Nearly Tight Low Stretch Spanning Trees

Ittai Abraham* Yair Bartal† Ofer Neiman‡

Abstract

We prove that any graph G with n points has a distribution $\mathcal{T}$ over spanning trees such that for any edge (u,v) the expected stretch $E_{T\sim\mathcal{T}}[d_T(u,v)/d_G(u,v)]$ is bounded by $\tilde{O}(\log n)$. Our result is obtained via a new approach of building "highways" between portals and a new strong diameter probabilistic decomposition theorem.

1 Introduction

Let $G=(V,E)$ be a finite graph. For any subgraph $H=(V',E')$ of G let d_H be the induced shortest path metric with respect to H. In particular, for any edge $(u,v)\in E$ and any spanning tree T of G, $d_T(u,v)$ denotes the shortest path distance between u and v in T.

Given a distribution $\mathcal{T}$ over spanning trees of G, let $\mathrm{stretch}_{\mathcal{T}}(u,v) = \mathbb{E}_{T\sim\mathcal{T}}\left[\frac{d_T(u,v)}{d_G(u,v)}\right]$ and let $\mathrm{stretch}_{\mathcal{T}}(G) = \max_{(u,v)\in E}\mathrm{stretch}_{\mathcal{T}}(u,v)$. Let $\mathrm{stretch}(n) = \max_{G=(V,E)||V|=n}\inf_{\mathcal{T}}\{\mathrm{stretch}_{\mathcal{T}}(G)\}$.

Initial results were obtained by Alon, Karp, Peleg and West [3] showing that $\Omega(\log n) = \mathrm{stretch}(n) = \exp(O(\sqrt{\log n\log\log n}))$. The upper bound was significantly improved to $O((\log n)^2\log\log n)$ by Elkin, Emek, Spielman and Teng [11][1]. For the class of Series-Parallel graphs Emek and Peleg [12] obtained a bound of $\Theta(\log n)$. The main result of this paper is a new upper bound on $\mathrm{stretch}(n)$ that is tight up to polylogarithmic factors[2].

*School of Engineering and Computer Science, Hebrew University, Israel. Email: ittaia@cs.huji.ac.il.

†School of Engineering and Computer Science, Hebrew University, Israel and Center of the Mathematics of Information, Caltech, CA, USA. Email: yair@cs.huji.ac.il. Supported in part by a grant from the Israeli Science Foundation (195/02) and in part by a grant from the National Science Foundation (NSF CCF-065253).

‡School of Engineering and Computer Science, Hebrew University, Israel. Email: neiman@cs.huji.ac.il. Supported in part by a grant from the Israeli Science Foundation (195/02).

[1] In fact these result apply to a similar notion, $\mathrm{avg}-\mathrm{stretch}(n) = \max_{G=(V,E)||V|=n}\inf_{\mathcal{T}}\{\frac{1}{|E|}\sum_{(u,v)\in E}\frac{d_T(u,v)}{d_G(u,v)}\}$ which is equivalent up to a constant factor to $\mathrm{stretch}(n)$.

[2] [10] announced $\mathrm{stretch}(n) = O((\log n)^2)$, but this claim was subsequently withdrawn by the authors

Theorem 1.

$$\mathrm{stretch}(n) = O\left(\log n\cdot\log\log n\cdot(\log\log\log n)^3\right)$$

Remark 1. *For ease of presentation we show a slightly weaker bound of*

$$\mathrm{stretch}(n) = O\left(\log n\cdot(\log\log n)^2\cdot\log\log\log n\right),$$

and prove the tighter bound in the full version [1].

Our result may be applied to improve the running time of the Spielman and Teng [17] solver for sparse symmetric diagonally dominant linear systems.

1.1 Techniques

We extend the star-decomposition technique of Elkin *et. al.*[11]. A star-decomposition of a graph is a partition of the vertices into clusters that are connected into a star: a central cluster is connected to every other cluster by a single edge. As in [11] given a subgraph over a cluster X, the central cluster X_0 is formed by cutting a ball with radius r_0 around a center x_0 and the remaining clusters $X_1, X_2, \ldots$, which are called cones, are formed iteratively. Let $Y_j = X\setminus\bigcup_{0\le k\le j}X_k$. The cone X_j is created by choosing an edge (y_j,x_j) such that $y_j\in X_0, x_j\in Y_{j-1}$ and defining X_j as the cone with radius r_j around x_j from the cluster Y_{j-1}, as all the points whose distance to x_0 going through the edge (x_j,y_j) does not increase too much relatively to the shortest path distance, formally $X_j = \{x\in Y_{j-1}\mid d_X(x_0,y_j)+d_X(y_j,x_j)+d_{Y_{j-1}}(x_j,x)-d_X(x_0,x)\le r_j\}$. Let $\mathrm{rad}_{x_0}(X) = \max_{x\in X}d(x_0,x)$, then typically the radius of the central ball is chosen so that $r_0\approx\mathrm{rad}_{x_0}(X)/c$ for a constant c. An important parameter of a star-decomposition is the radius of the cone. We say that the star-decomposition has parameter ϵ if for any $j\ge 1$, the radius r_j of the cone X_j is at most $\epsilon\cdot\mathrm{rad}_{x_0}(X)$.

Applying star-decompositions in a recursive manner induces a spanning tree T. For a point u denote by $X^{(i)}$ the cluster that contains u in the ith recursive invocation of the hierarchical star-decomposition algorithm.

The $O(\log^2 n\log\log n)$ bound of [11] is obtained by choosing $\epsilon\approx 1/\log n$ and showing:

1. $O(1)$ *radius stretch.* For any cluster X induced by the recursive invocation of the hierarchical star-decomposition algorithm, and any $z \in X$, $d_T(x_0, z) = O(\mathrm{rad}_{x_0}(X))$.
2. $O((\log n \cdot \log\log n)/\epsilon)$ *decomposition stretch.* For any edge (u, v),
$\sum_i \Pr[(u,v)$ is separated when star-decomposing $X^{(i)}] \cdot \mathrm{diam}(X^{(i)}) = O(\log n \log\log n)/\epsilon$.

Combining these two properties yields their result, noticing that if the end points of an edge (u, v) fall into different clusters in the partitioning of $X^{(i)}$ then $d_T(u, v)$ can be bounded by $d_T(u, x_0) + d_T(v, x_0) = O(\mathrm{diam}(X^{(i)}))$.

Good radius stretch is obtained by observing that in each recursive application of the star partition the radius of a cluster is stretched by at most $1 + 1/\log n$, and since there are $O(\log n)$ scales the total radius stretch is a constant. Good decomposition stretch is obtained by using a version of the decomposition of [5, 9].

Better radius stretch. In our scheme we perform a star-decomposition with a parameter $\epsilon \approx 1/\log\log n$, this significantly improves the decomposition stretch, by a factor of $\approx \log n/\log\log n$. A naive attempt to bound the radius stretch, by $1+1/\log\log n$ in each scale, will result in super logarithmic radius stretch over all scales.

We introduce a new approach to bound the radius stretch. We arrange all the points of X in a queue $Q = (z_1, z_2, \dots, z_n)$, and bound the distance $d_T(x_0, z_i)$ as a function of i by building "highways" – *low stretch paths*. Roughly speaking, the smaller value of i means the harder we try to give a better bound on $d_T(x_0, z_i)$. Therefore we try hardest for the first point z_1, and indeed by choosing the first portal edge (y_1, x_1) on a shortest path to z_1 and keeping z_1, y_1 in the head of the recursive queues we obtain a "highway" from x_0 to z_1, *i.e.* preserving the original distance. Surprisingly, this small change is enough to give a good bound on $d_T(x_0, z_i)$ for all $i > 1$, and we obtain $d_T(x_0, z_i) = O(\log\log i)\mathrm{rad}_{x_0}(X)$. The intuition is that since every cluster contains less points, z_i advances in the recursive queues, and when it becomes the first we get a "highway" to it. For this intuition to work one must delicately define the ordering of the queues $Q_0, \dots, Q_m$ for the clusters $X_0, \dots, X_m$ created by the star partition algorithm. Specifically, we obtain

1. $O(\log\log n)$ *radius stretch.* For any cluster X, and any $z \in X$, $d_T(x_0, z) = O(\log\log n)\mathrm{rad}_{x_0}(X)$.

