) Similar to (32),
\[
c^{2q} \cdot \mathbb{P}[e \in \mathcal{O}_{i, \geq q}] a_i^2 \leq \mathbb{E}_{e \sim \mathcal{O}_{i, \geq q}}[a_e^2] \leq \mathbb{E}_{e \sim \mathcal{O}}[b_e^2] \leq \frac{\mu^2}{\alpha} \leq \frac{1}{\alpha} \cdot 2a_i^2 \cdot (4 + q)^2 \cdot \mathbb{P}[e \in \mathcal{O}_i]^2,
\]
where the last inequality uses (33). Using \( c^t = 4/\alpha \), we obtain
\[
\mathbb{P}[e \in \mathcal{O}_{i, \geq q}] \leq \frac{\alpha}{16} \cdot (4 + q)^2 \cdot \mathbb{P}[e \in \mathcal{O}_i]^2 \leq \frac{1}{2} \cdot \mathbb{P}[e \in \mathcal{O}_i]^2,
\]
for \( \alpha < 1/200 \). Now, let \( j = \arg \max_{0 \leq j < q} \mathbb{P}[e \in \mathcal{O}_{i,j}] \). Then,
\[
\mathbb{P}[e \in \mathcal{O}_{i,j}] \cdot a_i^2 \geq \frac{a_i^2}{q^2} \cdot (\mathbb{P}[e \in \mathcal{O}_i] - \mathbb{P}[e \in \mathcal{O}_{i, \geq q}])^2 \geq \frac{\mathbb{P}[e \in \mathcal{O}_i]^2 \cdot a_i^2}{4q^2} \geq \frac{\mu^2}{8q^2(4 + q)}.
\]
The last inequality uses (33). Letting \( C_2(\alpha) = 8q^2(4 + q)^2 \) and \( O'(t) = \mathcal{O}_{i,j} \) both of the lemma’s conclusions follow from the above inequality. First, \( \mathcal{O}_{i,j} \) is a 2-similar. By the above inequality,
\[
\left( \sum_{e \in \mathcal{O}_{i,j}} (UX\chi_e)_e(e) \right)^2 \geq |O(t)|^2 \cdot \mathbb{P}[e \in \mathcal{O}_{i,j}]^2 a_i^2 \geq \frac{|O(t)|^2 \cdot \mu^2}{C_2(\alpha)} = \frac{1}{C_2(\alpha)} \cdot \left( \sum_{e \in O(t)} (UX\chi_e)_e(e) \right)^2.
\]
This proves (30). Similarly,
\[
\mathbb{P}[e \in \mathcal{O}_{i,j}] \cdot a_i^2 \geq \frac{\mu^2}{\mathbb{E}_{e \sim \mathcal{O}(t)} \| X\chi_e \|_2} \geq \frac{\alpha \cdot \mathbb{E}_{e \sim \mathcal{O}(t)} \| X\chi_e \|_2}{C_2(\alpha) \cdot \mathbb{P}[e \in \mathcal{O}_{i,j}] \| \mathcal{O}_{i,j} \|} \geq \frac{\alpha \cdot \sum_{e \in \mathcal{O}(t)} \| X\chi_e \|_2}{C_2(\alpha) \cdot \| \mathcal{O}_{i,j} \|} \geq \frac{\alpha}{C_2(\alpha)} \cdot \min_{e \in \mathcal{O}_{i,j}} \| X\chi_e \|_2.
\]
The second inequality uses the lemma’s assumption. This proves (31).

**Proof of Lemma 6.7.** First, by Lemma 6.8 for each node \( t \in T \), there is a 2-similar set \( O'(t) \subseteq \mathcal{O}(t) \) that satisfies (30) and (31). Let \( \tilde{\alpha} = \alpha / C_2(\alpha) \). For each \( t \in T \), let
\[
a_t = \min_{e \in \mathcal{O}'(t)} (UX\chi_e)_e^2(e) \quad \text{and} \quad b_t = \min_{e \in \mathcal{O}'(t)} \| X\chi_e \|_2.
\]
Without loss of generality, assume that \( \min_{t \in T} a_t \geq 1 \). Let \( \tilde{\lambda} < 1 \) be a function of \( \lambda \) that we fix later. For any integer \( i \geq 0 \), let
\[
T_i := \{ t \in T : \forall e \in \mathcal{O}'(t), \tilde{\lambda}^{i+1/2} \leq a_t < \tilde{\lambda}^{i-1/2} \}
\]
Note that, by definition, for all \( i \), \( T_i \cap T_{i+1} = \emptyset \).

Next, we construct a homogeneous family of bags of balls for each \( T_i \). First, we partition the nodes of each \( T_i \) into sets \( T_{i,j_a,j_b} \); for all integers \( 0 \leq j_a \leq \log(1/\lambda) \) and \( 0 \leq j_b \leq \log(\tilde{\alpha}) \) let
\[
T_{i,j_a,j_b} := \{ t \in T_i : 2^{j_a} \leq \frac{a_t}{\tilde{\lambda}^{i+1/2}} < 2^{j_a+1}, 2^{j_b} \leq \frac{b_t}{a_t} < 2^{j_b+1} \}.
\]

37
Observe that for all $i, j_a, j_b$, $\cup_{t \in T_{i, j_a, j_b}} \mathcal{O}'(t)$ is 4-similar. It follows that the number of groups in each $T_i$ is at most $O(\log^2(1/\lambda))$.

For a set $S \subseteq T$, let
\[
\Pi(S) := \sum_{t \in S} \frac{1}{|\mathcal{O}(t)|} \left( \sum_{e \in \mathcal{O}'(t)} (UX_Xe)_{\pi(e)} \right)^2.
\]

For each $T_i$ let
\[
T_{i, j_a(i), j_b(i)} = \arg\max_{T_{i, j_a, j_b}} \Pi(T_{i, j_a, j_b}).
\]

Since by lemma’s assumption,
\[
\tilde{\alpha}^\epsilon = \frac{\alpha^\epsilon}{C_2(\alpha)^\epsilon} \leq \frac{1}{400\beta \cdot C_1(\epsilon)},
\]
and $T_{i, j_a(i), j_b(i)}$ is 4-similar and all $t \in T_{i, j_a(i), j_b(i)}$ satisfy (31), we may invoke Lemma 6.6 for each set $T_{i, j_a(i), j_b(i)}$. This gives us either a homogeneous family of $\beta$-compact bags of balls $\text{FBag}_i(\delta_i, \Delta_i)$, or, for a set $S_i \subseteq T_{i, j_a(i), j_b(i)}$, a homogeneous family of $\alpha^{1+2\epsilon}$-assigned bags of balls, $\text{FBag}_i(\rho_i, S_i)$ such that
\[
\tilde{\alpha} \cdot \tilde{\lambda}^{i+1/2} \leq \tilde{\alpha} \cdot \min_{t \in T_i} a_t \leq \delta_i, \Delta_i, \rho_i \leq \frac{1}{\alpha^2} \max_{t \in T_i} a_t \leq \frac{\tilde{\lambda}^{-1/2}}{\tilde{\alpha}^2}.
\]

Without loss of generality we assume that $\Pi(\cup T_{2i}) \geq \Pi(\cup T_{2i+1})$. Consider the sequence $\text{FBag}_0, \text{FBag}_2, \text{FBag}_4, \ldots$. With probability $1/2$ we construct a geometric sequence of families of compact bags of balls by eliminating all families of assigned bags of balls in this sequence and with the remaining probability we construct a geometric sequence of families of assigned bags of balls by removing families of compact bags of balls. By (34), the decaying parameter of any of these two sequence is at least $\tilde{\lambda}/\tilde{\alpha}^3$. So, for $\tilde{\lambda} = \tilde{\alpha}^3 \cdot \lambda$ we obtain $\lambda$-geometric sequences. Furthermore, by Lemma 6.6, the sum of the radii of all balls in these sequences is at least
\[
\frac{\tilde{\alpha}^\epsilon}{400\beta C_1(\epsilon)} \sum_i \Pi(T_{2i, j_a(i), j_b(i)}) \geq \frac{\tilde{\alpha}^\epsilon}{400\beta C_1(\epsilon) \log \tilde{\lambda} \cdot \log \tilde{\alpha}} \sum_i \Pi(T_{2i}) \geq \frac{\tilde{\alpha}^\epsilon \cdot \Pi(T)}{2400\beta \cdot C_1(\epsilon) \log \lambda \log^{\epsilon} \tilde{\alpha}}.
\]

Now, by (30) of Lemma 6.8,
\[
\Pi(T) \geq \frac{1}{C_2(\alpha)} \cdot \sum_{t \in T} \frac{1}{|\mathcal{O}(t)|} \cdot \left( \sum_{e \in \mathcal{O}(t)} (UX_Xe)_{\pi(e)} \right)^2
\]

Therefore, one of the two geometric sequences satisfy conclusion 1 or 2 of the lemma. \qed

7 Charging Arguments

In this part we upper-bound the sum of radii of balls in a geometric sequence. Throughout this section we use $C_3, C_4 > 0$ as large universal constants. The following two propositions are the main statements that we prove in this section.

Proposition 7.1. Given a $k$-connected graph $G$, and a $\lambda$-geometric sequence of homogeneous families of $C_3$-compact bags of balls $\text{FBag}_1(\delta_1, \Delta_1), \text{FBag}_2(\delta_2, \Delta_2), \text{FBag}_3(\delta_3, \Delta_3), \text{FBag}_4(\delta_4, \Delta_4)$, etc. If $\lambda \leq 1/12$ and $C_5 \geq 36$, then,
\[
\frac{k}{4} \cdot \sum_{i \geq 1} \sum_{\text{Bag}(\delta_i, \Delta_i) \in \text{FBag}_i} \delta_i \cdot b \leq \sum_{\{u, v\} \in E} \|X_u - X_v\|_1.
\]
Proposition 7.2. Given a hierarchical decomposition $T(k, k\lambda, T)$ of $G$ and a $\lambda$-geometric sequence of homogeneous families of $24C_3/k$-assigned bags of balls, $\text{FBag}_1(\rho_1, T_1)$, $\text{FBag}_2(\rho_2, T_2)$, etc. such that $T_1, T_2$, etc. are disjoint subsets of $T$.

If $C_4 \geq 3$, $\lambda \leq 1/6C_4$ and $C_3 \geq (C_4 + 1) + 4(C_4 + 1)^2$, then

$$\frac{k}{4} \cdot \frac{C_4}{6C_3} \cdot \sum_{t \geq 1} \sum_{t \in T_i} \rho_i \cdot b_t \leq \sum_{\{u, v\} \in E} \|X_u - X_v\|_1.$$  

First, we use the above propositions to finish the proof of Theorem 6.1.

Proof of Theorem 6.1. Now, let $T' \subseteq T$ be the set of nodes satisfying (27). It follows that,

$$\alpha \geq \frac{\sum_{t \in T - T'} \frac{1}{|O(t)|} \cdot \left( \sum_{e \sim O(t)} (UX_{\chi_e})_\pi(e) \right)^2}{\sum_{T - T'} \sum_{e \sim O(t)} \|X_{\chi_e}\|_1} \geq \frac{\sum_{t \in T - T'} \frac{1}{|O(t)|} \cdot \left( \sum_{e \in O(t)} (UX_{\chi_e})_\pi(e) \right)^2}{2 \sum_{\{u, v\} \in E} \|X_u - X_v\|_1}. \quad (35)$$

The second inequality uses the fact that each edge is in at most two sets $O(t), O(t')$ for $t, t' \in T - T'$.

We apply Lemma 6.7 to $T'$. Let $C_4 = 3$, $\beta = 36$ and $C_3 = 68$. We choose $\alpha \propto \log^4(k)/k, \epsilon \propto \log \log(k)/\log(k)$ such that the following conditions are satisfied

$$\hat{\alpha}^\epsilon = \left( \frac{\alpha}{C_2(\alpha)} \right)^\epsilon \leq \frac{1}{200 \beta \cdot C_1(\epsilon)},$$

$$\hat{\alpha}^{1+2\epsilon} = \left( \frac{\alpha}{C_2(\alpha)} \right)^{1+2\epsilon} \geq \frac{24C_3}{k}.$$

Recall that $C_1(\epsilon)$ is an inverse polynomial of $\epsilon$ and $C_2(\alpha)$ is a poly-logarithmic function of $\alpha$ so the above assignment is feasible. Also let $\hat{\lambda} < \lambda/k$ such that $\hat{\lambda} < 1/6C_4$.

