Ben-Gurion University of the Negev

Department of Computer Science

Thesis submitted as partial fulfillment of the requirements for M.Sc. degree.

Terminal Embeddings

Author:
Arnold Filtser

Supervisors:
Prof. Michael Elkin
Dr. Ofer Neiman

February 23, 2015

Abstract

In this thesis we study *terminal embeddings*, in which one is given a graph $G = (V, E)$, and a subset $K \subseteq V$ of its vertices are specified as *terminals*. The objective is to embed the graph into a simpler graph, or a normed space, while approximately preserving all distances among pairs that contain a terminal. We devise such embeddings in various settings, and conclude that even though we have to preserve $\approx |K| \cdot |V|$ pairs, the distortion depends only on $|K|$, rather than on $|V|$.

We also strengthen this notion, and consider embeddings that must preserve the distances between all pairs, but have improved distortion guarantees for pairs containing a terminal. Surprisingly, in many settings we show that such embeddings exist, and have optimal distortion bounds in both regimes.

Acknowledgements

I would like to express my special appreciation and thanks to my advisers Prof. Michael Elkin and Dr. Ofer Neiman for their continuous support of my M.sc studies and research, for their patience, motivation, enthusiasm, and immense knowledge. I would like to specially thank Michael for suggesting the great research topic of terminal embeddings, which gave me a lot of interesting questions to think about, which lead to innumerous hours of fun. Moreover, thank you for the endless patience to read and repair my poor English and style. I would also like to specially thank Ofer for all the hours of fruitful discussions, for the constant availability, for suggesting great ideas when I was stuck and for working late with me before the deadline of conferences.

I would like to thank Ben-Gurion University, for the great research environment, for the funding, flexibility and thoughtfulness. Special thanks to many faculty members who revealed to me the endless beauty of math. A great thanks also to the technical and administrative staff that make this research pleasant and possible.

A special thanks to my family. Words cannot express how grateful I am to my mother and father for raising the love to math and knowledge in me, investing time and money in my education, believing in my abilities and pushing me through excellence. Thanks a lot also to my siblings, grandparents and all further family for love and support. I would also like to thank all of my friends and colleagues who supported me and incentive me to strive towards my goal. You make a great environment around me of curiosity and pursuit of truth.

I would like to express appreciation to my beloved wife Omrit for fruitful discussions, who spent sleepless nights with, was my inspiration and was always my support in the hard moments. I would like to thank my daughter Naamma, for her patience to hear my ideas in our walks, listening to me reading papers to her instead of proper children stories and accompanying me in conferences instead of going to kinder-garden.

At the end, I would like to express respect and admiration to all the math, and theoretical computer science communities. I will finish by citation of Sir Isaac Newton "If I have seen further it is by standing on ye sholders of Giants."

Contents

1 Introduction **4**

I General Transformations **8**

II Terminal Distortion By a Single Tree **13**

2 Introduction and preliminaries **13**

3 Algorithms **14**

4 Lower bounds **23**

5 NP-completeness of finding minimum weight spanning tree with terminal stretch $(2k-1)\,\alpha$ **32**

III Probabilistic Embedding **36**

6 Probabilistic Embedding into Ultrametrics with Strong Terminal Distortion **36**

7 Strong Terminal Embedding of Graphs into a Distribution of Spanning Trees **38**

IV Average stretch **56**

8 Terminal Scaling Embedding into an Ultrametric **56**

9 Terminal Scaling Embedding into a Spanning Tree **63**

V Conclusion, Future Work and Open Problems **74**

1 Introduction

Embedding of finite metric spaces is a very successful area of research, due to both its algorithmic applications and its natural geometric appeal. Given two metric space (X, d_X), (Y, d_Y), we say that X embeds into Y with distortion α if there is a map $f : X \to Y$ and a constant $c > 0$, such that for all $u, v \in X$,

$$d_X(u, v) \leq c \cdot d_Y(f(u), f(v)) \leq \alpha \cdot d_X(u, v) \ .$$

Some of the basic results in the field of metric embedding are: a theorem of Bourgain [Bou85], asserting that any metric space on n points embeds with distortion $O(\log n)$ into Euclidean space (which was shown to be tight by Linial et al. [LLR95]), and probabilistic embedding into a distribution of ultrametrics (or trees) by works of Bartal and Fakcharoenphol et al. [Bar96, FRT04], with expected distortion $O(\log n)$ (which is also tight [Bar96]).

In this thesis we consider a natural variant of embedding, in which the input consists of a finite metric space or a graph, and in addition, a subset of the points are specified as *terminals*. The objective is to embed the metric into a simpler metric (e.g., Euclidean metric), or into a simpler graph (e.g., a tree), while approximately preserving the distances between the terminals to *all other points*. We show that such embeddings, which we call *terminal embeddings*, can have improved parameters compared to embeddings that must preserve all pairwise distances. In particular, the distortion (and the dimension in embedding to normed spaces) depends only on the number of terminals, regardless of the cardinality of the metric space.

We also consider a strengthening of this notion, which we call *strong terminal embedding*. Here we want a distortion bound on *all pairs*, and in addition an improved distortion bound on pairs that contain a terminal. Such strong terminal embeddings may enhance the classical embedding results, essentially saying that one can obtain the same distortion guarantees on all pairs, with the option to select some of the points, and get improved bounds on the distances between any selected point to the rest of the points.