Better decomposition stretch. A relaxation of the spanning tree problem suggested by Bartal [4] is to consider a distribution of dominating tree metrics (in fact of ultrametrics) that do not necessarily span the graph. This relaxation has proven applicable for approximation algorithms, online problems and has contributed to recent solutions for the spanning tree problem (*i.e.* [11]). Initially $O(\log^2 n)$ approximation was obtained in [4] based on the truncated exponential distribution approach of [15]. This bounded was subsequently improved to $O(\log n \log\log n)$ in [5] and [9]. Finally an optimal $O(\log n)$ approximation was obtained by [13] based on the cutting scheme of [8]. Subsequently an $O(\log n)$ bound was also obtained using a truncated exponential distribution approach [6, 2].

However, all previous schemes that obtained the optimal $O(\log n)$ bound for the metric problem were insufficient for the spanning tree problem. Given a graph $G = (X, E)$, a sequence $x_1, x_2, \dots$ of cluster centers and a sequence $r_1, r_2, \dots$ of radiuses we can define a weak diameter decomposition by defining $W_i = B_X(x_i, r_i) \setminus \bigcup_{j<i} W_j$. We can define a strong diameter decomposition by defining $C_i = B_{X \setminus \bigcup_{j<i} C_j}(x_i, r_i)$. Observe that in a strong diameter decomposition, for any nonempty cluster C_i, we have that $x_i \in C_i$ and C_i is a connected component of G, this may not be the case for weak diameter decompositions. Indeed the techniques of [13, 6, 2] provide a weak diameter decomposition. It was not clear how to extend these results to strong diameter decompositions that are necessary for star-decompositions. We show how to obtain a strong diameter hierarchical decomposition theorem that obtains an optimal bound in the following sense:

2. $O(\log n \log(1/\epsilon)/\epsilon)$ *decomposition stretch.* For any edge (u, v),
$\sum_i \Pr[(u,v)$ is separated when star-decomposing $X^{(i)}] \cdot \mathrm{diam}(X^{(i)}) = O(\log n \log(1/\epsilon)/\epsilon)$.

As in [6, 2], our decomposition is based on the truncated exponential distribution with a parameter depending on the local growth rate of the space. The main technical difficulty arises since the space *changes* after each cluster is cut (the metric is derived from a graph, and some nodes and edges are removed at every cut). The idea is to define the local growth rate with respect to the current metric, and to show two things: that the expected sum of all growth rates (which are random variables) over all the scales telescopes to n, and that the probability to be cut is appropriately bounded in each scale. Dealing with the randomly changing graph raises some additional subtleties in the proof. Our strong diameter hierarchical decomposition theorem may be of independent interest.

1.2 Applications

One of the main applications of low stretch spanning trees is solving *sparse symmetric diagonally dominant linear systems of equations*. This approach was suggested by Boman and Hendrickson [7] and later improved by Spielman and Teng [17]. Spielman and Teng showed an algorithm that for such an n-by-n matrix A with m non-zero entries and an n-dimensional vector b, if $\epsilon > 0$ is the precision of the solution then the algorithm finds x' such that $\|x - x'\|_A \leq \epsilon$ where $Ax = b$, and the running time is $O\Big(m\Big(\log^{O(1)} m + \log(1/\epsilon)\Big) + n \cdot \text{avg} - \text{stretch}(n)\log(1/\epsilon)\Big)$. Improving the bound requires improvement of the second element, and we improve it by roughly an additional $O(\log\log n)$ factor over [11]. Actually, if the running time of our construction is reduced, we can obtain an $O(\log n)$ improvement. For planar graphs we obtain $O(n \cdot \log^2 n)$.

The *minimum communication cost spanning tree* problem introduced in [14], in which one is given a weighted graph $G = (V, E, w)$ and a matrix $A = a_{xy} \mid x, y \in V$, the objective is to find a spanning tree minimizing $c(T) = \sum_{x,y \in V} a_{xy} \cdot d_T(x, y)$. [16] showed an $O(2^{\sqrt{\log n \cdot \log\log n}})$ approximation ratio based on [3], and [11] improved to $O(\log^2 n \cdot \log\log n)$. Our results can be used to obtain $O(\log n \cdot \log\log n (\log\log\log n)^3)$ approximation ratio.

See [11] for details about more applications.

2 Highways

Let $G = (V, E)$ be a finite graph. For any $X \subseteq V$ let $d_X : X^2 \to \mathbb{R}^+$ be the shortest path metric induced by the subgraph on X. Let $\text{diam}(X) = \max_{y,z \in X}\{d_X(y, z)\}$. For $x \in X$ let $\text{rad}_x(X) = \max_{y \in X} d_X(x, y)$, we omit the subscript when clear from context (note that $\text{diam}(X)/2 \leq \text{rad}(X) \leq \text{diam}(X)$). For any $x \in X$ and $r \geq 0$ let $B_{X,d}(x, r) = \{y \in X \mid d_X(x, y) \leq r\}$. Let $c = 2^{16}$ be a constant. We use the uppercase letter Q to denote a *queue*, a sequence of points. Given a point x not in the queue we say that we enqueue x into Q meaning that we add x as the last element of the sequence and given a queue Q, the dequeue operation removes and returns the first element of the sequence.

In this paper we assume that the graph G is unweighted. The extension for weighted graphs appears in the full version [1]. It is standard and similar to the techniques of [11].

Definition 1 (cone metric[3])**.** *Given a graph $G = (V, E)$, subsets $Y \subset X \subseteq V$, points $x \in X \setminus Y$, $y \in Y$ define the* cone-metric *$\rho = \rho(X, Y, x, y) : Y^2 \to \mathbb{R}^+$ as $\rho(u, v) = |(d_X(x, u) - d_Y(y, u)) - (d_X(x, v) - d_Y(y, v))|$.*

Note that a ball $B_{Y,\rho}(y, r)$ in the cone-metric $\rho = \rho(X, Y, x, y)$ is the set of all points $z \in Y$ such that $d_X(x, y) + d_Y(y, z) - d_X(x, z) \leq r$.

Hierarchical-Star-Partition algorithm. See Figure 1 for the algorithm. Given an unweighted graph $G = (V, E)$, create a spanning tree $T = (V, E')$ by choosing some $x_0 \in V$, letting Q be an arbitrary ordering of $V \setminus \{x_0\}$ and calling: `hierarchical-star-partition`(V, x_0, Q).

$T =$ `hierarchical-star-partition`(X, x_0, Q):

1. If $\text{rad}_{x_0}(X) \leq 16c$ return $\text{BFS}(X)$.
2. $(X_0, \ldots, X_m, (y_1, x_1), \ldots, (y_m, x_m), Q_0, Q_1, \ldots, Q_m) =$ `star-partition`(X, x_0, Q);
3. For each $i \in [0, \ldots, m]$:
4. $T_i =$ `hierarchical-star-partition`(X_i, x_i, Q_i);
5. Let T be the tree formed by connecting T_0 with T_i using edge (y_i, x_i) for each $i \in [1, \ldots, m]$;

Figure 1. `hierarchical-star-partition` **algorithm**

Star-Partition algorithm. See Figure 2 for our `star-partition` algorithm. We highlight the main differences of our algorithm from that of [11]. In addition to X, x_0 it receives as input an ordering of the points in X, implemented as a queue data structure and denoted by Q. In addition to returning a star decomposition $X_0, X_1, \ldots, X_m$ it returns for each $0 \leq j \leq m$ an ordering of the points in X_j, implemented as a queue data structure and denoted by Q_j.

Given a star decomposition $X_0, X_1, \ldots, X_m$ we create the queue Q_j for $j > 0$ simply as the restriction of Q on $X_j \setminus \{x_j\}$. The queue Q_0 is created by first adding either z_1 or the portal y_1 which is chosen on a shortest path to z_1, thus making sure the distance from x_0 to z_1 is preserved in the recursion. Then interleaving three different queues $Q_0^{(\text{ball})}, Q_0^{(\text{fat})}, Q_0^{(\text{reg})}$.

- $Q_0^{(\text{ball})}$ is the restriction of Q on X_0.
- $Q_0^{(\text{reg})}$ is a queue of portals y_j ordered by the minimal point of Q that their cones X_j contains.
- $Q_0^{(\text{fat})}$ is a queue of portals y_j that lead to cones that contain "many" points relative to the ordering Q of the points in X_j.

The exact way these three queues are created is detailed in Line 5 of Figure 2.

2.1 Bounding the radius stretch

In this part we show that the radius stretch induced by the `hierarchical-star-partition` algorithm is at most $O(\log\log n)$.

The following two claims imply that the `star-partition` algorithm on a cluster X induces a partition on X and that radial distances are stretched by a most $1 + \epsilon$. These claims are essentially proven in [11].

[3] In fact, the cone-metric is a pseudo-metric.