Now, by Lemma 7.18 either there is a $\hat{\lambda}$-geometric sequence of $\beta$-compact bags of balls $\text{FBag}_1(\delta_1, \Delta_1), \ldots$, that satisfies (28), or there is a $\hat{\lambda}$-geometric sequence of $\hat{\alpha}^{1+2\epsilon}$-assigned bags of balls $\text{FBag}_1(\rho_1, S_1), \ldots$, that satisfies (29). Now, by Proposition 7.1 and Proposition 7.2 we get

$$\frac{\sum_{t \in T'} \frac{1}{|O(t)|} \cdot \left( \sum_{e \in O(t)} (UX_{\chi_e})_\pi(e) \right)^2}{\sum_{e \in E} \|X_{\chi_e}\|_1} \leq C_1(\epsilon)C_2(\alpha) \log^2 \hat{\alpha} \cdot \left| \log \hat{\lambda} \right| \frac{k \cdot \hat{\alpha}^\epsilon}{k \cdot \hat{\alpha}^\epsilon}.$$

Noting $\hat{\alpha}^\epsilon = 1/\text{polylog}(k)$, the theorem follows from the above equation together with (35). $\square$

Before getting into the proofs of above propositions, we give a simple example to show that above claims do not hold for any arbitrary $\lambda$-geometric sequence of homogeneous bags of balls. This example is designed based on the dual solution that we constructed in Theorem 1.7.

Example 7.3. Let $G$ be the graph illustrated in Figure 2, and let $X_0, X_1, \ldots, X_{2h}$ be an embedding of $G$ where $X_i = 1_{[i]}$. Now, for any $1 \leq j \leq h - 1$, let $\text{Bag}_j$ be the union of balls

$$B(X_{2^j}, 2^j), B(X_{2^{j+1}}, 2^j), \ldots, B(X_{2^{h-2}}, 2^j).$$

Note that the center of each of these balls is a vertex of $G$ and that for any $j$, all balls of $\text{Bag}_j$ have equal radius and are disjoint. So we get a $1/2$-geometric sequence of bags of balls (and similarly we obtain a $\lambda$-geometric sequence by letting $j$ be multiples of $-\log(\lambda)$). As alluded to in the proof of Theorem 1.7, the the sum of the radii of balls in the give sequence is $h \cdot 2^h$ while the sum of the $L_1$ length of edges of $G$ is only $k \cdot 2^h$. 

39
1. We process bags of balls in phases, we assume that phase $\ell$ starts at time $\tau_{\ell-1} + 1$ and ends at $\tau_{\ell}$. In phase $\ell$ we process the bags in $\text{FBag}_\ell(\delta_\ell, \Delta_\ell)$; in other words, we add larger balls earlier than smaller ones.

2. In addition to adding new balls, in each phase we may shrink or delete some of the already inserted balls but when we insert a ball of $\text{FBag}_\ell$ we never alter it until the end of phase $\ell$.

3. We keep the invariant that for any $\tau$, all balls in $Z_\tau$ are disjoint. This crucial property will not hold in our construction of the assigned bags of balls in the next section and it is the main reason that our second construction is more technical.

4. For any ball $B(x, r_1, r_2) \in Z_\tau$, there are vertices $u, v \in V$ such that $\|x - X_u\|_1 \leq r_1$ and $\|x - X_v\|_1 \geq r_2$.

---

Figure 5: Properties of the inductive charging argument for compact bags of balls.

The above example serve as crucial barrier to both of our proofs. In the proof of Proposition 7.1 we bypass this barrier using the compactness of bags of balls. Note that in the above example Bag$_j$ is not dense, and indeed the diameter of centers of balls of Bag$_j$ is $2^h$ which is the same as the sum of the radii of balls in Bag$_j$. In the proof of Proposition 7.2 we bypass the above barrier using the properties of hierarchical decomposition $T$.

### 7.1 Charging Argument for Compact Bags of Balls

In this section we prove Proposition 7.1. We construct a set of $L_1$ balls inductively from the given compact bags of balls. For integer $\tau \geq 0$ we use $Z_\tau$ to denote the set of balls in the construction at time $\tau$. Initially, we have $Z_0 = \emptyset$ and $Z_\infty$ is the final construction. We describe the main properties of our construction in Figure 5.

**Inductive Charging.** Before explaining our construction, we describe our inductive charging argument. First, by the following lemma, in our construction, we only need to lower-bound the sum of the radii of all balls of $Z_\infty$ by (a constant factor of) the sum of radii of all balls in the given sequence of compact bags of balls.

**Lemma 7.4.** For any $\tau \geq 0$,

$$
k \cdot \sum_{B(x, r_1, r_2) \in Z_\tau} (r_2 - r_1) \leq \sum_{\{u, v\} \in E} \|X_u - X_v\|_1.
$$

**Proof.** We simply use $k$-connectivity of $G$. First, by property 4 of Figure 5 for each ball $B = B(x, r_1, r_2) \in Z_\tau$ there are vertices $u, v \in V$ such that $\|x - X_u\|_1 \leq r_1$ and $\|x - X_v\|_1 \geq r_2$. Since $G$ is $k$-edge connected, there are at least $k$ edge-disjoint paths between $u, v$. Each of these paths must cross $B$ and, by the triangle inequality, the length of the intersection with $B$ is at least $r_2 - r_1$. Finally, since by property 3 of Figure 5, balls of $Z_\tau$ are disjoint, this argument does not over-count the $L_1$-length of any edge of $G$. 

40
Say at the end of our construction, for any ball \( B(x, r_1, r_2) \in \mathcal{Z}_\infty \) we allocate \( r_2 - r_1 \) tokens. Our goal is to distribute these tokens between all bags of balls such that each \( \text{Bag}(\delta_i, \Delta_i, b) \) receives at least \( b \cdot \delta_i / 4 \) tokens. We prove this by an inductive argument. For a ball \( B(x, r_1, r_2) \in \mathcal{Z}_\tau \),

\[
\text{token}_\tau(B) := \begin{cases} 
\delta_i - 6\Delta_{\ell+1} & \text{if } B \in \text{FBag}_\ell \\
[(r_2 - r_1) - 6\Delta_\ell]^+ & \text{otherwise}
\end{cases}
\]  

(36)

Instead of allocating \( r_2 - r_1 \) tokens to a ball at time \( \tau \), we allocate \( r_2 - r_1 - 6\Delta_\ell \). The term \( 6\Delta_\ell \) takes into account the fact that we shrink balls in \( \mathcal{Z}_\tau \) later in our construction. We prove the following lemma inductively.

**Lemma 7.5.** At any time \( \tau_{\ell-1} + 1 \leq \tau \leq \tau_\ell \), if we allocate \( \text{token}_\tau(B) \) tokens to any ball \( B \in \mathcal{Z}_\tau \), then, we can distribute these tokens among the bags of balls that we processed by time \( \tau \) such that each \( \text{Bag}(\delta, \Delta, b) \) receives at least \( \delta \cdot b / 4 \) tokens.

It is easy to see that Proposition 7.1 follows by applying the above lemma to the final set of balls \( \mathcal{Z}_\infty \) and using Lemma 7.4,

\[
\frac{1}{\ell} \sum_{i \geq 1} \sum_{\text{Bag}(\delta_i, \Delta_i, b) \in \text{FBag}_i} \delta_i \cdot b \leq \sum_{B(x, r_1, r_2) \in \mathcal{Z}_\tau} r_2 - r_1 \leq \frac{1}{\ell} \sum_{\{u, v\} \in E} \|X_u - X_v\|_1.
\]

Construction. It remains to prove Lemma 7.5. First, we need some definitions. We say a ball \( B = B(X_u, \delta_\ell) \in \text{Bag}(\delta_\ell, \Delta_\ell, b) \) is in the interior of a ball \( B' = B(x, r_1, r_2) \in \mathcal{Z}_\tau \) if

\[
r_1 + \delta_\ell + \Delta_\ell \leq \|X_u - x\|_1 \leq r_2 - \delta_\ell - \Delta_\ell.
\]

In this case we call \( B \) an interior ball. Since balls in \( \mathcal{Z}_\tau \) are disjoint, \( B \) can be in the interior of at most one ball of \( \mathcal{Z}_\tau \).

**Fact 7.6.** Any ball \( B \in \text{Bag}(\delta_\ell, \Delta_\ell, b) \) is in the interior of at most one ball of \( \mathcal{Z}_\tau \).

Say Lemma 7.5 holds at time \( \tau_{\ell-1} < \tau \), we show it also holds at time \( \tau + 1 \). At time \( \tau \), we process a bag of balls in \( \text{FBag}_\ell \) that has at least one interior ball (and is not processed yet). If there is no such bag then we run the post processing algorithm that we describe later. Say at time \( \tau \) we are processing \( \text{Bag}^*(\delta_\ell, \Delta_\ell, b) = B_1(X_{u_1}, \delta_\ell), \ldots, B_6(X_{u_b}, \delta_\ell) \) of \( \text{FBag}_\ell \) and assume that one of these balls, say \( B_1 \), is in the interior of a ball \( B(x, r_1, r_2) \in \mathcal{Z}_\tau \).

First, we show that all balls of \( \text{Bag}^* \) are inside of \( B_1 \). Recall that, since \( \text{Bag}^* \) is \( C_3 \)-compact,

\[
b \cdot \delta \geq C_3 \cdot \Delta_\ell.
\]  

(37)

Let

\[
r_1' = \min_{1 \leq i \leq b} \|x - X_{u_i}\|_1 \quad \text{and} \quad r_2' = \max_{1 \leq i \leq b} \|x - X_{u_i}\|_1
\]

It follows that

\[
r_2' \leq \|x - X_{u_1}\|_1 + \text{diam}\{X_{u_1}, \ldots, X_{u_b}\} \leq (r_2 - \delta_\ell - \Delta_\ell) + \Delta_\ell \leq r_2 - \delta_\ell,
\]

and similarly, \( r_1' \geq r_1 + \delta_\ell \). Therefore, all balls of \( \text{Bag}^* \) are inside of \( B_1 \) and by property 3 of Figure 5 they do not touch any other ball of \( \mathcal{Z}_\tau \).
Now, we construct $Z_{\tau + 1}$. We remove $B$ and add two new balls $B'_1 = B(x, r_1, r'_1 - \delta\ell)$ and $B'_2 = B(x, r_2, r_2')$. In addition, we add all of the balls of Bag*. It is easy to see that balls in $Z_{\tau + 1}$ are disjoint. We send $\delta\ell/4$ tokens of each of $B_1, \ldots, B_b$ to Bag*. We send the rest of their tokens and all of the tokens of $B'_1, B'_2$ to $B$ and we re-distribute them by the induction hypothesis. It follows that Bag* receives at least $b \cdot \delta\ell/4$ tokens and $B$ receives $\text{token}_{\ell}(B)$.

$$\text{token}_{\ell}(B_1') + \text{token}_{\ell}(B_2') + \sum_{i=1}^{b} \text{token}_{\ell}(B_i) \geq r'_2 - r'_1 - 2\delta\ell - 12\Delta\ell + b \cdot (\delta\ell - 6\Delta\ell+1) \geq \text{token}_{\ell}(B) + b \cdot \delta\ell(1 - 6\lambda) - 6\Delta\ell \geq \text{token}_{\ell}(B) + b \cdot \delta\ell/2 - C_3 \Delta\ell/4 \geq \text{token}_{\ell}(B) + b \cdot \delta\ell/4,$$

where the first inequality uses (36), the second inequality uses $\Delta\ell+1 \leq \lambda \cdot \delta\ell$, the third inequality uses that $\lambda < 1/12$ and $C_3 \geq 24$. The last inequality is the only place that we use the compactness of Bag*; it simply holds by (37). Therefore, Lemma 7.5 holds at time $\tau + 1$.

**Post Processing.** Let $\tau_{\ell}$ be the time that we have processed all bags with at least one interior ball of $\text{FBag}_{\ell}$ and let $\text{FBag}'_{\ell}$ be the set of bags that we have not processed yet, i.e., all balls of $\text{FBag}'_{\ell}$ are border balls with respect to $Z_{\tau_{\ell}}$. As alluded to, at the end of phase $\ell$, i.e., at time $\tau_{\ell}$, we shrink all balls of $Z_{\tau}$ except those balls that were in $\text{FBag}_{\ell}$. Given a ball $B = B(x, r_1, r_2) \in Z_{\tau_{\ell}}$, the shrink$_\ell$ operator is defined as follows:

$$\text{shrink}_{\ell}(B) := \begin{cases} 
B & \text{if } B \in \text{FBag}_{\ell} 
B(x, 0) & \text{otherwise.}
\end{cases}$$

At time $\tau_{\ell}$, for any ball $B \in Z_{\tau_{\ell}}$ we add shrink$_\ell(B)$ to $Z_{\tau_{\ell}+1}$. In addition, we add all of all bags of $\text{FBag}'_{\ell}$ to $Z_{\tau_{\ell}+1}$. This is the end of phase $\ell$ and we consider $Z_{\tau_{\ell}+1}$ as our construction in the beginning of phase $\ell + 1$.

Let us verify that balls of $Z_{\tau_{\ell}+1}$ are disjoint, i.e., $Z_{\tau_{\ell}+1}$ satisfies property 3 of Figure 5. For any two balls $B = B(x, r_1, r_2) \in Z_{\tau_{\ell}}$ and $B' = B(x', \delta\ell) \in \text{FBag}'_{\ell}$, we show that shrink$_\ell(B)$ and $B$ are non-intersecting. First, if $B \in \text{FBag}_{\ell}$, then shrink$_\ell(B) = B$, so by the definition of the homogeneous bag of balls, any two balls of $\text{FBag}_{\ell}$ are non-intersecting, so shrink$_\ell(B), B'$ are non-intersecting. Now, suppose $B \notin \text{FBag}_{\ell}$. Since $B' \in \text{FBag}'_{\ell}$, $B'$ is not in the interior of $B$, i.e., either $\|x - X_{u'}\|_1 < r_1 + \delta\ell + \Delta\ell$ or $\|x - X_{u'}\|_1 > r_2 - \delta\ell - \Delta\ell$. In both cases, $B'$ does not intersect shrink$_\ell(B)$.

It remains to distribute the tokens. We send all tokens of all balls of bags of $\text{FBag}'_{\ell}$ their corresponding bag. Therefore, any $\text{Bag}(\delta\ell, \Delta\ell, b) \in \text{FBag}'_{\ell}$, receives at least

$$b \cdot (\delta\ell - \text{loss}(\ell + 1)) \geq b \cdot \delta\ell(1 - 6\lambda) \geq b \cdot \delta\ell/2$$

tokens. In addition, for every ball $B \in Z_{\tau_{\ell}}$, we send all tokens of shrink$_\ell(B)$ to $B$. Since

$$\text{token}_{\ell}(B) = \text{token}_{\ell+1}(\text{shrink}_{\ell}(B)),$$

$B$ receives exactly the same number of tokens. This completes the proof of Proposition 7.1.
7.2 Charging Argument for Assigned Bags of Balls

In this part we prove Proposition 7.2. Before getting into the details of the proof we illustrate our ideas to bypass the barrier of Example 7.3. The first observation is that, unlike the previous section, we can not construct a family of disjoint balls in $Z_{\infty}$, where the sum of radii of balls of $Z_{\infty}$ is a constant fraction of the sum of radii of all balls in the given geometric sequence. Instead, we let balls of $Z_{\infty}$ intersect and we employ a ball labeling technique that uses the hierarchical decomposition, $T$.

Let us give a simple example to show the crux of our analysis. Suppose a node $t_1 \in T$ has exactly two children, $t_2, t_3$. Say at time $t_{i-1} < t \leq t_i$ we are processing Bag$_{t_2}(\rho_t, b_t)$. Suppose $Z_t$ has a large ball, $B = B(x, r) \in$ Bag$_t$ as shown in the left of Figure 6 such that $t$ is an ancestor of $t_1$. Say Bag$_{t_2}$ has four balls $B_1, \ldots, B_4$. Because Bag$_{t_2}$ is not compact, if we remove the part of $B$ that intersects with balls of Bag$_{t_2}$ and add $B_1, \ldots, B_4$, the sum of the radii of balls in $Z_{t+1}$ is the same as that sum in $Z_t$. Instead, we add a new ball that intersects $B_1, \ldots, B_4$ as shown in the right of Figure 6.

Say the center of each $B_i$ is $X_{u_i}$ for $u_i \in V(t_2)$; each $X_{u_i}$ corresponds to a blue dot in Figure 6. By the definition of assigned bag of balls, for each $i$ there is a vertex $v_i \in V(t_1) - V(t_2) = V(t_3)$ such that $\|X_{u_i} - X_{v_i}\| \leq \rho_t$ (each $X_{v_i}$ corresponds to a red dot in Figure 6). We add all balls of Bag$_{t_2}$ and a new ball from the closest red vertex to the center of $B$ to the farthest one. We also break $B$ into two balls and remove part of it that intersects with either of these 5 new balls.

Observe that, the sum of radii of balls of $Z_{t+1}$ is $\Omega(\rho_t \cdot b_{t_2})$ more than this sum in $Z_t$. The only problem is that, the balls of $Z_{t+1}$ are intersecting. So, it is not clear if, analogous to Lemma 7.4, we can charge the sum of radii of balls of $Z_{t+1}$ with the sum of $L_1$ length of edges of $G$. Our idea is to label balls with different subset of edges of $G$. Although the red ball and the blue balls intersect, we charge their radii with disjoint subset of edges of $G$; we charge the radius of the red ball with $k$ edge-disjoint paths supported on $E(V(t_1)) - E(V(t_2))$ going across this ball and we charge the radius of each blue ball with $k$ edge-disjoint paths supported on $E(V(t_2))$ going across that ball.

7.2.1 Ball Labeling

In this part we define a valid labeling of balls in our construction (see Figure 7). In the previous construction, we used the disjointness property of balls in the construction in two places; namely in the proofs of Lemma 7.4 and Fact 7.6. We address both of these issues by our ball labeling technique.

At the end of this part, we reduce Proposition 7.2 to a “simpler” statement, that is arranging a set of balls with a valid labeling such that the sum of the radii of all balls in the construction is a constant fraction of the sum of the radii of all balls in the given geometric sequence (see Proposition 7.9).

Basic Label. In the proof of Lemma 7.4 we used the disjointness property to charge the sum of the radii of balls of a set $Z_t$ by the sum of the $L_1$ length of edges of $G$ with no over-counting. Let us give a simple example to show the difficulty in extending this argument to the new setting where balls may intersect. Suppose $Z_t$ is a union of 10 identical copies of $B(x, r)$ with the guarantee that there is a vertex of $G$ at $x$ and one at distance $r$ of $x$. Then, the sum of the $L_1$ length of edges of $G$ can be as small as $k \cdot r$, as $G$ may just be $k$ edge-disjoint paths from a vertex at $x$ to a vertex at distance $r$ of $x$. 

43
Figure 6: A simple example of the ball labeling technique. Balls are illustrated as squares to emphasize that these are $L_1$ balls. The grey ball $B$ in the left is one the balls of $Z_\tau$. Small $L_1$ balls with blue vertices as their centers represent balls of $\text{Bag}_{t_1}$ that we are processing at time $\tau$. Each red vertex together with the closest blue vertex are the endpoints of an edge of $O(t_1)$. The right figure shows new balls added to $Z_{\tau+1}$. In particular, each blue vertex is in $V(t_2)$ and each red vertex is in $V(t_3)$ where $t_2, t_3$ are the only children of $t_1$.

A ball $B = B(x, r_1, r_2)$, can be labeled with $t \in T$, denoted by $t(B) = t$, if there are vertices $u, v \in V(t)$ such that $\|x - X_u\|_1 \leq r_1$ and $\|x - X_v\|_1 \geq r_2$. Recall that, by definition of $T$, for any node $t \in T$, $G[V(t)]$ is $k$ edge connected. Therefore, if $B$ is labeled with $t$, then $k$-edge disjoint paths supported on $E(t)$ cross $B$. For any ball $B \in \text{Bag}_t$ we let $t(B) = t$. Furthermore, when we shrink or divide a ball into smaller ones the label of the shrunk ball or the new subdivisions remain unchanged.

Now, we address the issue of intersecting balls. In a valid ball labeling, we make sure that, for any two intersecting balls $B, B'$, the two families of $k$-edge disjoint paths that cross $B, B'$ are supported on disjoint subsets of edges. Say $t, t' \in T$ are ancestor-descendent if either $t$ is an ancestor of $t'$ or $t'$ is an ancestor of $t$. We also assume that $t, t$ are ancestor-descendent. The simplest definition of a valid ball labeling is to make sure that for any two intersecting balls $B, B'$, $t(B), t(B')$ are not ancestor-descendent. Unfortunately, this simple definition is not enough for our inductive argument, and as we elaborate next, we may richen the label of some of the balls $B$ by “dis-allowing some of descendants of $t(B)$”.

**Child-Avoiding Balls** As alluded to in Figure 6, we may add new balls to $Z_\tau$ that do not exist in the given geometric sequence. We may treat these balls differently and call them child-avoiding balls. We insert a child-avoiding ball only when we shrink or remove part of a ball that already exists in $Z_\tau$. Say we add a child-avoiding $B'$ to $Z_{\tau+1}$ by removing part of a ball $B \in Z_\tau$. Then, $t(B') = t(B)$ and $B'$ has an additional label $t_d(B')$ where $t_d(B')$ is a descendent of $t(B')$. For example, say $B'$ is the the red ball in the right of Figure 6; then $t(B') = t, t_d(B') = t_2$.

For any child-avoiding ball $B = B(x, r_1, r_2)$ there must be vertices $u, v \in V(t(B)) - V(t_d(B))$ such that $\|X_u - x\|_1 \leq r_1$, $\|X_v - x\|_1 \geq r_2$ and that there are at least $k/2$ edge-disjoint paths from $u$ to $v$ in the induced graph $G[V(t(B)) - V(t_d(B))]$. 

44
Non-Insertable Balls In Fact 7.6 we used the disjointness property to argue that any ball of FBag$_t$ is in the interior of at most one ball of Z$_t$. Here, this fact may not necessarily hold. Say at time $\tau$ a ball $B \in \text{Bag}_t$ is in the “interior” of two balls $B_1, B_2$, i.e., the center of $B$ is far from the boundary and $t$ is ancestor-descendent of both of $t(B_1), t(B_2)$. Then, $B_1, B_2$ are intersecting. Since $B_1, B_2$ are intersecting, $t(B_1), t(B_2)$ are not ancestor-descendant. Now, if either of $t(B_1)$ or $t(B_2)$ is an ancestor of $t$ we reach to a contradiction. But a bad case happens when both of $t(B_1), t(B_2)$ are descendants of $t$.

In general, the above scenario occurs only if for a node $t$ we process the bags assigned to the descendants of $t$ earlier than Bag$_t$. To address this issue we use the third property of the hierarchical decomposition, $\mathcal{T}$. For any ball $B$ in our construction with $t(B) = t$, we define $P(B) \subset \mathcal{T}$ to be the set of descendants $t$ such that the $k$-edge disjoint paths across $B$ can be routed on $E(t) - \cup_{t' \in P(B)} E(t')$. We will prune Bag$_t$ to the set of balls $B$ such that $P(B)$ includes all descendants of $t$ that are processed earlier than $t$. We use the third property of the hierarchical decomposition, $\mathcal{T}$, to show that the pruning step only removes a small fraction of balls.

For a node $t \in T_\ell$, we say a node $t'$ is a preceder of $t$ if $t'$ is a descendent of $t$ and $t' \in T_i$ for $i < \ell$. For any node $t$ and any ball $B = B(X_u, r) \in \text{Bag}_t$ we say $B$ is non-insertable by $t'$ if $t'$ is a preceder of $t$ and an endpoint of an edge of $\mathcal{P}(t')$ is in $B$. We say $B$ is insertable otherwise. For any insertable ball $B \in \text{Bag}_t$ we let $P(B)$ be the set of preceders of $t$. In other words, a ball $B = (X_u, r) \in \text{Bag}_t$ is insertable if and only if for any preceder of $t$, $t'$, all endpoints of the edges of $\mathcal{P}(t')$ are outside of $B$. Observe that, by the definition of assigned bags of balls, if $B$ is insertable, then for any $t' \in P(B)$, $u \notin V(t')$. In particular, since $B \in \text{Bag}_t$, there is an edge $\{u, v\} \in \mathcal{O}(t)$ for $v \notin V(t)$. Therefore, if $u \in V(t')$, $\{u, v\} \in $ which is a contradiction.

In the following lemma we show that for any node $t$, the sum of radii of all balls that are non-insertable by $t$ is $o(\rho_t b_t)$.