$(X_0,\dots,X_m,(y_1,x_1),\dots,(y_m,x_m),Q_0,Q_1,\dots,Q_m) = \texttt{star-partition}(X,x_0,Q)$:

1. Let $j=2$; Denote the (ordered) elements of Q by $Q=(z_1,z_2,\dots,z_k)$; Let $\epsilon=\epsilon(X)\in(0,\frac{1}{170c}]$;
2. Creating the ball X_0:
   (a) Choose r_0 uniformly at random from the interval $[1/(16c),1/(8c)]$;
   (b) Let $X_0=B(x_0,r_0\cdot \mathrm{rad}_{x_0}(X))$; Let $Y_0=X\setminus X_0$;
3. Creating the first cone X_1:
   (a) If $z_1\in Y_0$ let $z=z_1$ otherwise let $z\in Y_0$ be an arbitrary point. Let (y_1,x_1) be an edge such that $y_1\in X_0$, $x_1\in Y_0$ and $d_X(x_0,z)=d_X(x_0,y_1)+d_X(y_1,x_1)+d_{Y_0}(x_1,z)$ (*i.e.* an edge on a shortest path from x_0 to z);
   (b) Let $\rho=\rho(X,Y_0,x_0,x_1)$ be the cone-metric;
   (c) Choose r_1 uniformly at random from the interval $[\epsilon/4,\epsilon/2]$;
   (d) Let $X_1=B_{(Y_0,\rho)}(x_1,r_1\cdot\mathrm{rad}_{x_0}(X))$; Let $Y_1=Y_0\setminus X_1$;
4. Creating the remaining cones $X_2,\dots,X_m$:
   (a) While $Y_{j-1}\neq\emptyset$:
      i. Let $(x_j,y_j,r_j) = \texttt{cone-cut}(X,x_0,X_0,Y_{j-1},\epsilon)$; (has the property that $r_j\le\epsilon/2$)
      ii. Let $\rho=\rho(Y_{j-1}\cup X_0,Y_{j-1},x_0,x_j)$;
      iii. Let $X_j=B_{(Y_{j-1},\rho)}(x_j,r_j\cdot\mathrm{rad}_{x_0}(X))$; $Y_j=Y_{j-1}\setminus X_j$;
      iv. Let $j=j+1$;
5. Creating the queues $Q_0^{(\text{ball})},Q_0^{(\text{fat})},Q_0^{(\text{reg})}$, $Q_1,\dots,Q_m$:
   (a) For $i=1,\dots,|X|-1$:
      i. If $z_i\in X_0$ then enqueue z_i into $Q_0^{(\text{ball})}$;
      ii. Otherwise let $\ell\ge 1$ be such that $z_i\in X_\ell$:
         - If $z_i\neq x_\ell$ then enqueue z_i into Q_ℓ.
         - If $y_\ell\notin Q_0^{(\text{reg})}$ then enqueue y_ℓ into $Q_0^{(\text{reg})}$.
         - If $|X_\ell\cap\{z_1,\dots,z_i\}|>\sqrt{i}$ and $y_\ell\notin Q_0^{(\text{fat})}$ then enqueue y_ℓ into $Q_0^{(\text{fat})}$.
6. Creating the queue Q_0:
   (a) Denote $Q_0^{(\text{ball})}=z_1^1,\dots,z_{m_1}^1$, $Q_0^{(\text{fat})}=z_1^2,\dots,z_{m_2}^2$, $Q_0^{(\text{reg})}=z_1^3,\dots,z_{m_3}^3$.
   (b) Create Q_0 by interleaving the three queues $Q_0^{(\text{ball})},Q_0^{(\text{fat})},Q_0^{(\text{reg})}$ such that:
      - If $z_1\in X_0$ then z_1 is the first element of Q_0. Otherwise y_1 is the first element of Q_0.
      - For any $x\in X$, $\ell\in\{1,2,3\}$, $1\le i\le n$ if $x=z_i^\ell$ then x is in the first $3i$ elements of Q_0.

Figure 2. `star-partition` **algorithm**

Claim 1. *For any graph X, $x_0\in X$, $j>0$ let $Y_{j-1}\subseteq X$ be the unassigned points of X after creating j clusters $X_0,\dots,X_{j-1}$ using the* `star-partition` *algorithm, then for any $z\in Y_{j-1}$ all the shortest paths from z to x_0 are fully contained in $Y_{j-1}\cup X_0$, in particular*

$$d_{Y_{j-1}\cup X_0}(x_0,z)=d_X(x_0,z).$$

Claim 2. *Let $X_0,\dots,X_m$ be the clusters created by the* `star-partition` *algorithm on (X,x_0,Q), then for any $1\le j\le m$*

$$\mathrm{rad}_{x_0}(X_0)+d(y_j,x_j)+\mathrm{rad}_{x_j}(X_j)\le(1+\epsilon)\mathrm{rad}_{x_0}(X),$$

Corollary 3. *For any $0\le j\le m$, $\mathrm{rad}_{x_j}(X_j)<(1-\frac{1}{20c})\mathrm{rad}_{x_0}(X)$*

Lemma 4. *Let $X\subseteq V$ be a connected component of $G(V,E)$. Let $x_0\in X$ and $Q=(z_1,\dots,z_{|X|-1})$ be any ordering of $X\setminus\{x_0\}$. Let T be any spanning tree of G returned by the algorithm* `hierarchical-star-partition`(X,x_0,Q) *with parameter $\epsilon=\epsilon(X)=\frac{1}{170c\log\log(|X|)}$, then*

$$d_T(x_0,z_i)\le\begin{cases} d_X(x_0,z_i) & i=1\\ i\cdot\mathrm{rad}_{x_0}(X) & 1<i<c\\ c\cdot\log\log i\cdot\mathrm{rad}_{x_0}(X) & \textit{otherwise}\end{cases}$$

(where $c=2^{16}$)

Proof. The proof is by induction on the radius of X. In the base case when $\mathrm{rad}_{x_0}(X)\le 16c$ create a breadth first tree centered in x_0, and since in such a tree for every $z\in X$, $d_X(x_0,z)=d_T(x_0,z)$ the claim holds. Now we turn to the inductive step. Note that Corollary 3 guarantees that for all $j=0,\dots,m$ we have $0\le\mathrm{rad}_{x_j}(X_j)<\mathrm{rad}_{x_0}(X)$.

The main idea of the proof is to consider a single application of the star-partition algorithm, partitioning X into $X_0,X_1,\dots,X_m$. Assuming that $z_i\in X_j$ the path between x_0 to z_i will be the path going through the edge (y_j,x_j). Then use the induction hypothesis on the sub-path x_0,y_j in X_0 and the sub path x_j,z_i in X_j. Since by Claim 2 the radius may increase by a factor of at most $1+\epsilon$, we need to "gain" in one of the two sub paths. This "gain" will occur since our construction guarantees that either the position of z_i in the queue of X_j will improve or the position of y_j in X_0 will improve, thus the induction hypothesis will give the required bounds.

There are three main cases to consider, when $i=1$, $i<c$ and $i\ge c$. The case $i=1$ is simple. The case $1<i<c$ subdivides into three more cases:

1. The first case is $z_i\in X_0$. This case is relatively straightforward.
2. The second case is that the first i points of the queue are all in X_1. Here we gain in the central ball because the portal y_1 leading to X_1 will be the first element in Q_0.

3. The remaining case is that not all of the first i points are in X_1, then there are at most $i-1$ points in the cone X_j among $z_1,\dots,z_i$, so by the construction of Q_j, we gain just enough in the cone (because the bound that needs to be shown is weak - linear in i) and $Q_0^{(\text{reg})}$ guarantees that we do not lose too much in the central ball.

The interesting case is when $i \geq c$, this last case also subdivides into three more cases:

1. One first is that $z_i \in X_0$. Again. this case is relatively straightforward and uses the construction of $Q_0^{(\text{ball})}$.
2. The second case is that $z_i \in X_j$ and X_j is a "thin" cone - contains less than $\sqrt{i}$ of the first i points. Here we gain in the cone because the position of z_i in Q_j is at most $\sqrt{i}$, and $Q_0^{(\text{reg})}$ guarantees that we do not lose too much in the central ball.
3. The third case is that $z_i \in X_j$ and X_j is a "fat" cone - contains more than $\sqrt{i}$ of the first i points. Here we gain in the central ball, using the construction of $Q_0^{(\text{fat})}$ and Claim 5 to show that the portal y_j leading to the cone is in position $\leq i^{9/10}$ in Q_0.

We continue with the formal proof of the lemma, according to the three main cases. Let $\Delta = \mathrm{rad}_{x_0}(X)$ and for all $0 \leq j \leq m$, $\Delta_j = \mathrm{rad}_{x_j}(X_j)$.

Case 1: In this case $i = 1$. Note that $z_1 \in X_0 \cup X_1$. If $z_1 \in X_0$ then by the construction z_1 is going to be the first in Q_0 therefore by the induction hypothesis on X_0 it follows that $d_T(x_0, z_1) \leq d_X(x_0, z_1)$. If on the other hand $z_1 \in X_1$, then again from the construction the point y_1, which was chosen such that y_1, x_1 are on a shortest path from x_0 to z_1, will be the first in Q_0, and z_1 will be the first in X_1, so by induction $d_T(x_0,z_1) = d_T(x_0,y_1) + d_T(y_1,x_1) + d_T(x_1,z_1) \leq d_X(x_0,y_1) + d_X(y_1,x_1) + d_X(x_1,z_1) = d_X(x_0,z_1)$.

Case 2: The second case to consider is when $1 < i < c$.

1. First assume that $z_i \in X_0$. Then z_i will be at most i in the ordering of $Q_0^{(\text{ball})}$ and hence at most $3i$ in the ordering of Q_0. By the induction hypothesis on X_0 : $d_T(x_0, z_i) \leq c\log\log(3i)\cdot\Delta_0 \leq i\cdot\Delta$, using that $\Delta_0 \leq \Delta/(8c)$, and that $\log\log(3i) \leq 2i$.
2. Now assume that $\{z_1,\dots,z_i\} \subseteq X_1$. As y_1 is the first in Q_0, by the induction hypothesis on X_0 and X_1 we have that $d_T(x_0,y_1) \leq d_X(x_0,y_1) \leq \Delta_0$ and $d_T(x_1,z_i) \leq i\cdot\Delta_1$, so

$$
\begin{aligned}
& d_T(x_0,z_i) \\
& \leq d_T(x_0,y_1) + d_T(y_1,x_1) + d_T(x_1,z_i) \\
& \leq \Delta_0 + i\cdot\Delta_1 + d_X(y_1,x_1) \\
& \leq i(\Delta_0+1+\Delta_1) - (i-1)\Delta_0 \\
& \leq i(1+\epsilon)\Delta - (i-1)\Delta/(16c) \\
& \leq i\cdot\Delta + i\cdot\Delta/(170c) - i\cdot\Delta/(32c) \\
& \leq i\cdot\Delta.
\end{aligned}
$$

In the fourth inequality using Claim 2 and that $\Delta_0 \geq \Delta/(16c)$ (note that by the stop condition of hierarchical-star-partition $\Delta \geq 16c$, so $\Delta_0 \geq 1$) and in the fifth that $i-1 \geq i/2$.
3. Now assume that $z_i \in X_j$ where not all of $z_1,\dots,z_i$ are in X_j (note that $z_1 \in X_0 \cup X_1$, therefore there is no case for $\{z_1,\dots,z_i\} \subseteq X_j$ where $j > 1$). First note that z_i must be at most the $i-1$ element in Q_j. By the insert sequence to $Q_0^{(\text{reg})}$ we have that y_j is at most the $3i$ element in Q_0. Using the induction hypothesis on X_0 and X_j we get that

$$
\begin{aligned}
& d_T(x_0,z_i) \\
& \leq d_T(x_0,y_j) + d_T(y_j,x_j) + d_T(x_j,z_i) \\
& \leq c\log\log(3i)\cdot\Delta_0 + (i-1)\cdot\Delta_j + d_X(y_j,x_j) \\
& \leq (i-1)(\Delta_0 + 1 + \Delta_j) + 5c\cdot\Delta_0 \\
& \leq (i-1)(1+\epsilon)\Delta + 5c\cdot\Delta/(8c) \\
& \leq i\cdot\Delta - \Delta + (i-1)\cdot\Delta/(170c) + 5\Delta/8 \\
& \leq i\cdot\Delta.
\end{aligned}
$$

The third inequality follows since $\log\log(3i) \leq \log\log(3c) \leq 5$. The fourth using Claim 2 and that $\Delta_0 \leq \Delta/(8c)$.

Case 3: In the third case $i \geq c$.

1. First assume that $z_i \in X_0$. Then z_i will be at most i in the ordering of $Q_0^{(\text{ball})}$, hence at most $3i$ in the ordering of Q_0. By the induction hypothesis on X_0 we get that $d_T(x_0,z_i) \leq c\log\log(3i)\cdot\Delta_0 \leq 2c\log\log i\cdot\Delta_0 \leq c\log\log i\cdot\Delta$, using that for $i \geq c$, $3i < i^2$.
2. Next assume that $z_i \in X_j$ such that $|X_j \cap \{z_1,\dots,z_i\}| \leq \sqrt{i}$, then z_i will be at most the $\sqrt{i}$ in Q_j, and y_j will be at most the i-th in $Q_0^{(\text{reg})}$ and hence at most $3i$ in the ordering of Q_0. By the induction hy-

pothesis on X_0 and X_j:

$$\begin{aligned}
&d_T(x_0, z_i)\\
&\le \quad d_T(x_0, y_j) + d_T(y_j, x_j) + d_T(x_j, z_i)\\
&\le \quad c\log\log(3i)\cdot\Delta_0 + c\log\log(\sqrt{i})\cdot\Delta_j + 1\\
&\le \quad c(\log\log i + 1)\cdot\Delta_0 + c(\log\log i - 1)\cdot\Delta_j + 1\\
&\le \quad c(\log\log i - 1)\,(\Delta_0 + 1 + \Delta_j) + 2c\cdot\Delta_0\\
&\le \quad c(\log\log i - 1)(1+\epsilon)\Delta + \Delta/4\\
&\le \quad c\log\log i\cdot\Delta + c\log\log i\cdot\epsilon\Delta - c\Delta + \Delta/4\\
&\le \quad c\log\log i\cdot\Delta,
\end{aligned}$$

the fifth inequality using Claim 2 and that $\Delta_0 \le \Delta/(8c)$, the sixth that $\epsilon \le 1/(170c\log\log i)$.

3. The last subcase is where $z_i \in X_j$ such that $|X_j \cap \{z_1, \ldots, z_i\}| > \sqrt{i}$, then z_i will be at most the i in Q_j and by Claim 5 y_j will be at most the $i^{9/10}$ in Q_0. Now by the induction hypothesis, for $t \ge 2$

$$\begin{aligned}
&d_T(x_0, z_i)\\
&\le \quad d_T(x_0, y_j) + d_X(y_j, x_j) + d_T(x_j, z_i)\\
&\le \quad c\log\log i^{9/10}\cdot\Delta_0 + c\log\log i\cdot\Delta_j + 1\\
&\le \quad c\log\log i(\Delta_0 + 1 + \Delta_j) + c\log(9/10)\cdot\Delta_0\\
&\le \quad c\log\log i\cdot\Delta + \epsilon\cdot c\log\log i\cdot\Delta - c\cdot\Delta_0/10\\
&\le \quad c\log\log i\cdot\Delta + \Delta/170 - \Delta/160\\
&\le \quad c\log\log i\cdot\Delta,
\end{aligned}$$

the fourth inequality using Claim 2 and the fifth that $\Delta_0 \ge \Delta/(16c)$ and $\epsilon \le 1/(170c\log\log i)$.

□

The following claim shows that a portal y_j leading to a point z_i that belongs to a "fat" cone will be located in an improved position in the queue of the central ball Q_0.

Claim 5. *For any $i \ge 2^{16}$, if $z_i \in X_j$ such that $|X_j \cap \{z_1, \ldots, z_i\}| > \sqrt{i}$ then y_j will be at position at most $i^{9/10}$ in Q_0.*

Proof. We will show that y_j will be in the first $(3/2)i^{2/3}+1$ elements of $Q_0^{(\text{fat})}$. Since $i \ge 2^{16}$ it follows that y_j will be in the first $3\cdot((3/2)i^{2/3}+1) < i^{9/10}$ elements of Q_0.