**Lemma 7.7.** For any node $t$ such that Bag$_t(\rho_t, b_t) \in \text{FBag}_t$,

$$\sum_{i \geq 1} \sum_{B \in \text{FBag}_i} \mathbb{I}[B \text{ is non-insertable by } t] \cdot \rho_i \leq \frac{4\rho_t \cdot b_t}{C_3}. $$

**Proof.** For any $i \geq 1$ let $b_i$ be the number of non-insertable ball by $t$ in FBag$_t$. By definition $b_i = 0$ for $i > \ell$. We show that for all $i > \ell$,

$$b_i \leq 2 |\mathcal{P}(t)|. \tag{39}$$

Then,

$$\sum_{i \geq 1} \sum_{B \in \text{FBag}_i} \mathbb{I}[B \text{ is non-insertable by } t] \cdot \rho_i \leq \sum_{i > \ell} b_i \cdot \rho_i \leq 2 |\mathcal{P}(t)| \sum_{i > \ell} \rho_i \leq 4\lambda \cdot |\mathcal{P}(t)| \cdot \rho_t \leq \frac{4|\mathcal{O}(t)|}{k} \cdot \rho_t \leq \frac{4b_t \rho_t}{C_3},$$

where the second to last inequality uses $\mathcal{T}$ is a $\mathcal{T}(k, k\lambda, T)$ hierarchical decomposition of $G$, i.e., that $t \in T$ and $\lambda |\mathcal{P}(t)| \leq |\mathcal{O}(t)|$. The last inequality uses that Bag$_t$ is a $C_3/k$-assigned bag of balls.

It remains to prove (39). Fix $i > \ell$. For any ball $B = B(X_u, \rho_i) \in \text{Bag}_t$ that is non-insertable by $t$ at least one endpoint of an edge of $\mathcal{P}(t)$ is in $B$. Since all balls of FBag$_i$ are disjoint, $b_i \leq 2 |\mathcal{P}(t)|$. 

45
Any set of balls has a valid ball labeling if it satisfies the following properties.

1. Any non child-avoiding ball \( B = B(x, r_1, r_2) \) with labels \( t(B) = t, P(B) \) satisfies the following: There are vertices \( u \in V(t) - \cup_{t' \in P(B)} V(t') \) and \( v \in V(t) \) such that \( \|x - X_u\|_1 \leq r_1 \) and \( \|x - X_v\|_1 \geq r_2 \). Furthermore, for any \( t' \in P(B) \) and \( \{u', v'\} \in O(t') \), \( \|x - X_u\|_1, \|x - X_v\|_1 \geq r_2 \).

2. Any child-avoiding ball \( B = B(x, r_1, r_2) \) with labels \( t(B) = t, t_d(B) = t_d, P(B) \) satisfies the following: \( t_d \) is a descendent of \( t \), there are vertices \( u \in V(t) - V(t_d) - \cup_{t' \in P(B)} V(t') \) and \( v \in V(t) - V(t_d) \) such that \( \|x - X_u\|_1 \leq r_1 \) and \( \|x - X_v\|_1 \geq r_2 \) and there are at least \( k/2 \) edge-disjoint paths in the induced subgraph \( G[V(t) - V(t_d)] \) from \( u \) to \( v \). Furthermore, for any \( t' \in P(B) \) and \( \{u', v'\} \in O(t') \), \( \|x - X_u\|_1, \|x - X_v\|_1 \geq r_2 \).

3. For any two balls \( B_1, B_2 \) one of the following four conditions is satisfied: i) \( B_1 \) and \( B_2 \) do not intersect. ii) \( t(B_1) \) and \( t(B_2) \) are not ancestor-descendent, iii) \( P(t(B_1)) \) contains (an ancestor of) \( t(B_2) \), or \( B_1 \) is child-avoiding and \( t_d(B_1) \) is (an ancestor of) \( t(B_2) \), iv) Conversely, \( P(t(B_2)) \) contains (an ancestor of) \( t(B_1) \), or \( B_2 \) is child-avoiding and \( t_d(B_2) \) is (an ancestor of) \( t(B_1) \).

Figure 7: Properties of a valid ball labeling

By the above lemma it is sufficient to prove Proposition 7.2 with the assumption that all balls in the given geometric sequence are insertable (see Proposition 7.9 at the end of this part).

In Figure 7 we define a valid labeling of balls. In our inductive argument we make sure that at any time \( \tau \), \( Z_\tau \) has a valid labeling. Note that, by definition, any insertable ball satisfies property 1. Therefore, for any insertable \( B \in Bag_t \), \( B \) can be added to \( Z_\tau \), if for any \( B' \in Z_\tau \), \( \{B, B'\} \) has a valid labeling.

The following lemma extends Lemma 7.4 to the new setting where the balls of \( Z_\tau \) may intersect.

**Lemma 7.8.** For any set of balls \( Z \) with a valid labeling we have,

\[
\frac{k}{2} \sum_{B(x_1, r_1, r_2) \in Z} (r_2 - r_1) \leq \sum_{\{u, v\} \in E} \|X_u - X_v\|_1.
\]

**Proof.** For any ball \( B \in Z \) we choose at least \( k/2 \) edge disjoint paths, \( P_1, \ldots, P_{k/2} \), going across it, and we charge \( B \) with \( \sum_{i=1}^{k/2} \|P_i \cap B\|_1 \). We choose these paths such that for any two intersecting balls, the corresponding paths are supported on disjoint subset of edges of \( G \). More precisely, we charge \( B \) with \( k/2 \) paths that are supported on \( E(t(B)) - \cup_{t' \in P(B)} E(t) \), if \( B \) is not child-avoiding, and \( E(t(B)) - E(t_d(B)) - \cup_{t' \in P(B)} E(t) \), otherwise. Since \( Z \) has a valid labeling, our counting argument does not over-count any edge of \( G \) and we are done.

It remains to construct the paths. First, assume that \( B = B(x, r_1, r_2) \in Z \) is not a child-avoiding ball. Then, there are vertices \( u \in V(t(B)) - \cup_{t' \in P(B)} V(t) \) and \( v \in V(t(B)) \) such that \( \|x - X_u\| \leq r_1 \) and \( \|x - X_v\| \geq r_2 \) as promised by property 1 of Figure 7. Since \( G[V(t(B))] \) is \( k \) connected, there are \( k \) edge disjoint paths from \( u \) to \( v \). We cut each of these \( k \) paths once they
cross $B$ and reach to a vertex at distance $r_2$ of $x$. Since for each $t \in P(B)$, there is no endpoint of an edge of $P(t)$ in $B(x, r_2)$ and $u \notin V(t)$, none of the edges of the chosen paths are in $E(t)$ and we are done.

Finally, if $B$ is a child-avoiding ball, we use a similar argument. The difference is that $G[V(t(B)) - V(t_d(B))]$ is not necessarily $k$ connected, but by property 2 we are guaranteed that there are vertices $u \in V(t(B)) - V(t_d(B)) - \cup_{t \in P(B)} V(t)$ and $v \in V(t(B)) - V(t_d(B))$ such that there are at least $k/2$ edge disjoint paths in $G[V(t(B)) - V(t_d(B))]$ from $u$ to $v$. Similarly, we cut each of these paths once it reaches to a vertex at distance $r_2$ and we guarantee that the corresponding paths do not use any edge of $\cup_{t \in P(B)} E(t)$.

By the above lemma, to prove Proposition 7.2, it is enough to construct a set of balls with a valid labeling such that the sum of the radii of balls in our construction is proportional to sum of the radii of balls in the given geometric sequence of assigned bags of balls.

**Proposition 7.9.** Given a hierarchical decomposition $T(k, k\lambda, T)$ of $G$ and a $\lambda$-geometric sequence of homogeneous families of $12C_3/k$-assigned bags of balls, $\text{FBag}_1(\rho_1, T_1)$, $\text{FBag}_2(\rho_2, T_2)$, etc., such that each $T_i$ is a subset of $T$. If all balls of all bags in the sequence are insertable, $C_4 \geq 3$, $\lambda \leq 1/6C_4$ and $C_3 \geq (C_4 + 1) + 4(C_4 + 1)^2$, then there is a set $Z$ of balls with a valid labeling such that

$$\frac{C_4}{6C_3} \sum_{i \geq 1} \sum_{\text{Bag}_i(\rho_i, b_i) \in \text{FBag}_i(\rho_i, T_i)} \rho_i \cdot b_i \leq \sum_{B(x, r_1, r_2) \in Z} (r_2 - r_1).$$

It is easy to see that the above proposition together with Lemma 7.8 implies Proposition 7.2.

**Proof of Proposition 7.2.** For any $i \geq 1$ and any $\text{Bag}_i(\rho_i, b_i) \in \text{FBag}_i$, we remove all non-insertable balls in $\text{Bag}_i$. If at least half of the balls of $\text{Bag}_i$ are insertable then we will have a $12C_3/k$-assigned bag of balls. Otherwise, we remove $\text{Bag}_i$ from our geometric sequence and we remove $t$ from $T_i$. The resulting geometric sequence satisfies the conditions of Proposition 7.9.

By Lemma 7.7, the sum of the radii of balls that we removed, which is at most twice the sum of the radii of all non-insertable balls, is at most half of the radii of all balls in the given geometric sequence,

$$\sum_{i \geq 1} \sum_{B \in \text{FBag}_i} \mathbb{I}[B \text{ is non-insertable}] \cdot \rho_i = \sum_{i \geq 1} \sum_{t \in T_i} \sum_{j \geq 1} \sum_{B \in \text{FBag}_j} \mathbb{I}[B \text{ is non-insertable by } t] \cdot \rho_j \leq \sum_{i \geq 1} \sum_{t \in T_i} \frac{4b_t \cdot \rho_i}{C_3} \leq \sum_{i \geq 1} \sum_{t \in T_i} \frac{b_t \cdot \rho_i}{4},$$

where the last inequality uses $C_3 \geq 16$. Therefore, the proposition follows by Lemma 7.8.

**7.2.2 Order of Processing**

In the rest of this section we prove Proposition 7.9. So from now on, we assume all balls of all bags in the sequence are insertable and that every bag is $12C_3/k$-assigned bag of balls.

Similar to the previous section, we give an inductive proof. In this part we describe general properties of our construction and we use them to prove two essential lemmas. We process families of bags of balls in phases, and in phase $\ell$ we process $\text{FBag}_\ell$. See Figure 8 for the essential properties of our inductive construction. In the rest of this part we use these properties to prove lemmata 7.12 and 7.13.

The following fact follows simply by property (3).
1. Phase $\ell$ starts at $\tau_{\ell-1} + 1$ and ends at $\tau_{\ell}$. In phase $\ell$ we process assigned bags of balls in $F_{Bag_{\ell}}$ in an increasing order of the distance the node that they are assigned to from the root of $T$. For example, if $Bag_{t_1}, Bag_{t_2} \in F_{Bag_{\ell}}$ and $t_1$ is an ancestor of $t_2$, we first process $t_1$ and then $t_2$.

2. Any ball of $F_{Bag_{\ell}}$ that we insert (in phase $\ell$) remains unchanged by the end of phase $\ell$. All other balls may be shrunk or be splitted into several balls but their labels remain invariant.

3. Say at time $\tau_{\ell-1} < \tau < \tau_{\ell}$ we process $Bag_{t}$. We construct $Z_{\tau+1}$ inductively by constructing $Z_{\tau,0} = Z_{\tau}, Z_{\tau,1}, \ldots, Z_{\tau,\infty} = Z_{\tau+1}$. We make sure that each set $Z_{\tau,s}$ has a valid labeling. When we construct $Z_{\tau,s+1}$, we insert several new balls where only some of them are in $Bag_{t}$. Any new ball $B = B(x, r_1, r_2) \in Z_{\tau,s+1}$, that is not in $Bag_{t}$, must be inside a non-child-avoiding ball $B' = B(x, r'_1, r'_2) \in Z_{\tau,s}$, i.e., $r'_1 \leq r_1, r_2 \leq r'_2$, where $t(B')$ is an ancestor of $t$ and $P(B')$ does not have (an ancestor of) $t$. If $B$ is child-avoiding, then $t(B) = t(B')$, $t_d(B) = t$ and $P(B) = P(B')$; otherwise, $t(B) = t$ and $P(B) \subseteq P(B')$.

4. At time $\tau_{\ell}$ we process the border balls of all bags of $F_{Bag_{\ell}}$.

**Figure 8: Properties of our Inductive Construction**

**Lemma 7.10.** Say we process $Bag_{t} \in F_{Bag_{\ell}}$ at time $\tau$. For any $s \geq 0$ and any ball $B \in Bag_{t}$ and $B' \in Z_{\tau,s}$, if $\{B, B'\}$ does not have a valid labeling, then $t(B')$ is equal to $t$ or is an ancestor of $t$.

**Proof.** Let $t' = t(B')$. If $\{B, B'\}$ does not have a valid labeling then by property 3 of Figure 7, $t, t'$ are ancestor-descendent. So, we just need to show that $t'$ is not a descendent of $t$.

First, by properties 2, and 3 of Figure 8, $t(B') \in T_i$ for $i \leq \ell$. Furthermore, if $t' \in T_\ell$ either $t' = t$ or $Bag_{t'}$ is processed by time $\tau$.