Let $y_{i_1}, \ldots, y_{i_s}$ with $i_1 < i_2 < \cdots < i_s$ be a set of s points that were inserted into $Q_0^{(\text{fat})}$ before considering the point z_i, we need to show that $s \le (3/2)i^{2/3}$. Let $z_{i'_1}, \ldots, z_{i'_s}$ be the set of points in Q such that y_{i_k} was inserted because $z_{i'_k} \in X_{i_k}$ and X_{i_k} was a "fat" cone, *i.e.* $|X_{i_k} \cap \{z_1, \ldots, z_{i'_k}\}| \ge \sqrt{i'_k}$. Let $A_{i_k} = X_{i_k} \cap \{z_1, \ldots, z_{i'_k}\}$ denote the set that caused y_{i_k} to enter $Q_0^{(\text{fat})}$, and note that $|A_{i_k}| \ge \sqrt{i'_k} \ge \sqrt{k}$. For any $1 \le k < \ell \le s$ we have that $A_{i_k} \cap A_{i_\ell} = \emptyset$, since we do not insert a point y_{i_ℓ} that already appear in $Q_0^{(\text{fat})}$, which implies $X_{i_k} \cap X_{i_\ell} = \emptyset$. Note that all the sets A_{i_k} contain points from $z_1, \ldots, z_i$, so we have that $\sum_{k=1}^{s} |A_{i_k}| \le i$. Hence $\sum_{k=1}^{s} \sqrt{k} \le \sum_{k=1}^{s} |A_{i_k}| \le i$. We also bound the sum from below

$$\sum_{k=1}^{s} \sqrt{k} \ge \int_1^s \sqrt{x}dx = [(2/3)x^{3/2}]_1^s \ge (2/3)s^{3/2},$$

therefore $i \ge (2/3)s^{3/2}$ or $s \le (3/2)i^{2/3}$. □

2.2 Improving the radius stretch

The factor of $c\log\log i$ that was chosen as a bound on the radius increase in Lemma 4 was somewhat arbitrary. In fact we can replace it with almost any other monotone increasing function of i, the position in the queue. This will reduce the "gain" in the induction, therefore the parameter ϵ will have to be adjusted accordingly. Define $\log^{(0)} n = n$ and for integer $k \ge 1$, $\log^{(k)} n = \log\left(\log^{(k-1)} n\right)$. Specifically, we show in [1] that by setting a new parameter $t = (\log^* n)/2$, and letting $c = O(t)$ and $\epsilon = \frac{1}{170c\cdot\prod_{k=2}^{t}\log^{(k)} n}$ we get a radius stretch of $O(c^2)$ (the fact that c is no longer a constant and that the gain is so small somewhat complicates the proof of Lemma 4). Then the decomposition stretch becomes $O((c^2\log n\cdot\log\log\log n)/\epsilon)$, hence the final stretch is at most

$$O\left(\log^{(1)} n\cdot\log^{(2)} n\cdots\log^{(t)} n\cdot\left(\log^{(t)} n\right)^4\cdot\log^{(3)} n\right).$$

3 Strong Diameter Probabilistic Partitions

Consider a graph $G = (V, E)$, a connected cluster $X \subseteq V$, $x_0 \in X$ and let $\Delta = \text{rad}_{x_0}(X)$. Fix some edge $(u, v) \in E$. Let $X^{(i)} = X^{(i)}(u)$ be a random variable that indicates which cluster contains u in the i-th step of the hierarchical application of the star-partition algorithm[4]. In a similar manner let $x_0^{(i)}$ be the random variable indicating the center of the cluster $X^{(i)}$, and when $X^{(i)}$ is partitioned denote the central ball as $X_0^{(i)}$ and cones as $X_1^{(i)}, \ldots X_m^{(i)}$ where m is a random variable depending on $X^{(i)}$. Let $\mathcal{E}_j(X^{(i)}, u, v)$ be the event that $u, v \in X^{(i)}$ and in the star-partition of the cluster $X^{(i)}$ with center $x_0^{(i)}$ into $X_0^{(i)}, \ldots, X_m^{(i)}$, $u \in X_j^{(i)}$, $v \notin X_j^{(i)}$. Let $\mathcal{E}(X^{(i)}, u, v)$ be the event that $\exists\, 0 \le j \le m$ such that $\mathcal{E}_j(X^{(i)}, u, v)$. Some notation:

[4]We abuse notation and think of $X^{(i)}$ as a function to subsets of X (instead of $\mathbb{R}$). We also refer to $X^{(i)}$ as an event.

$(x, y, r) = \texttt{cone cut}(X, x_0, X_0, Y, \epsilon)$:

- Let $p \in Y$ be the point minimizing $\frac{|X|}{|B_{(Y,d_Y)}(z, \epsilon \cdot \Delta/16)|}$ over all $z \in Y$; Let χ denote that minimum;
- Let (y, x) be an edge such that $x \in Y$, $y \in X_0$ and $d_X(x_0, y) + d_X(y, x) + d_Y(x, p) = d_X(x_0, p)$ (*i.e.* y and x lie on some shortest path between x_0 and p);
- Choose $r \in [\epsilon/4, \epsilon/2]$ according to the following random process:
  - Divide the interval $[\epsilon/4, \epsilon/2]$ into $N = \lceil 2 \log \chi \rceil$ equal length intervals $S_1, \ldots, S_N$; Let $h = 1$;
  - LOOP: Toss a fair coin; If it turns out head and $h < N$ then let $h = h + 1$ and goto LOOP;
  - Choose r uniformly at random from the interval S_h.
- Return (x, y, r).

Figure 3. `cone-cut` **algorithm**

$\mathbb{E}_{X^{(i)}}[f(X^{(i)})]$ will stand for $\sum_{X'} \Pr[X^{(i)} = X'] f(X')$.

Let $\mathcal{T}$ be the support of the distribution over spanning trees induced by the hierarchical star partition algorithm. Let $\mathcal{T}^{(i)} \subseteq \mathcal{T}$ be the set of spanning trees for which event $\mathcal{E}(X^{(i)}, u, v)$ occurs.

$$
\begin{aligned}
&\mathbb{E}[d_T(u, v)] \\
&\quad \le \sum_{i \ge 1} \sum_{T \in \mathcal{T}^{(i)}} \Pr[T] \cdot d_T(u, v) \\
&\quad \le \sum_{i \ge 1} \mathbb{E}_{X^{(i)}} \left[\Pr[\mathcal{E}(X^{(i)}, u, v)] \max_{T \in \mathcal{T}^{(i)}} \{ d_T(u, v) \} \right] \\
&\quad \le O(\log \log n) \sum_{i \ge 1} \mathbb{E}_{X^{(i)}} \Big[\Pr[\mathcal{E}(X^{(i)}, u, v)] \mathrm{rad}_{x_0^{(i)}}(X^{(i)}) \Big]
\end{aligned}
$$

The last inequality holds since for any $T \in \mathcal{T}^{(i)}$, $d_T(u, v) \le d_T(u, x_0^{(i)}) + d_T(x_0, v) \le 2\mathrm{rad}_{x_0^{(i)}}(T)$ and using Lemma 4 we get that $\mathrm{rad}_{x_0^{(i)}}(T) \le O(\log \log n \cdot \mathrm{rad}_{x_0^{(i)}}(X^{(i)}))$.

In what follows we bound $\mathbb{E}_{X^{(i)}} \left[\Pr[\mathcal{E}(X^{(i)}, u, v)] \cdot \mathrm{rad}_{x_0^{(i)}}(X^{(i)}) \right]$. Let $\epsilon = \frac{1}{170c \cdot \log \log |X|}$ and $k = 20c(\ln(1/\epsilon) + 5)$. The main lemma to prove is the following

Lemma 6. *There is a universal constant C such that for any graph $G = (V, E)$, any edge $(u, v) \in E$ and any connected cluster $X^{(i)} \subseteq V$ we have that*

$$
\begin{aligned}
&\mathbb{E}_{X^{(i)}} \left[\Pr[\mathcal{E}(X^{(i)}, u, v)] \cdot \mathrm{rad}_{x_0^{(i)}}(X^{(i)}) \right] \\
&\quad \le C d(u, v)/\epsilon \left(\mathbb{E}_{X^{(i)}}[\log |X^{(i)}|] - \mathbb{E}_{X^{(i+k)}}[\log |X^{(i+k)}|] \right)
\end{aligned}
$$

Once this lemma is proved, a telescopic sum argument yields that

$$
\begin{aligned}
&\mathbb{E}[d_T(u, v)] \\
&\quad \le O(\log \log n) \sum_{i \ge 1} \mathbb{E}_{X^{(i)}} \Big[\Pr[\mathcal{E}(X^{(i)}, u, v)] \mathrm{rad}_{x_0}(X^{(i)}) \Big] \\
&\quad \le O(\log \log n) \cdot d(u, v)/\epsilon \sum_{i=1}^{k} \mathbb{E}_{X^{(i)}}[\log |X^{(i)}|] \\
&\quad \le O(\log n \cdot \log \log n) \cdot d(u, v) \cdot \log(1/\epsilon)/\epsilon \\
&\quad = O(\log n \cdot (\log \log n)^2 \cdot \log \log \log n) \cdot d(u, v)
\end{aligned}
$$

As we stated in the introduction, the algorithm of Figure 3 and proof of Lemma 6 are based on the truncated exponential distribution approach of [6, 2]. The main technical difficulty arises since the space *changes* after each cluster is cut. Dealing with the randomly changing graph raises some additional subtleties in the proof.