Now, if $t' \in T_\ell$, then by property 1 of Figure 8, $t'$ is not a descendent of $t$ and we are done. Otherwise, $t' \in T_i$ and $i < \ell$. If $t'$ is a descendent of $t$, then it is a preceder of $t$ and since $B$ is an insertable ball, $t' \in P(B)$. So $\{B, B'\}$ has a valid labeling, which can not be the case. \qed

We need to use slightly bigger constants in the definition of interior balls.

**Definition 7.11.** We say a ball $B \in Bag_{t}$ is in the interior of a ball $B' = B(x, r_1, r_2) \in Z_{\tau}$ if $\{B, B'\}$ does not have a valid labeling and,

$$r_1 + C_3 \cdot \rho_\ell < ||x - X_u||_1 < r_2 - C_3 \cdot \rho_\ell.$$ 

We say $B$ is an interior ball if $B$ is in the interior of a ball. If $B$ is not an interior ball, we call it a border ball. Similar to the previous construction we insert all border balls of phase $\ell$ at time $\tau_{\ell}$.

In the following lemma we show that when we are processing $Bag_{t}$ (at time $\tau$) any ball in this bag that is in the interior of a ball of $Z_{\tau,s}$ has a valid labeling with any other ball of $Z_{\tau,s}$

**Lemma 7.12.** Say we process $Bag_{t} \in F_{Bag_{\ell}}$ at time $\tau$. For any $s \geq 0$, and any ball $B \in Bag_{t}$ and $B' \in Z_{\tau,s}$, if $\{B, B'\}$ does not have a valid labeling, then for any ball $B'' \in Z_{\tau,s}$ that intersects $B'$, $\{B, B''\}$ has a valid labeling.
Consequently, if \( B \) is in the interior of \( B' \in Z_{\tau,s} \), then \( \{B, B''\} \) has a valid labeling for any \( B'' \in Z_{\tau,s} \) where \( B'' \neq B' \).

Proof. Let \( t(B') = t' \); Fix a ball \( B'' \in Z \) and let \( t(B'') = t'' \). Assume, for the sake of contradiction, that \( \{B, B''\} \) does not have a valid labeling. \( t' \) is not a descendent of \( t \). Otherwise, First, by Lemma 7.10, \( t' \) is equal to \( t \) or is an ancestor of \( t \). Similarly, \( t'' \) is equal to \( t \) or is an ancestor of \( t \). Therefore, \( t', t'' \) are ancestor-descendent.

Since \( B', B'' \in Z_{\tau,s} \), \( \{B', B''\} \) has a valid labeling. If \( t' = t'' = t \), then we reach to a contradiction. So, without loss of generality assume that \( t' \) is an ancestor of \( t'' \). By property 3 of Figure 7, \( P(B') \) has (ancestor of) \( t'' \), or \( B' \) is child-avoiding and \( t_d(B') \) is (ancestor of) \( t'' \). But, this implies that \( \{B, B'\} \) has a valid labeling which is a contradiction.

In the following lemma we show that once a ball \( Bag_{\ell} \) becomes a border ball, it remains to be a border ball by the end of phase \( \ell \).

**Lemma 7.13.** Say we process \( Bag_{\ell} \in FBag_{\ell} \) at time \( \tau \). For any ball \( B \in Bag_{\ell} \) if \( B \) is an interior ball w.r.t. \( Z_{\tau,s} \) for \( \tau' \geq \tau \) and \( s > 0 \), then, it is also in the interior of a ball of \( Z_{\tau',s-1} \).

Proof. Say we process \( Bag_{\ell'} \) at time \( \tau' \), and \( B \) is in the interior of a ball \( B' \in Z_{\tau',s} \) Since \( B \) is in the interior of \( B' \), \( \{B, B'\} \) does not have a valid labeling, so \( t', t'' \) are ancestor-descendent and by property 1 of Figure 8, \( t \) is equal to \( t \) or is an ancestor of \( t' \).

Since balls of \( FBag_{\ell} \) are disjoint, \( B' \notin Bag_{\ell} \), so by property 3 of Figure 8, \( B' \) is inside a non child-avoiding ball \( B'' \in Z_{\tau,s-1} \) such that \( t(B'') \) is an ancestor of \( t' \) and \( P(B'') \) does not have (ancestor of) \( t' \). Let \( t' = t(B'') \), and note that \( t, t'' \) are ancestor-descendent. We show that \( B \) is in the interior of \( B'' \).

Since \( B' \) is inside \( B'' \), we only need to show that \( \{B, B''\} \) does not have a valid labeling. First, assume that \( B' \) is a child-avoiding ball. Then, we must have \( t(B') = t'' \). But then by property 3 of Figure 8, \( t_d(B') = t' \) and \( P(B') = P(B'') \). Since \( t \) is equal to \( t' \) or is an ancestor of \( t' \), if \( \{B, B''\} \) has a valid labeling, then \( \{B, B'\} \) has a valid labeling. Therefore, \( \{B, B''\} \) does not have a valid labeling and we are done.

Otherwise, \( B' \) is not a child-avoiding ball and by property 3 of Figure 8, \( t(B') = t' \). For the sake of contradiction, suppose \( \{B, B''\} \) has a valid labeling. Since \( t \) and \( t'' \) are ancestor-descendent, by property 3 of Figure 7 one of the following two cases holds: (i) \( P(B'') \) has (ancestor of) \( t \); (ii) \( P(B) \) has (ancestor of) \( t'' \). In case (i), since \( t' \) is equal to \( t \) or is a descendent of \( t \), \( P(B') \) has (ancestor of) \( t' \) which is a contradiction. In case (ii), since \( t'' \) is an ancestor of \( t' \), \( P(B) \) has an ancestor of \( t(B') \), so \( \{B, B'\} \) has a valid labeling which is a contradiction. So, \( \{B, B''\} \) does not have a valid labeling.

### 7.2.3 The Construction.

At time \( \tau_{\ell-1} \leq \tau \leq \tau_{\ell} \) we allocate \( token_{\ell}(B) \) tokens to any ball \( B = (x, r_1, r_2) \in Z_{\tau} \), where

\[
    token_{\ell}(B) = \begin{cases} 
        \rho_{\ell} - C_4 \cdot \rho_{\ell+1} & \text{if } B \in FBag_{\ell} \\
        \left[ r_2 - r_1 - C_4 \cdot \rho_{\ell} \right]^+ & \text{if } B \text{ is not child-avoiding} \\
        \left[ \frac{r_2 - r_1 - C_4 \cdot \rho_{\ell}}{2(1+C_4)} \right]^+ & \text{otherwise.} 
    \end{cases}
\]

Note that we allocate significantly smaller number of tokens to the child-avoiding balls; roughly speaking we allocate \( 1/C_4 \) fraction of their radius.
Say we process Bag$_t$ at time $\tau_{t-1} + 1 \leq \tau < \tau_t$. The details of our construction is described in Algorithm 3. We process Bag$_t$ in several steps; we start with $Z = Z_\tau$ and in each iteration of the loop we may add/remove several balls to/from $Z$. We use $Z_{\tau,s}$ to denote the set $Z$ after the $s$ iteration of the loop, i.e., $Z_{\tau,0} = Z_\tau$ and $Z_{\tau,\infty} = Z_{\tau+1}$.

After processing each Bag$_t \in FBag_t$ we keep the border balls of Bag$_t$ in a set Bor$_t$ and we insert them at the end of phase $\ell$. Initially, we let Bor$_t$ be the set of border balls of Bag$_t$ with respect to $Z_{\tau,0}$ and Int$_t$ be the set of interior balls. We update these sets in each iteration of the loop. We use Bor$_{t,s}$, Int$_{t,s}$ to denote the sets Bor$_t$, Int$_t$ after the $s$ iterations of the loop respectively.

Algorithm 3 Construction of $Z_{\tau+1}$ by processing Bag$_t$($\rho_\ell, b_\ell$).

**Input:** $Z_\tau$ and Bag$_t \in FBag_t$.

**Output:** $Z_{\tau+1}$

1. Let $Z = Z_\tau$, $t^*$ be father of $t$ and Bor$_t$, Int$_t$ be the border balls and interior balls of Bag$_t$ respectively. Also, let $O'(t) = \{u, v \in O(t) : \|X_u - X_v\|_1 < \rho_\ell\}$.
2. **while** $|\text{Int}_t| \geq b_\ell/2$ **do**
3.   **if** $\exists B' \in \text{Int}_t$ s.t. $B$ is in interior of a child-avoiding ball $B \notin Z$, **then**
4.     Let $B' = B(X_u, \rho_\ell)$ and $B = B(x, r_1, r_2)$.
5.     **Update** $Z$: Remove $B$ and add $B_1 = B(x, r_1, \|X_u - x\|_1 + \rho_\ell)$ and $B_2 = B(x, \|X_u - x\|_1 + \rho_\ell, r_2)$ with the same label as $B$. Add $B'$ to $Z$ and remove it from Int$_t$. **Goto** step 18.
6. **end if**
7. Let $S_1, \ldots, S_j$ be a natural decomposition of $G[V(t^*) - V(t)]$ into $k/2$ connected subgraphs as defined in Definition 2.9. \textbf{Goto} **Lemma** 7.15 we show that $j \leq 2|O(t)|/k$.
8. Let $U \subseteq V(t)$ be the centers of balls of Int$_t$, $V_i := \{v \in S_i : \{u, v\} \in O'(t)\}$ be the vertices of $S_i$ that are incident to an edge of $O'(t)$ and $U_i := \{u \in U : \exists v \in S_i, \{u, v\} \in O'(t)\}$. \textbf{Recall} that every vertex of $U$ is incident to an edge of $O'(t)$, so $\cup_{i=1}^{j} U_i = U$. Also, since Bag$_t$ is a 12$C_3$/k-assigned bag, $|U| = |\text{Int}_t| \geq b_\ell/2 \geq \frac{6C_3|O(t)|}{k}$.
9. Let $i = \text{argmax}_{1 \leq i \leq j} |U_i|$. \textbf{Goto} $|U_i| \geq |U|/j \geq 3C_3$.
10. Let $B = B(x, r_1, r_2) \in Z$ such that a ball of Int$_t$ with center in $U_i$ is in the interior of $B$.
11. We define $r'_1 = r_1 \vee \min_{v \in V_i} \|x - X_v\|_1$ and $r'_2 = r_2 \wedge \max_{v \in V_i} \|x - X_v\|_1$.
12. Let $U_B \subseteq U$ be the vertices of $U$ in the ball $B(x, r'_1 - \rho_\ell, r'_2 + \rho_\ell)$ and Int$_B$ be balls of Int$_t$ where their centers are in $U_B$. \textbf{We may have $U_i \nsubseteq U_B$ but all vertices of $U_B$ are in $B$.}
13. **if** $|\text{Int}_B| : \rho_\ell > 3(r'_2 - r'_1)$ **then** \textbf{Goto} \textbf{i.e., if Int$_B$ is a $3$-compact bag of balls}.
14. **Update** $Z$: Remove $B$ and add $B_1 = B(x, r_1, r'_1 - 2\rho_\ell)$ and $B_2 = B(x, r'_2 + 2\rho_\ell, r_2)$ with the same label as $B$. Add all balls of Int$_B$ to $Z$ and remove them from Int$_t$.
15. **else**
16. **Update** $Z$: Remove $B$ and add $B_1 = B(x, r_1, r'_1)$ and $B_2 = B(x, r'_2, r_2)$ to $Z$, with the same label as $B$. Add a new (non child-avoiding) ball $B_3 = B(x, r'_1 + \rho_\ell, r'_2 - \rho_\ell)$ with $t(B_3) = t$ and $P(B_1)$ has all nodes $t' \in P(B)$ such that $t'$ is a descendant of $t$. Add a child-avoiding ball $B_4 = B(x, r'_1, r'_2)$ with $t(B_4) = t(B)$, $t_d(B_4) = t$ and $P(B_3) = P(B)$. Remove all balls of Int$_B$ from Int$_t$. \textbf{Note that no balls of Int$_t$ - Int$_B$ is in the interior of $B_1$ or $B_2$.}
17. **end if**
18. Move all balls of Int$_t$ that become a border ball w.r.t. $Z$ to Bor$_t$.
19. **end while**

return $Z$.