We begin with some definitions and an informal description of the algorithm and the proof idea. Fix the edge $(u, v) \in E$, a scale i and $X = X^{(i)}$. Let $Y \subseteq X$ be a random variable indicating that there exists $0 < j \le m$ such that $Y = Y_{j-1}$ in the star partition of X. Define the local growth rate around $x \in Y$ with respect to Y as

$$
\chi(X, Y, x) = \frac{|X|}{|B_{Y,d_Y}(x, \epsilon\Delta/16)|}
$$

The algorithm for the partition is as follows: Choose a radius for the central ball around x_0 from a uniform distribution in a range of size $\approx \Delta/c$. The center x_1 is chosen on a shortest path to z_1, the first point in the queue, and then the radius for the cone is again sampled from a uniform distribution in a range of size $\approx \epsilon\Delta$. For $j > 1$ the jth center x_j is chosen on a shortest path to the point $p_j \in Y_{j-1}$ minimizing $\chi_j = \chi(X, Y_{j-1}, p_j)$, and then the radius of the cone is chosen from a truncated exponential distribution, with parameter χ_j.

Denote the event that $Y = Y_{j-1}$ and $u \in X_j$ as $\mathcal{Z}_j(X, Y, u)$, and let $\mathcal{Z}(X, Y, u)$ be the event that $\exists\, 0 \le j < m$ such that $\mathcal{Z}_j(X, Y, u)$. Note that fixing Y_{j-1} determines deterministically p_j and therefore also x_j and χ_j. Similarly let $\mathcal{Z}_j(X, Y)$ be the event that $Y = Y_{j-1}$ and $\mathcal{Z}(X, Y)$ the event that $\exists\, 0 \le j < m$ such that $\mathcal{Z}_j(X, Y)$. Let $N(j)$ be the random variable that is the number of partitions $S_1, \ldots, S_{N(j)}$ of the interval $[\epsilon/4, \epsilon/2]$ for the jth cone. Let $0 \le h(j) \le N(j)$ be the random variable that is the index of the interval $S_{h(j)}$ from which the radius r_j is uniformly chosen for X_j. Some more notation:

$\mathbb{E}_{Y \subseteq X}[f(Y)]$ will stand for $\sum_{Y \subseteq X} \Pr[\mathcal{Z}(X, Y)] \cdot f(Y)$ (we write $\mathbb{E}_Y$ when X is implicit).

$\mathbb{E}_{Y \subseteq X, j}[f(Y)]$ will stand for $\sum_{Y \subseteq X} \Pr[\mathcal{Z}_j(X, Y)] \cdot f(Y)$ (we write $\mathbb{E}_{Y,j}$ when X is implicit).

$\mathbb{E}_{Y \subseteq X, u}[f(Y)]$ will stand for $\sum_{Y \subseteq X} \Pr[\mathcal{Z}(X,Y,u)] \cdot f(Y)$ (we write $\mathbb{E}_{Y,u}$ when X is implicit).

In all events we remove the parameter X when clear from context. We divide the event $\mathcal{E}(u,v)$ into three cases (by symmetry we can define all these events with respect to u).

- The first is the event that u falls into one of the first two clusters (the central ball X_0 or the first cone X_1). This event is denoted by $\mathcal{G}(X,u)$.
- The second is the event that u is contained in cluster X_j for some $j > 1$, such that the cone distance between u and the center x_j is in the last interval *i.e.* that $\rho(x_j,u)/\Delta \in S_{N(j)}$. This event is denoted by $\mathcal{F}(X,u)$. We partition the event $\mathcal{F}(X,u)$ using the different values of j: For any $j > 1$ let $\mathcal{F}_j(X,u)$ be the event that $\rho(x_j,u)/\Delta \in S_{N(j)}$, and note that $\mathcal{F}(X,u)$ is simply that there exists $j > 1$ such that $\mathcal{F}_j(X,u)$ and also $u \in X_j$.
- The third is the completion of the first two events, that the cluster X_j containing u has $j > 1$ and $\rho(x_j,u)/\Delta \notin S_{N(j)}$.

The probability of the first event can be bounded simply by the inverse of the range from which the radius is drawn, so we obtain probability at most $\approx \frac{d(u,v)}{\epsilon\Delta}$.

For the second event we note that reaching the tail of the exponential distribution requires that $N-1$ fair coin tosses turned out head, which is bounded by $\approx \frac{1}{2^N} \approx \frac{1}{\chi_j^2}$, then since we choose uniformly from the last interval, the probability that we separate u,v is $\approx \frac{\log\chi_j \cdot d(u,v)}{\epsilon\Delta\chi_j^2} \leq \frac{d(u,v)}{\epsilon\Delta\chi_j}$. Since the parameter χ_j is a random variable which depends on the previous cone cuts, the proof becomes a bit more involved as we need to give a different bound for every possible $Y = Y_{j-1}$. We show that for every star-partition $\sum_{j>1} \chi_j^{-1} \leq 1$, hence this also holds in expectation and the second event probability is bounded by $\approx \frac{d(u,v)}{\epsilon\Delta}$. This is shown in Claim 7

Bounding the third event relies on the memoryless property of the exponential distribution. The major technical difficulty is that the bound we show depends on the parameter χ. Hence we can only show the bound given some subspace Y from which we cut the next cone. The bound on the probability obtained here is $\approx \frac{\log\chi \cdot d(u,v)}{\epsilon\Delta}$. This is shown in Claim 8.

The last step is to sum over all scales i, and use a telescopic sum argument on the expectation of the values of the $\log\chi$ showing that they sum to $O(\log(1/\epsilon)\cdot\log n)$. This is shown in the proof of Lemma 6.

Claim 7. *For any cluster $X \subseteq V$, edge $u,v \in X$, $(u,v) \in E$, we have* $\Pr[\mathcal{F}(u) \wedge \mathcal{E}(u,v)] \leq 48d(u,v)/(\epsilon\Delta)$.

Proof. Note that we can only bound the probability of event such as $\mathcal{E}_j(u,v)$ given that some $Y = Y_{j-1}$ is fixed *i.e.* that event $\mathcal{Z}_j(X,Y)$ occurred (because the parameters x_j and χ_j that govern the next cone creation are random variables depending on Y. So fix some $Y = Y_{j-1}$ and note that indeed p_j, x_j and $\chi_j = \chi(X,Y,p_j)$ are determined deterministically.

$$\begin{aligned}
&\Pr[\mathcal{F}(u) \wedge \mathcal{E}(u,v)] \\
&\quad = \Pr[\exists j > 1, \mathcal{F}_j(u) \wedge \mathcal{E}_j(u,v)] \\
&\quad \leq \sum_{j\geq 2} \Pr[\mathcal{E}_j(u,v) \mid \mathcal{F}_j(u)] \\
&\quad = \sum_{j\geq 2}\sum_{Y\subseteq X} \Pr[Z_j(Y)] \cdot \Pr[\mathcal{E}_j(u,v) \mid \mathcal{F}_j(u) \wedge \mathcal{Z}_j(Y)] \\
&\quad = \sum_{j\geq 2} \mathbb{E}_{Y,j}\left[\Pr[\mathcal{E}_j(u,v) \mid \mathcal{F}_j(u)]\right]
\end{aligned}$$

The first equation holds since the probability to be cut by a cluster whose radius is "large" is the probability that some cluster X_j with large radius separates u,v. The first inequality holds by the union bound and the second equation since for every event A and pairwise disjoint events $B_1,\ldots,B_\ell$ with $\sum_{i=1}^{\ell}\Pr[B_i] = 1$ it holds that $\Pr[A] = \sum_{i=1}^{\ell}\Pr[B_i]\cdot\Pr[A \mid B_i]$. Here the events B are $\mathcal{Z}_j(X,Y)$ which are disjoint for different subgraphs Y. Note that events $\mathcal{F}_j(u)$ and $\mathcal{Z}_j(X,Y)$ tell us nothing of the radius of the next cone X_j, therefore the probability of $\mathcal{E}_j(u,v)$ given the subspace Y_{j-1} and that $\rho(x_j,u)/\Delta \in S_{N(j)}$ (where $\rho = \rho(X, Y \cup X_0, d', x_0, x_j)$ is the cone metric), is the probability that $h(j) = N(j)$ (recall that the random variable $h(j)$ is the index of the interval $S_{h(j)}$ from which the radius is uniformly chosen for X_j) and that the uniform choice in the interval $S_{N(j)}$ hits the place that separates u,v. To bound the first one

$$\Pr[h(j) = N(j)] = 2^{-(N(j)-1)} \leq 2^{-2\log\chi_j+2} = 4/\chi^2,$$

and the probability of the second event is $\frac{d(u,v)}{\Delta|S_{N(j)}|}$. Note that $|S_{N(j)}| = \frac{\epsilon}{4\lceil 2\log\chi_j\rceil} \geq \frac{\epsilon}{8\log\chi_j+4} \geq \min\{1, \frac{1}{\log\chi_j}\}\frac{\epsilon}{12}$. These two events are independent, hence

$$\begin{aligned}
&\Pr[\mathcal{F}(u) \wedge \mathcal{E}(u,v)] \\
&\quad \leq \frac{48d(u,v)}{\epsilon\cdot\Delta}\sum_{j\geq 1}\mathbb{E}_{Y,j}\left[\max\left\{\frac{1}{\chi_j^2}, \frac{\log\chi_j}{\chi_j^2}\right\}\right] \\
&\quad \leq \frac{48d(u,v)}{\epsilon\cdot\Delta}\sum_{j\geq 1}\mathbb{E}_{Y,j}[\chi_j^{-1}]
\end{aligned}$$