The following is the main result of this part. It is easy to see that Proposition 7.9 follows simply
Lemma 7.14. For any \( \tau, s \geq 0 \) the following holds. The set \( \mathcal{Z}_{\tau,s} \) has a valid labeling. If we allocate \( \text{token}_t(B) \) tokens to any ball \( B(x, r_1, r_2) \in \mathcal{Z}_\tau \), then we can distribute these tokens among the bags of balls that we have processed by time \( \tau \) such that for any \( i < \ell \), any \( t \in T_i \) receives at least \( \frac{C_2}{\alpha_3} \cdot b_i \cdot \rho_\ell \) tokens, any \( t \in T_\ell \), that is processed by time \( \tau \), receives at least \( \frac{C_2}{\alpha_3} \cdot (b_\ell - |\text{Bor}_t|) \cdot \rho_\ell \) tokens, and the node \( t \) we are processing at time \( \tau \) receives at least \( \frac{C_2}{\alpha_3} \cdot (b_\ell - |\text{Int}_t| - |\text{Bor}_t|) \cdot \rho_\ell \) tokens.

We prove the above lemma by an induction on \( \tau, s \). From now on, we assume that all conclusions of the lemma hold for \( \tau, s \) and we prove the same properties holds for \( \tau, s + 1 \). We construct \( \mathcal{Z}_{\tau,s+1} \) (from \( \mathcal{Z}_{\tau,s} \) in three steps of the loop, i.e., steps 5, 14, 16. We analyze each of these steps in the following three cases.

Case 1: A ball \( B' \in \text{Int}_{t,s} \) is in interior of a child-avoiding ball \( B = B(x, r_1, r_2) \in \mathcal{Z}_{\tau,s} \).

In this case by Lemma 7.12, for any ball \( B'' \in \mathcal{Z}_{\tau,s} \) such that \( B \neq B'' \), \( \{B', B''\} \) has a valid labeling. Since by definition, \( B' \) does not intersect any of \( B_1, B_2, \mathcal{Z}_{\tau,s+1} \) has a valid labeling. We send all tokens of \( B_1 \) and \( B_2 \) and \( \rho_\ell/2 \) of the tokens of \( B' \) to \( B \) and we re-distribute them by the induction hypothesis. We send the rest of the tokens of \( B' \) to \( t \). Then, \( B \) receives,

\[
\text{token}_t(B_1) + \text{token}_t(B_2) + \frac{\rho_\ell}{2} \geq \frac{(r_1 - r_2 - \rho_\ell) - 2C_4 \cdot \rho_\ell + \rho_\ell(1 + C_4)}{2(1 + C_4)} = \text{token}_t(B).
\]

On the other hand, \( t \) receives

\[
\text{token}_t(B') - \frac{\rho_\ell}{2} \geq \rho_\ell - C_4 \cdot \rho_{\ell+1} - \frac{\rho_\ell}{2} \geq \rho_\ell/4.
\]

new tokens, where we used \( \rho_{\ell+1} \leq \lambda \cdot \rho_\ell \) and \( \lambda \leq 1/4C_4 \). Since \( |\text{Bor}_{t,s+1}| + |\text{Int}_{t,s+1}| = |\text{Bor}_{t,s}| + |\text{Int}_{t,s}| + 1 \) we are done by induction.

Now suppose that the above does not happen. Consider the induced graph \( G[V(t^*) - V(t)] \). Note that this graph may be disconnected. Let \( S_1, S_2, \ldots, S_j \) be a natural decomposition of this graph as defined in Definition 2.9. In the following lemma we show that \( j \leq 2|\mathcal{O}(t)|/k \).

Lemma 7.15. For any node \( t \in \mathcal{T}(k,..) \) with father \( t^* \), let \( S_1, S_2, \ldots, S_j \) be a natural decomposition of \( G[V(t^*) - V(t)] \) into \( k/2 \)-connected subgraphs. Then,

\[
j \leq \frac{2|\mathcal{O}(t)|}{k}.
\]

Proof. By definition of \( \mathcal{T} \), \( G[V(t^*)] \) is \( k \)-connected. Therefore, for any \( 1 \leq i \leq j \),

\[
\partial_{G[V(t^*)]}(S_i) \geq k.
\]

Therefore,

\[
j \cdot k \leq \sum_{i=1}^{j} \partial_{G[V(t^*)]}(S_i) = \partial_{G[V(t^*)]}(V(t)) + \sum_{i=1}^{j} \partial_{G[V(t^*) - V(t)]}(S_i) = |\mathcal{O}(t)| + \sum_{i=1}^{j} \partial_{G[V(t^*) - V(t)]}(S_i).
\]

But, by Lemma 2.10, the second term in the RHS is at most \((j - 1)(k/2 - 1)\). Therefore, \( j \leq 2|\mathcal{O}(t)|/k \).
Lemma 7.16. If $U_i \not\subseteq U_B$, then $r_2' - r_1' \geq (C_3 - 1) \cdot \rho_t$.

Proof. First, we show that there is a vertex $v \in V_i$ such that $X_v \not\in B$. For the sake of contradiction assume $V_i \subseteq B$. We show that any vertex $u \in U_i$ is in $U_B$ which is a contradiction. Fix a vertex $u \in U_i$. By definition, there is vertex $v \in V_i$ such that $\{u, v\} \in \mathcal{O}'(t)$. Since $X_v \in B$, by the definition of $r_1', r_2'$, $r_1' \leq \|X_v - x\|_1 \leq r_2'$. So, $X_u \in B(x, r_1' - \rho_t, r_2' + \rho_t)$, i.e., $u \in U_B$. This is a contradiction.

Now, let $v \in V_i$ such that either $\|X_v - x\|_1 \geq r_2$ or $\|X_v - x\|_1 \leq r_1$. Here, we assume the former; the other case can be analyzed similarly. Then, we have $r_2' = r_2$. But by definition of $B$, there is a ball $B' = (X_u, \rho_t) \in \text{Int}_{t,s}$ where $u \in U_i$ such that $B'$ is in the interior of $B$. Since $u \in U_i$, there is a vertex $v \in V_i$ such that $\|X_u - X_v\|_1 < \rho_t$. Therefore,

$$r_1' \leq \|x - X_v\|_1 \leq \|x - X_u\|_1 + \rho_t \leq r_2 - C_3 \rho_t + \rho_t.$$ 

where the last inequality uses that $B'$ is in the interior of $B$. So, $r_2' - r_1' \geq (C_3 - 1) \rho_t$. 

Case 2: $|\text{Int}_B| \cdot \rho_t \geq 3(r_2' - r_1')$.

First, we show that $Z_{t,s+1}$ has a valid labeling. Then, we distribute the tokens. To show that $Z_{t,s+1}$ has a valid labeling, first we argue that all balls of $\text{Int}_B$ are in the interior of $B$. Fix a ball $B' \in \text{Int}_B$, we show $B'$ is in the interior of $B$. First, $\{B', B\}$ does not have a valid labeling. Because i) $B'$, $B$ intersect by the definition of $\text{Int}_B$ and ii) a ball of $\text{Bag}_t$ is in the interior of $B$ and all balls of $\text{Bag}_t$ have similar labels. Second, since $\text{Int}_B \subseteq \text{Int}_{t,s}$, $B'$ is an interior ball. Therefore, by Lemma 7.12, $B'$ is in the interior of $B$.

Now, by Lemma 7.12, for any $B' \in \text{Int}_B$ and $B'' \in Z_{t,s}$ where $B'' \neq B$, $\{B', B''\}$ has a valid labeling. Furthermore, by construction, $B_1, B_2$ do not intersect any balls of $\text{Int}_B$. Hence, $Z_{t,s+1}$ has a valid labeling.

Next, we describe the distribution of tokens allocated to balls of $Z_{t,s}$. Before that, we show that $|\text{Int}_B| \geq 3(C_3 - 1)$. We consider two cases. If $U_i \subseteq U_B$. Then, by the assumption that $\text{Bag}_t$ is a $12C_3/k$-assigned bag of balls,

$$|\text{Int}_B| = |U_B| \geq |U_i| \geq 3C_3.$$ 

The last inequality holds by the definition of $i$ and Lemma 7.15. Otherwise, $U_i \not\subseteq U_B$. Then, by Lemma 7.16,

$$|\text{Int}_B| \geq \frac{3(r_2' - r_1')}{\rho_t} \geq \frac{3(C_3 - 1) \cdot \rho_t}{\rho_t} = 3(C_3 - 1).$$ 

Therefore, $|\text{Int}_B| \geq 3(C_3 - 1)$.

Now, we send all tokens of $B_1, B_2$ and $3/4$ of the tokens of each ball of $\text{Int}_B$ to $B$ and we re-distribute them by the induction hypothesis. $B$ receives,

$$\text{token}_t(B_1) + \text{token}_t(B_2) + \frac{3}{4}|\text{Int}_B| \cdot (\rho_t - 4C_4 \cdot \rho_t + \frac{5}{6}|\text{Int}_B| \cdot \rho_t)$$

\[
\geq r_2 - r_1 - 2\rho_t - (r_2' - r_1') - 2C_4 \cdot \rho_t + \frac{3}{4} \cdot \frac{5}{6}|\text{Int}_B| \cdot \rho_t
\]

\[
\geq \text{token}_t(B) - (2 + C_4) \cdot \rho_t + \frac{7}{24}|\text{Int}_B| \cdot \rho_t
\]

\[
\geq \text{token}_t(B) - (2 + C_4) \cdot \rho_t + \frac{7}{8}(C_3 - 1) \cdot \rho_t \geq \text{token}_t(B).
\]
where the first inequality uses $\rho_{\ell+1} < \lambda \rho_{\ell}$ and $\lambda < 1/6C_4$, the second inequality uses the assumption $3(r'_2 - r'_1) < |\text{Int}_B| \cdot \rho_{\ell}$, the third inequality uses $|\text{Int}_B| \geq 3(C_3 - 1)$ and the last inequality uses $C_3 \geq 2C_4 + 4$. On the other hand, each ball $B' \in \text{Int}_B$ sends
\[
\frac{1}{4} \text{token}_{\ell}(B') \geq \frac{1}{4} \cdot \frac{5}{6} \rho_{\ell}.
\]
to $t$. So, $t$ receives $|\text{Int}_B| \cdot \rho_{\ell}/4$ new tokens. Since
\[
|\text{Bor}_{t,s+1}^\Delta| + |\text{Int}_{t,s+1}^\Delta| = |\text{Bor}_{t,s}^\Delta| + |\text{Int}_{t,s}^\Delta| + |\text{Int}_B^\Delta|,
\]
we are done by induction.

**Case 3:** $|\text{Int}_B| \cdot \rho_{\ell} \leq 3(r'_2 - r'_1)$.

We need to verify that $B_3, B_4$ has a valid labeling, $Z_{t,s+1}$ has a valid labeling and that $B_3, B_4$ satisfy property 3 of Figure 8. We start with the latter. First, since Case 1 does not happen, $B$ is not a child-avoiding ball. Since the label of $B_3, B_4$ is defined as required in property 3 and $B_3, B_4$ are inside $B$ we only need to show that $t(B)$ is an ancestor of $t$. By Lemma 7.10, either $t(B)$ is equal to $t$ or is an ancestor of $t$. So, we just need to argue that $t(B) \neq t$. If $t(B) = t$, then either $B \in \text{Bag}_t$ or $B$ is constructed in an earlier step of the algorithm, say in $s' \leq s$ iteration of the loop. The former cannot happen because balls of $\text{Bag}_t$ do not intersect. The latter does not happen because whenever we construct a new ball in step 16 we delete all balls of $\text{Int}_t$ that intersects with the new ball. Therefore $t(B) \neq t$.

In the following claim we show that $Z_{t,s+1}$ has a valid labeling.

**Claim 7.17.** The set $Z_{t,s+1}$ has a valid labeling.

**Proof.** We start showing that $B_3$ satisfies property 1 of Figure 7. Since $B$ has a valid labeling and $P(B_3) \subseteq P(B)$, for any $t' \in P(B_3)$, endpoints of edges of $P(t')$ have distance at least $r_2 \geq r_2' - \rho_{\ell}$ of $x$. By definition of $U_i$ there are vertices $u_1, u_2 \in U_i$ such that $\|x - X_{u_1}\| < r'_1 + \rho_{\ell}$ and $\|x - X_{u_2}\| \geq r'_2 - \rho_{\ell}$. It remains to argue that for any $t' \in P(B_3)$, $u_1 \notin V(t')$. This is because, $u_1 \notin U_i$ is incident to an edge $e$ of $O'(t)$. If $u_1 \in V(t')$ then $e \in P(t')$ so an endpoint of an edge of $P(t')$ is in $B$ which is a contradiction.

Now, we show $Z_{t,s+1} - B_4$ has a valid labeling. First, $\{B_1, B_2, B_3\}$ has a valid labeling because no pair of them intersect. Second, fix a ball $B' \neq B$ in $Z_{t,s}$ such that $B'$ intersects $B_3$, we show $\{B_3, B'\}$ has a valid labeling. If $t(B), t(B')$ are not ancestor-descendent, then $t, t(B')$ is not ancestor-descendent, so $\{B_3, B'\}$ has a valid labeling. Otherwise, since $B, B' \in Z_{t,s}$ intersect and $B$ is not child-avoiding, by 3 of Figure 7 one of the following holds: i) $P(B')$ has (an ancestor of) $t(B)$, or $B'$ is child-avoiding and $t_d(B')$ is (an ancestor of) $t(B)$, or ii) There is $t' \in P(B)$ that is (an ancestor of) $t(B')$. In case (i), since $t(B)$ is an ancestor of $t(B_3)$, $\{B_3, B'\}$ has a valid labeling. In case (ii), if $t'$ is an ancestor of $t$, then $\{B, B_3\}$ has a valid labeling, but this cannot happen because a ball of $\text{Bag}_t$ is in the interior of $B$. Otherwise, $t'$ is a descendant of $t$. But then, by definition of $P(B_3)$, $t' \in P(B_3)$, so $\{B_3, B'\}$ has a valid labeling.

Now we analyze the label of $B_4$. First, we show $B_4$ satisfies property 2 of Figure 7. Since $B$ has a valid labeling and $P(B_4) = P(B)$, for any $t' \in P(B_4)$, endpoints of all edges of $P(t')$ have distance at least $r_2 \geq r_2'$ from $x$. By the definition of $V_i$ there are vertices $v_1, v_2 \in V_i$ such that $\|x - X_{v_1}\| \leq r'_1$ and $\|x - X_{v_2}\| \geq r'_2$. Since $V_i \subseteq S_i$ and $S_i$ is $k/2$-connected in $G[V(t^*) - V(t)]$, there are $k/2$ edge disjoint paths from $v_1$ to $v_2$ in $G[V(t(B)) - V(t)]$. So, we just need to argue
Lemma 7.16, intersect. 

Finally, we show \( \mathcal{Z}_{\tau,s+1} \) has a valid labeling. \( \{B_3, B_4\} \) has a valid labeling since \( t_d(B_4) = t = t(B_3) \). For any \( B' \in \mathcal{Z}_{\tau,s} \) such that \( B' \neq B \), \( \{B, B'\} \) has a valid labeling since \( \{B, B'\} \) has a valid labeling and \( B_4 \) is inside \( B \). Furthermore, \( \{B_1, B_2, B_4\} \) has a valid labeling since no pair of them intersect. \( \square \)

It remains to distribute the tokens. First, we show that \( r'_2 - r'_1 \geq (C_3 - 1) \cdot \rho_\ell \). If \( U_i \nsubseteq U_B \), then by Lemma 7.16, \( r'_2 - r'_1 \geq (C_3 - 1) \cdot \rho_\ell \). Otherwise, since \( \text{Bag}_t \) is a 12\( C_3/k \)-assigned bag of balls,

\[
r'_2 - r'_1 \geq \frac{1}{3} \left| \text{Int}_B \right| \cdot \rho_\ell \geq \frac{1}{3} \left| U_i \right| \cdot \rho_\ell \geq C_3 \cdot \rho_\ell.
\]

We send all tokens of \( B_1, B_2, B_3 \) and \( 2(C_4 + 2) \rho_\ell \) tokens of \( B_4 \) to \( B \) and we re-distribute them by the induction hypothesis. We send the rest of the tokens of \( B_4 \) to \( t \). Ball \( B \) receives,

\[
\sum_{i=1}^{3} \text{token}_\ell(B_i) + (2C_4 + 2) \cdot \rho_\ell \geq r_2 - r_1 - 2\rho_\ell - 3C_4 \rho_\ell + (2C_4 + 2) \rho_\ell = \text{token}_\ell(B).
\]

On the other hand, \( t \) receives,

\[
\text{token}_\ell(B_4) - (2C_4 + 2) \rho_\ell = \frac{r'_2 - r'_1 - C_4 \cdot \rho_\ell - 4(1 + C_4)^2 \cdot \rho_\ell}{2(1 + C_4)} \geq \frac{r'_2 - r'_1 - (C_3 - 1) \rho_\ell / 2}{2(1 + C_4)} \geq \frac{r'_2 - r'_1}{4(1 + C_4)} \geq \left| \text{Int}_B \right| \cdot \rho_\ell \geq \frac{C_4 \left| \text{Int}_B \right| \cdot \rho_\ell}{6C_3},
\]

new tokens. In the first inequality we used \( C_3 \geq (C_4 + 1) + 4(C_4 + 1)^2 \), the second inequality uses \( r'_2 - r'_1 \geq (C_3 - 1) \cdot \rho_\ell \), and the third inequality uses the assumption \( r'_2 - r'_1 \geq \frac{1}{3} \left| \text{Int}_B \right| \cdot \rho_\ell \).

### 7.2.4 Post Processing

Say we have processed all \( \text{Bag}_t \in \text{FBag}_\ell \) and we are the end of phase \( \ell \), i.e., time \( \tau_\ell \). We may add several border balls. We need to send at least \( \frac{C_4}{6C_3} | \text{Bor}_\ell | \cdot \rho_\ell \) new tokens to any node \( t \in T_\ell \). We define the shrink operator as follows: For any ball \( B = B(x, r_1, r_2) \in \mathcal{Z}_{\tau_\ell} \),

\[
\text{shrink}_\ell(B) = \begin{cases} 
B & \text{if } B \in \text{FBag}_\ell \\
B(x, r_1 + (C_3 + 1) \rho_\ell, r_2 - (C_3 + 1) \rho_\ell) & \text{if } B \notin \text{FBag}_\ell \text{ and } r_2 - r_1 > 2(C_3 + 1) \rho_\ell \\
B(x, 0) & \text{otherwise.}
\end{cases}
\]

Let

\[
b = \sum_{t \in T_\ell} | \text{Bor}_\ell |,
\]

\[
\text{excess} = \sum_{B \in \mathcal{Z}_{\tau_\ell}} (\text{token}_{\ell+1}(B) - \text{token}_\ell(B)).
\]
Think of excess as the additional number of tokens that we gain for each ball \( B \in \mathcal{Z}_{\tau_\ell} \) when we go to the new phase \( \ell + 1 \). Our idea is simple. If excess is very large then we do not add any of the border balls and we just distribute excess between all nodes of \( T_\ell \). Otherwise, we shrink balls of \( \mathcal{Z}_{\tau_\ell} \) and we add the border balls.

**Case 1:** \( \text{excess} \geq \frac{C_4}{6C_3} \cdot b \cdot \rho_\ell \).

In this case, we do not add any of the border balls and we simply let \( \mathcal{Z}_{\tau_{\ell + 1}} = \mathcal{Z}_{\tau_\ell} \).

Now, observe that for any ball \( B \in \mathcal{Z}_{\tau_\ell} \), we have \( \text{token}_{\ell + 1}(B) - \text{token}_\ell(B) \) additional token that is in no use for \( B \). We distribute these tokens between the nodes of \( T_\ell \) proportional to their number of border balls. More precisely, for any ball \( B \in \mathcal{Z}_{\tau_\ell} \) and \( t \in T_\ell \), we send \( \frac{|\text{Bor}_t|}{b} \cdot (\text{token}_{\ell + 1}(B) - \text{token}_\ell(B)) \) tokens to \( t \). Therefore, \( t \) receives

\[
\sum_{B \in \mathcal{Z}_{\tau_\ell}} \frac{|\text{Bor}_t|}{b} \cdot (\text{token}_{\ell + 1}(B) - \text{token}_\ell(B)) = \frac{|\text{Bor}_t|}{b} \cdot \text{excess} \geq \frac{C_4}{6C_3} \cdot |\text{Bor}_t| \cdot \rho_\ell,
\]

and we are done.

**Case 2:** \( \text{excess} < \frac{C_4}{6C_3} \cdot b \cdot \rho_\ell \).

For each ball \( B \in \mathcal{Z}_{\tau_\ell} \) we include shrink\(_\ell(B)\) in \( \mathcal{Z}_{\tau_{\ell + 1}} \). We also add all balls of \( \text{Bor}_t \) for all \( t \in T_\ell \) to \( \mathcal{Z}_{\tau_{\ell + 1}} \). By Lemma 7.13 any border ball \( B \in \text{Bor}_t \) is not in the interior of any ball of \( \mathcal{Z}_{\tau_\ell} \). By the definition of the shrink operator, and using the fact that balls of \( \text{FBag}_\ell \) do not intersect, any ball of \( \cup_{t \in T_\ell} \text{Bor}_t \) do not intersect any ball of \( \mathcal{Z}_{\tau_{\ell + 1}} \). So, \( \mathcal{Z}_{\tau_{\ell + 1}} \) has a valid labeling.

It remains to distribute the tokens. First, we prove a technical lemma.

**Lemma 7.18.** If \( \text{excess} < \frac{C_4}{6C_3} \cdot b \cdot \rho_\ell \), then

\[
b \cdot \rho_\ell \geq 2 \sum_{B \in \mathcal{Z}_{\tau_\ell}} (\text{token}_\ell(B) - \text{token}_{\ell + 1}(\text{shrink}_\ell(B))).
\]

**Proof.** It is sufficient to show that for any ball \( B = B(x, r_1, r_2) \in \mathcal{Z}_{\tau_\ell} \)

\[
\text{token}_{\ell + 1}(B) - \text{token}_\ell(B) \geq \frac{C_4}{3C_3} \cdot (\text{token}_\ell(B) - \text{token}_{\ell + 1}(\text{shrink}_\ell(B))). \tag{41}
\]

First, note that if \( \text{token}_\ell(B) = 0 \) then the above holds trivially. We consider three cases. i) \( B \in \text{FBag}_\ell \). In this case both sides of the above inequality is zero. This is because \( \text{shrink}_\ell(B) = B \) and \( \text{token}_\ell(B) = \text{token}_{\ell + 1}(B) \). ii) \( B \) is not a child-avoiding ball. Since \( \text{token}_\ell(B) > 0 \), \( r_2 - r_1 > C_4 \cdot \rho_\ell \). Therefore,

\[
\text{token}_{\ell + 1}(B) - \text{token}_\ell(B) = C_4 \cdot (\rho_\ell - \rho_{\ell + 1}) \geq \frac{2}{3} \cdot C_4 \cdot \rho_\ell
\]

\[
\text{token}_\ell(B) - \text{token}_{\ell + 1}(\text{shrink}_\ell(B)) \leq 2(C_3 + 1)\rho_\ell + C_4 \cdot (\rho_{\ell + 1} - \rho_\ell) \leq 2C_3 \cdot \rho_\ell.
\]

using \( \rho_{\ell + 1} \leq \rho_\ell / 3 \) and \( C_4 \geq 3 \). So, (41). iii) \( B \) is a child-avoiding ball. Equation (41) is equivalent to case (ii) up to a \( 2(1 + C_4) \) factor in both sides of the inequality.  

\[\square\]
For any ball $B \in \text{Bor}_t$ and any ball $B' \in Z_{\tau t}$, we send

$$\frac{\rho_t}{2} \cdot \frac{\text{token}_t(B') - \text{token}_{t+1}(\text{shrink}_t(B'))}{\sum_{B'' \in Z_{\tau t}} \text{token}_t(B'') - \text{token}_{t+1}(\text{shrink}_t(B''))}$$
tokens to $B'$ and we send the remaining tokens to $t$. For any ball $B \in Z_{\tau t}$, also send all of the tokens of $\text{shrink}_t(B)$ to $B$.

Therefore, by Lemma 7.18, any ball $B \in Z_{\tau t}$ receives at least

$$\text{token}_{t+1}(\text{shrink}_t(B)) + b \cdot \frac{\rho_t}{2} \cdot \frac{\text{token}_t(B) - \text{token}_{t+1}(\text{shrink}_t(B))}{\sum_{B' \in Z_{\tau t}} \text{token}_t(B') - \text{token}_{t+1}(\text{shrink}_t(B'))} \geq \text{token}_{t+1}(\text{shrink}_t(B)) + (\text{token}_t(B) - \text{token}_{t+1}(\text{shrink}_t(B))) = \text{token}_t(B),$$

that we re-distribute by the induction hypothesis. On the other hand, any $t \in T_t$ receives,

$$|\text{Bor}_t| \cdot (\rho_t - \rho_t/2 - C_4 \rho_{t+1}) \geq |\text{Bor}_t| \cdot \rho_t/4$$

new tokens, and we are done with the induction. This completes the proof of Lemma 7.14.

Acknowledgment.

We would like to thank James R Lee, Satish Rao, Amin Saberi, Luca Trevisan and Umesh Vazirani for enlightening and stimulating discussions.