For any $\bar{Y} = (\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_n) \subset X^n$ let $\mathcal{Z}(\bar{Y})$ be the event $\bigwedge_{1\leq j\leq n}\mathcal{Z}(\bar{Y}_j, j)$ (where $\bar{Y}_j$ is the jth component of $\bar{Y}$). Observe that for any j and $Y \subset X$ we have

$\Pr[\mathcal{Z}(Y,j)] = \sum_{\bar{Y}\subset X^n, \bar{Y}_j = Y} \Pr[\mathcal{Z}(\bar{Y})]$. Therefore

$$\begin{aligned}\sum_{j>1} \mathbb{E}_{Y,j}[\chi_j^{-1}] &= \sum_{j>1}\sum_{Y\subseteq X} \Pr[Z(Y,j)]\cdot \chi_j^{-1} \\ &= \sum_{j\geq 1}\sum_{\bar{Y}\subset X^n} \Pr[\mathcal{Z}(\bar{Y})]\cdot\chi_j^{-1} \\ &= \sum_{\bar{Y}\subset X^n} \Pr[\mathcal{Z}(\bar{Y})] \sum_{j\geq 1}\chi_j^{-1}\end{aligned}$$

Now it is enough to show that for any $X_0, X_1, \dots, X_m$ that may occur in the start-partition algorithm (*i.e.* $\Pr[\mathcal{Z}(\bar{Y})] > 0$, given that $\bar{Y}_j = X \setminus \bigcup_{\ell<j} X_\ell$) we have $\sum_{j=1}^m \chi_j^{-1} \leq 1$. This holds because for any $2 \leq \ell < j \leq m$ we have that $B_{Y_\ell, d_{Y_\ell}}(p_\ell, \epsilon\Delta/16) \subseteq X_\ell$, and $Y_j \cap X_\ell = \emptyset$, *i.e.* $B_{Y_\ell, d_{Y_\ell}}(p_i, \epsilon\Delta/16) \cap B_{Y_j, d_{Y_j}}(p_j, \epsilon\Delta/16) = \emptyset$. Therefore

$$\sum_{j=1}^m \chi_j^{-1} \leq |X|^{-1} \sum_{j=1}^m B_{Y_j, d_j}(p_j, \epsilon\Delta/16) \leq 1.$$

□

Claim 8. *For any cluster $X \subseteq V$, edge $u, v \in X$, $(u,v) \in E$, subgraph $Y \subset X$ we have*

$$\begin{aligned}&\Pr[\mathcal{E}(u,v) \wedge \neg\mathcal{F}(u) \mid \neg\mathcal{G}(u) \wedge \mathcal{Z}(Y,u)] \\ &\quad\leq\ 12 d(u,v)\max\{1, \log\chi(X,Y,u)\}/(\epsilon\cdot\Delta)\end{aligned}$$

Proof. If $d(u,v) \geq \epsilon\cdot\Delta/12$ the the claim is trivial, so assume it is smaller. Let $j > 1$ be such that the next cone to be cut is X_j (the value of j is not relevant, we fix it in order to simplify the notation), and recall that fixing $Y = Y_{j-1}$ determines deterministically p_j, x_j and χ_j. Let $\rho = \rho(X_0 \cup Y, Y, x_0, x_j)$ be the appropriate cone metric on Y by which the next cone is cut.

$$\begin{aligned}&\Pr[\mathcal{E}(u,v) \wedge \neg\mathcal{F}(u) \mid \mathcal{Z}(Y,u)] \\ &\quad\leq\ \Pr[\mathcal{E}_j(u,v) \wedge \neg\mathcal{F}_j(u) \mid \mathcal{Z}(Y,u) \wedge \mathcal{Z}(Y)] \\ &\quad\leq\ \Pr[\mathcal{E}_j(u,v) \mid \rho(x_j,u)/\Delta \notin S_{N(j)} \wedge \mathcal{Z}(Y,u) \wedge \mathcal{Z}(Y)] \\ &\quad\leq\ \frac{\Pr[\mathcal{E}_j(u,v) \mid \rho(x_j,u)/\Delta \notin S_{N(j)} \wedge \mathcal{Z}(X,Y)]}{\Pr[\mathcal{Z}(Y,u) \mid \rho(x_j,u)/\Delta \notin S_{N(j)} \wedge \mathcal{Z}(X,Y)]}\end{aligned}$$

The first inequality holds since event $\mathcal{Z}(Y,u)$ implies that $u \in X_j$ so the events $\mathcal{E}(u,v)$ and $\mathcal{E}_j(u,v)$ are equivalent (the same holds for $\neg\mathcal{F}(u)$), and because $\mathcal{Z}(Y,u) \subseteq \mathcal{Z}(X,Y)$. The second is by the definition of $\mathcal{F}(u)$ (given that $u \in X_j$ it cannot be that $\rho(x_j,u)/\Delta$ falls in the interval $S_{N(j)}$), and since for any events A, B, $\Pr[A \wedge B] \leq \Pr[A \mid B]$. The third is by Bayes rule and since $\mathcal{E}_j(u,v) \wedge \mathcal{Z}(Y,u) = \mathcal{E}_j(u,v)$. Let ℓ be such that $\rho(x_j,u)/\Delta \in S_\ell$.

First we bound the denominator, noting that there is no prior information given about the distribution for the next choice of radius. Since $\ell < N(j)$ we can bound $\Pr[\mathcal{Z}(Y,u) \mid \rho(x_j,u)/\Delta \notin S_{N(j)} \wedge \mathcal{Z}(X,Y)] \geq 2^{-\ell}$, since with this probability the radius for the cone X_j will be chosen from $S_m \cdot \Delta$ with $m > \ell$ so it will large enough to contain u. The numerator $\Pr[\mathcal{E}_j(u,v) \mid \rho(x_j,u)/\Delta \notin S_{N(j)} \wedge \mathcal{Z}(X,Y)]$ can be bounded by $\frac{1}{2^{\ell-1}} \cdot \frac{1}{2} \cdot \frac{d(u,v)}{\Delta|S_\ell|}$, which is the probability that we reach the ℓ-th interval, not continue to the next one (note that the next interval exists because $\ell < N(j)$) and when choosing r_j uniformly from S_ℓ, it happens to be the place that separates u, v. The probability for the first event is $2^{-(\ell-1)}$, the second is $1/2$, and the third is $\frac{d(u,v)}{\Delta|S_\ell|}$. Since $|S_\ell| \geq \min\{1, \frac{1}{\log\chi_j}\} \cdot \frac{\epsilon}{12}$ it follows that $\Pr[\mathcal{E}_j(u,v) \mid \rho(x_j,u)/\Delta \notin S_{N(j)} \wedge \mathcal{Z}(X,Y)] \leq \frac{12 d(u,v)\max\{1,\log\chi_j\}}{\epsilon\cdot\Delta\cdot 2^\ell}$. We conclude that

$$\Pr[\mathcal{E}(u,v) \wedge \neg\mathcal{F}(u) \mid \mathcal{Z}(Y,u)] \leq \frac{12 d(u,v)\max\{1,\log\chi_j\}}{\epsilon\cdot\Delta}.$$

□

Proof of Lemma 6. Fix any $i \geq 1$ and $X^{(i)} = X^{(i)}(u)$. As described before we partition the event $\mathcal{E}(u,v) = \mathcal{E}(X^{(i)}, u, v)$, given a fixed cluster $X^{(i)}$ into the three cases.

$$\begin{aligned}&\Pr[\mathcal{E}(u,v)] \\ &\quad=\ \Pr[\mathcal{E}(u,v) \wedge \mathcal{F}(u)] + \Pr[\mathcal{E}(X^{(i)},u,v) \wedge \neg\mathcal{F}(u)] \\ &\quad=\ \Pr[\mathcal{E}(u,v) \wedge \mathcal{F}(u)] + \Pr[\mathcal{E}(u,v) \wedge \mathcal{G}(u)] \\ &\qquad + \Pr[\mathcal{E}(u,v) \wedge \neg\mathcal{F}(u) \wedge \neg\mathcal{G}(u)]\end{aligned}$$

The last equality holds since event $\mathcal{G}(u)$ implies that $\neg\mathcal{F}(u)$. We claim that the following hold:

$$\Pr[\mathcal{E}(X^{(i)},u,v) \wedge \mathcal{F}(u) \mid X^{(i)}] \leq 48 d(u,v)/(\epsilon\Delta) \quad (1)$$

$$\Pr[\mathcal{E}(u,v) \wedge \mathcal{G}(u) \mid X^{(i)}] \leq 5 d(u,v)/(\epsilon\Delta) \quad (2)$$

$$\begin{aligned}&\Pr[\mathcal{E}(u,v) \wedge \neg\mathcal{F}(u) \wedge \neg\mathcal{G}(u) \mid X^{(i)}] \quad (3) \\ &\quad\leq\ 12 d(u,v)/(\epsilon\Delta)\cdot\mathbb{E}_{Y,u}[\max\{1,\log\chi(X,Y,u)\}]\end{aligned}$$

(1) holds directly from Claim 7. (2) since the radius of the central ball is chosen uniformly from interval of length $\Delta/(16c) \geq \epsilon\Delta$, and for the first cone from interval of length $\epsilon\Delta/4$. (3) holds by using Claim 8 and writing

$$\begin{aligned}&\Pr[\mathcal{E}(u,v) \wedge \neg\mathcal{F}(u) \wedge \neg\mathcal{G}(u)] \\ &\quad\leq\ \mathbb{E}_{Y,u}\left[\Pr[\mathcal{E}(u,v) \wedge \neg\mathcal{F}(u) \mid \neg\mathcal{G}(u)]\right] \\ &\quad\leq\ \frac{12 d(u,v)}{\epsilon\cdot\Delta}\mathbb{E}_{Y,u}[\max\{1,\log\chi(X,Y,u)\}]\end{aligned}$$

Combining these three equation yields that for $C = 65$

$$\Pr[\mathcal{E}(u,v)] \leq C\cdot d(u,v)/(\epsilon\Delta)\cdot\mathbb{E}_{Y,u}[\max\{1,\log\chi(X,Y,u)\}].$$

Recall that $k = 20c(\ln(1/\epsilon)+5)$, and Corollary 3 suggests that for any cluster X and any $j \geq 0$ that $\mathrm{rad}_{x_j}(X_j) \leq$

$(1-1/(20c))\mathrm{rad}_{x_0}(X)$, hence for any event $X^{(i+k)}$, given that $X^{(i)}$ happened

$$\mathrm{rad}(X^{(i+k)}) \le (1-1/(20c))^k \cdot \mathrm{rad}(X^{(i)}) \le \epsilon \cdot \mathrm{rad}(X^{(i)})/32,$$

therefore $\mathrm{diam}(X^{(i+k)}) \le \epsilon \cdot \mathrm{rad}(X^{(i)})/16$ and by definition $u \in X^{(i+k)}$, so fixing any Y such that event $\mathcal{Z}(X^{(i)}, Y, u)$ occurred then if $X^{(i+k)} \subseteq Y$ also $X^{(i+k)} \subseteq B_{Y,d_Y}(u, \epsilon \cdot \mathrm{rad}(X^{(i)})/16)$.

$$\begin{aligned}
&\mathbb{E}_{Y,u}[\log \chi(X^{(i)}, Y, u)] \\
&\quad = \log |X^{(i)}| - \mathbb{E}_{Y,u}[\log |B_{Y,d_Y}(u, \epsilon \cdot \mathrm{rad}(X^{(i)})/16|] \\
&\quad \le \log |X^{(i)}| - \\
&\qquad \mathbb{E}_{Y,u}\left[\sum_{X^{(i+k)} \subseteq Y} \Pr[X^{(i+k)} \mid \mathcal{Z}(X^{(i)}, Y, u)] \log |X^{(i+k)}|\right] \\
&\quad = \log |X^{(i)}| - \sum_{X^{(i+k)} \subseteq X^{(i)}} \Pr[X^{(i+k)} \mid X^{(i)}] \cdot \log |X^{(i+k)}|
\end{aligned}$$

We conclude that

$$\begin{aligned}
&\mathbb{E}_{X^{(i)}}\left[\Pr[\mathcal{E}(X^{(i)}, u, v)]\right] \\
&\quad \le \mathbb{E}_{X^{(i)}}[\log |X^{(i)}|] - \\
&\qquad \mathbb{E}_{X^{(i)}}\left[\sum_{X^{(i+k)} \subseteq X^{(i)}} \Pr[X^{(i+k)} \mid X^{(i)}] \log |X^{(i+k)}|\right] \\
&\quad = \mathbb{E}_{X^{(i)}}[\log |X^{(i)}|] - \\
&\qquad \left[\sum_{X^{(i)}} \Pr[X^{(i)}] \sum_{X^{(i+k)} \subseteq X^{(i)}} \Pr[X^{(i+k)} \mid X^{(i)}] \log |X^{(i+k)}|\right] \\
&\quad = \mathbb{E}_{X^{(i)}}[\log |X^{(i)}|] - \mathbb{E}_{X^{(i+k)}} \log |X^{(i+k)}|
\end{aligned}$$

□

Acknowledgments: We would like to thank Michael Elkin for initial discussions on the problem, Harald Räcke and Yuval Emek for comments on a preliminary version.

References

[1] I. Abraham, Y. Bartal, and O. Neiman. Nearly tight low stretch spanning trees, Arxiv 0808.2017, 2008.

[2] I. Abraham, Y. Bartal, and O. Neiman. Advances in metric embedding theory. In *thirty-eighth annual ACM symposium on Theory of computing*, pages 271–286, New York, NY, USA, 2006. ACM Press.

[3] Noga Alon, Richard M. Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM J. Comput.*, 24(1):78–100, 1995.

[4] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 184–193. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[5] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *30th Annual ACM Symposium on Theory of Computing*, pages 183–193, 1998.

[6] Y. Bartal. Graph decomposition lemmas and their role in metric embedding methods. In *12th Annual European Symposium on Algorithms*, pages 89–97, 2004.

[7] Erik Boman, Bruce Hendrickson, and Stephen Vavasis. Solving elliptic finite element systems in near-linear time with support preconditioners, 2004.

[8] Gruia Calinescu, Howard J. Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Symposium on Discrete Algorithms*, pages 8–16, 2001.

[9] Moses Charikar, Chandra Chekuri, Ashish Goel, and Sudipto Guha. Rounding via trees: deterministic approximation algorithms for group steiner trees and k-median. In *thirtieth annual ACM symposium on Theory of computing*, pages 114–123, New York, NY, USA, 1998. ACM Press.

[10] Kedar Dhamdhere, Anupam Gupta, and Harald Räcke. Improved embeddings of graph metrics into random trees. In *seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 61–69, New York, NY, USA, 2006. ACM.

[11] Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. Lower-stretch spanning trees. In *thirty-seventh annual ACM symposium on Theory of computing*, pages 494–503, New York, NY, USA, 2005. ACM Press.

[12] Yuval Emek and David Peleg. A tight upper bound on the probabilistic embedding of series-parallel graphs. In *seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1045–1053, New York, NY, USA, 2006. ACM.

[13] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *thirty-fifth annual ACM symposium on Theory of computing*, pages 448–455. ACM Press, 2003.

[14] T.C. Hu. Optimum communication spanning trees. *SIAM Journal on Computing*, pages 188–195, 1974.

[15] Nathan Linial and Michael Saks. Decomposing graphs into regions of small diameter. In *second annual ACM-SIAM symposium on Discrete algorithms*, pages 320–330, Philadelphia, PA, USA, 1991. Society for Industrial and Applied Mathematics.

[16] D. Peleg and E. Reshef. Deterministic polylogarithmic approximation for minimum commu- nication spanning trees. In *25th International Colloq. on Automata, Languages and Programming*, pages 670–681, 1998.

[17] Daniel A. Spielman and Shang-Hua Teng. Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems. In *thirty-sixth annual ACM symposium on Theory of computing*, pages 81–90, 2004.