References

[AGM++10] Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An $o(\log n/ \log \log n)$-approximation algorithm for the asymmetric traveling salesman problem. In SODA, pages 379–389, 2010. 1, 58, 59, 60

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expanders, geometric embeddings and graph partitioning. J. ACM, 56:5:1–5:37, April 2009. 10

[Blä02] M. Bläser. A new approximation algorithm for the asymmetric tsp with triangle inequality. In SODA, pages 638–645, 2002. 1

[BSS14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. SIAM Review, 56(2):315–334, 2014. 3

[BV06] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2006. 13, 24

[CA13] Nik Weaver Charles Akemann. A lyapunov type theorem from kadison-singer. 2013. 4

[CGK06] Moses Charikar, Michel X. Goemans, and Howard Karloff. On the integrality ratio for the asymmetric traveling salesman problem. MOR, 31(2):245–252, 2006. 1

[Chr76] Nicos Christofides. Worst case analysis of a new heuristic for the traveling salesman problem. Report 388, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA, 1976. 1

56
AshokK. Chandra, Prabhakar Raghavan, WalterL. Ruzzo, Roman Smolensky, and Prasoon Tiwari. The electrical resistance of a graph captures its commute and cover times. *computational complexity*, 6(4):312–340, 1996.

Jeff Erickson and Anastasios Sidiropoulos. A near-optimal approximation algorithm for asymmetric TSP on embedded graphs. In *SOCG*, page 130, 2014.

A. M. Frieze, G. Galbiati, and F. Maffioli. On the worst-case performance of some algorithms for the asymmetric traveling salesman problem. *Networks*, 12:23–39, 1982.

Wai Shing Fung, Ramesh Hariharan, Nicholas J. A. Harvey, and Debmalya Panigrahi. A general framework for graph sparsification. In *STOC*.

U. Feige and M. Singh. Improved approximation ratios for traveling salesman tours and path sin directed graphs. In *APPROX*, pages 104–118, 2007.

Arpita Ghosh, Stephen Boyd, and Amin Saberi. Minimizing effective resistance of a graph. *SIAM Rev.*, 50(1):37–66, February 2008.

Luis A. Goddyn. Some open problems i like. available at http://www.math.sfu.ca/~goddyn/Problems/problems.html, 2004.

M. Held and R.M. Karp. The traveling salesman problem and minimum spanning trees. *Operations Research*, 18:1138–1162, 1970.

Nicholas J. A. Harvey and Neil Olver. Pipage rounding, pessimistic estimators and matrix concentration. In *SODA*, pages 926–945, 2014.

F. Jaeger. On circular flows in graphs, finite and infinite sets. *Colloquia Mathematica Societatis Janos Bolyai 37*, pages 391–402, 1984.

David R. Karger. Random sampling in cut, flow, and network design problems. *MOR*, 24(2):383–413, May 1999.

H. Kaplan, M. Lewenstein, N. Shafrir, and M. Sviridenko. Approximation algorithms for asymmetric TSP by decomposing directed regular multigraphs. *J. ACM*, pages 602–626, 2005.

László Mikhó Lovász, Carsten Thomassen, Yezhou Wu, and Cun-Quan Zhang. Nowhere-zero 3-flows and modulo k-orientations. *J. Comb. Theory, Ser. B*, 103(5):587–598, 2013.

L. Mirsky. A trace inequality of john von neumann. *Monatshefte für Mathematik*, 79(4):303–306, 1975.

Adam Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem. In *FOCS*, 2013.
Proof of Theorem 1.2. For a feasible vector $x$ of LP (1), let $c(x) = \sum_{u,v} c(u,v) \cdot x_{u,v}$. We use the following theorem that is proved in [AGM+10].

**Theorem A.1.** For a feasible solution $x$ of LP (1) and a spanning tree $T$ such that for any $S \subseteq V$,

$$|T(S,\overline{S})| \leq \alpha \cdot \sum_{u \in S, v \notin S \text{ or } u \notin S, v \in S} x_{u,v} + x_{v,u} =: \alpha \cdot x(S,\overline{S}),$$

and $\sum_{(u,v) \in T} c(u,v) \leq \beta \cdot c(x)$. Then, there is a polynomial time algorithm that finds a tour of length $O(\alpha + \beta) \cdot c(x)$.

Suppose for $C \geq 1$ any $C \cdot \log(n)$-connected graph has a $f(\log(n))/\log(n)$-thin tree. Given a feasible solution $x$ of LP (1), for a constant $C' \geq 4C$, we sample $C' \cdot \log(n)$ arcs where the probability of choosing each arc $(u,v)$ is proportional to $x_{u,v}$. We drop the direction of the arcs and we call the sampled graph $G = (V,E)$. Since $x(S,\overline{S}) \geq 1$ for all $S \subseteq V$, it follows by the seminal work of Karger [Kar99] that for a sufficiently large $C'$, with high probability, for any $S \subseteq V$, $|E(S,\overline{S})|$ has value between $1/2$ and $2$ times $C' \log(n) x(S,\overline{S})$. For a set $F \subseteq E$, $c(F) := \sum_{(u,v) \in F} \min\{c(u,v), c(v,u)\}$. Since above happens with high probability, by Markov inequality we also assume that $c(E) \leq C' \cdot \log(n) \cdot c(x)$.

Since $x(S,\overline{S}) \geq 2$ and $C' > 4C$, $G$ is $2C \log(n)$ connected. Let $\alpha = \frac{f(\log(n))}{\log(n)}$. By the assumption of the theorem, $G$ has an $\alpha$-thin tree, say $T_1$. Because of the thinness of $T_1$, $G(V,E - T_1)$ is...
$2C \log(n)(1-\alpha) \geq C \cdot \log(n)$-connected. Therefore, it has a $\alpha$-thin tree. By repeating this argument, we can find $\ell = \frac{1}{2\alpha}$ edge-disjoint $2\alpha$-thin spanning trees in $G, T_1, \ldots, T_j$.

Without loss of generality, assume that $c(T_1) = \min_{1 \leq i \leq j} c(T_i)$. We show that $T_1$ satisfies the conditions of the above theorem. First, since $c(T_1) = \min_{1 \leq i \leq j} c(T_i)$,

$$c(T_1) = \frac{c(E)}{\ell} \leq \frac{C' \cdot \log(n) \cdot c(x)}{\ell} = 2C' \cdot f(\log(n)) \cdot c(x).$$

On the other hand, since $T_1$ is $\alpha$-thin with respect to $G$, for any set $S \subseteq V$,

$$|T_1(S, \overline{S})| \leq \alpha \cdot |E(S, \overline{S})| \leq 2C' \log(n) \cdot \alpha \cdot x(S, \overline{S}) = 2C' \cdot f(\log(n)) \cdot x(S, \overline{S}).$$

Therefore, the theorem follows from the application of the above theorem of [AGM+10].

**Proof of Theorem 1.6.** Akemann and Weaver proved the following extension of [MSS13].

**Theorem A.2.** Given vectors $x_1, \ldots, x_m \in \mathbb{R}^d$, such that $A = \sum_{i=1}^{m} x_i x_i^\top \preceq I$ and $\|x_i\|^2 \leq \epsilon$ for all $1 \leq i \leq m$. For any $0 < \beta < 1$, there is a set $S \subseteq [m]$ such that

$$\left\| \sum_{i \in S} x_i x_i^\top - \beta \cdot A \right\| \leq C \cdot \epsilon^{1/4},$$

for some universal constant $C > 0$.

We assign a vector $x_e = \frac{1}{\sqrt{2}} \chi_e$ to each edge $e \in E'$. First, by (i) of Lemma 2.4,

$$\sum_{e \in E'} x_e x_e^\top = \frac{1}{2} L_G^{1/2} L_{E'} L_G^{1/2} \preceq \frac{1}{2} L_G^{1/2} L_G \frac{1}{2} L_G = I.$$

Let $\epsilon$ be the maximum effective resistance of edges of $E'$,

$$\epsilon = \max_{e \in E'} \mathcal{R}e(i)(e) = \max_{e \in E'} \chi_e L_G^{1/2} L_G \chi_e = \max_{e \in E'} \|x_e\|^2.$$

Let $\beta = 2C \cdot \epsilon^{1/4}$. Then, by the above theorem, there is a set $F \subseteq E'$ such that

$$\left\| \sum_{e \in F} x_e x_e^\top - \beta \cdot \sum_{e \in E'} x_e x_e^\top \right\| \leq C \cdot \epsilon^{1/4}. \quad (42)$$

We show that $F$ is connected and $2\beta$-spectrally thin. First, by the above inequality,

$$\sum_{e \in F} x_e x_e^\top \preceq \beta \cdot \sum_{e \in E'} x_e x_e^\top + \beta \cdot I.$$

Therefore,

$$L_F = \frac{1}{\sqrt{2}} \sum_{e \in F} x_e x_e^\top L_G \leq \beta \cdot \frac{1}{\sqrt{2}} \left( \sum_{e \in E'} x_e x_e^\top + I \right) L_G^{1/2} = \beta \cdot (L_{E'} + L_G) \leq 2\beta \cdot L_G.$$

So, $F$ is $2\beta$ spectrally thin with respect to $G$. It remains to show that $F$ is connected. Fix a set $S \subseteq V$. By (42),

$$y^\top \left( \sum_{e \in F} x_e x_e^\top \right) y \geq \beta \cdot y^\top \left( \sum_{e \in E'} x_e x_e^\top \right) - C \cdot \epsilon^{1/4} \|y\|^2$$

Letting $y = L_G^{1/2} 1_S$ we get,

$$|F(S, \overline{S})| \geq \beta \cdot |E'(S, \overline{S})| - C \cdot \epsilon^{1/4} |E(S, \overline{S})| \geq 2C \cdot \epsilon^{1/4} |E(S, \overline{S})| - C \cdot \epsilon^{1/4} |E(S, \overline{S})| > 0$$

Therefore, any spanning tree of $F$ is a $2\beta$-spectrally thin tree of $G$.\qed
Proof of Proposition 1.10. Let $0 \prec D \preceq L_G$ and $F \subseteq E$ be as defined in Theorem 1.8. By Theorem 2.8, we assume $F$ is a union of $\ell$ disjoint spanning trees for $\ell = \Omega(k)$. We use the following theorem of [AGM+10].

**Theorem A.3 ([AGM+10, SV14]).** For any graph $G = (V, E)$ and any point $z$ in the interior of the spanning tree polytope of $G$, there are $\gamma : E \to \mathbb{R}$, and a $\gamma$-uniform random spanning tree distribution $\mu$, such that for any edge $e \in E$, $\mathbb{P}_{T \sim \mu} [e \in T] = z_e$.

Now, we let $z_e = 1/\ell$ for any $e \in F$, i.e., $z$ is a convex combination of $\ell$ spanning trees. By slightly perturbing $z$ we can assume $z$ is in the interior of the spanning tree polytope of $(V, F)$. So, by the above theorem, there is a $\gamma$-uniform distribution of random spanning trees, $\mu$, where the marginal of each edge is at most $1/\ell - 1$.

The rest of the proof is similar to the proof of Theorem 3.2. Let $A = L_G + D$. For any edge $e \in F$, we define $x_e = A^{-1/2} \chi_e$. Then,

$$\sum_{e \in F} x_e x_e^\top = \sum_{e \in F} A^{-1/2} \chi_e \chi_e^\top A^{-1/2} \preceq A^{-1/2} L_F A^{-1/2} \preceq I.$$ 

Letting $\epsilon \asymp \sqrt{k}$ polylog($k$), such that for any edge $e \in E$,

$$\frac{\|x_e\|^2}{\mathbb{P}_{T \sim \mu} [e \in T]} = \frac{\chi_e^\top A^{-1} \chi_e}{1/(\ell - 1)} \leq \frac{\chi_e^\top A^{-1} \chi_e}{1/(\ell - 1)} = \epsilon.$$ 

Since, Conjecture 1.9 holds true, there is a spanning tree $T$ in the support of $\mu$ such that for any $y \in \mathbb{R}^n$,

$$y^\top \left( \sum_{e \in T} x_e x_e^\top \right) y \leq O\left( \frac{1 + \epsilon}{k} \right) \cdot \|y\|^2.$$ 

We show that $T$ is $\tilde{O}(1/\sqrt{k})$-thin. For any set $S \subseteq V$, and $y = A^{1/2} 1_S$ we have,

$$|T(S, \overline{S})| = 1_S^\top A^{1/2} \left( \sum_{e \in T} x_e x_e^\top \right) A^{1/2} 1_S \leq O\left( \frac{1 + \epsilon}{k} \right) \cdot 1_S^\top A 1_S \leq \tilde{O}\left( \frac{1}{\sqrt{k}} \right) : |E(S, \overline{S})|.$$ 

The last inequality uses that $D \preceq L_G$. 

\[\square\]