As a possible motivation for studying such embeddings, consider a scenario in which a certain network of clients and servers is given as a weighted graph (where edges correspond to links, weights to communication/travel time). In some settings, it could be that we only care about distances between clients and servers, and that there are few servers. We would like to have a simple structure, such as a tree spanning the network, so that the client-server distances in the tree are approximately preserved. It is well known that embedding a graph into a tree may cause distortion linear in the number of points [RR98]. However, if one only cares about client-server distances, we show that it is possible to obtain distortion $2k - 1$, where k is the number of servers, and that this is tight. Furthermore, we study possible tradeoffs between the distortion and the total weight of the obtained tree. This generalizes the notion of *shallow light trees*.

We then address probabilistic approximation of metric spaces and graphs by ultrametrics and spanning trees. This line of work started with the results of [AKPW95, Bar96, EEST08], and culminated in the $O(\log n)$ expected distortion for ultrametrics by [FRT04], and $\tilde{O}(\log n)$ for spanning trees by [AN12]. These embeddings found numerous algorithmic applications, in various settings, see [FRT04, AN12] and the references therein for details. In their work on Ramsey partitions, Mendel and Naor [MN06] implicitly show that there exists a probabilistic embedding into ultrametrics with expected terminal distortion $O(\log k)$ (see Section 1.3 for the formal definitions). Here we generalize this result by obtaining a *strong* terminal embedding with the same expected $O(\log k)$ distortion guarantee for all pairs containing a terminal, and $O(\log n)$ for all other pairs. We also show a similar result that extends the embedding of [AN12] into spanning trees, with $\tilde{O}(\log k)$ expected distortion for pairs containing a terminal, and $\tilde{O}(\log n)$ for all pairs. In [ABN07], it was shown that the average distortion (taken over all pairs) in an embedding into a single tree can be bounded by $O(1)$ (in contrast to the $\Omega(\log n)$ lower bound for the average stretch over edges). Here we extend and simplify their result, and obtain $O(1)$ average terminal distortion, that is, the average is over pairs containing a terminal. We do this both in the ultrametric and in the spanning tree settings.

We show that there exists a general phenomenon; essentially any known metric embedding result can be transformed into a terminal embedding, while paying only a constant blow-up in the distortion. An immediate implication is a terminal embedding of any finite metric into any ℓ_p space with terminal distortion $O(\log k)$, using only $O(\log k)$ dimensions. We also show that embeddings into normed spaces have their *strong* terminal embedding counterpart, albeit there is an additional requirement: we assume that there is a Lipschitz extension of the black-box embedding to the non-terminals (to be defined formally later).[1] Once again, many of the known embeddings into normed spaces satisfy this requirement. On the other hand, these results do not apply in a graph setting (e.g., when we want an embedding into a spanning subgraph), and we provide specific embeddings for many such scenarios.

This thesis is based on a paper that is currently under submission [EFN]. In this paper one can find results that do not appear in the thesis (about distance oracles and general spanners that preserve distance from the terminals). On the other hand, there are results that appear in this thesis, but do not appear in the paper. (Mainly results from Part II.) There are some additional results that do not appear in any of them, we mention some of them in Part V.

1.1 Related Work

Already in the pioneering work of [LLR95], an embedding that has to provide a distortion guarantee for a subset of the pairs is presented. Specifically, in the context of the Sparsest-Cut problem, [LLR95] devised a non-expansive embedding of an arbitrary metric into ℓ_1, with distortion at most $O(\log k)$ for a set of k specified demand pairs.

In the context of preserving distances just between the terminals, [Gup01] showed that given a tree and a set of terminals in it, one can find another tree (with only the terminals as the vertices) that preserves distances between the terminals up to a factor of 8. In [CXKR06] this factor was shown to be tight, and that one can in fact obtain a tree which is a minor of the original tree. Recently, [EGK+10] showed that for planar graphs there exists a distribution of minors over the terminals, which in expectation preserves the distances between terminals up to a constant factor. Even more recently, [KKN14] showed that any graph with a set of terminals, one can obtain a minor on the terminals with polylogarithmic distortion.

In their work on the requirement cut problem, among other results, [GNR10] obtain for any metric with k specified terminals, a distribution over trees with expected expansion $O(\log k)$ for all pairs, and which is non-contractive for terminal pairs. (Note that this is different from our setting, as the extra guarantee holds for terminals only, not for pairs containing a terminal. However, their techniques are quite similar to ours.)

The notion of preserving certain quantities over a set of terminals appears in other contexts as

[1] The basic idea leading to the second transformation was communicated to us by an anonymous referee, and we are grateful to her.

well. For instance, [Moi09] studied cut and flow-sparsifiers. Given a graph and a set of terminals, the objective is to find a graph only on the terminals that approximates all minimum cuts. There are interesting duality relations between preserving distances and cuts, and we refer the reader to [Räc08, AF09] for more on this.

1.2 Organization

Our general transformations which convert ordinary embeddings into terminal ones are presented in Part I. The tradeoff between terminal distortion and lightness in a single tree embedding is shown in Part II. We also show in Part II several lower bounds, and prove NP-hardness of certain optimization variants of the studied problems. The probabilistic embeddings into ultrametrics and spanning trees with strong terminal distortion are given in Part III. The embeddings into a single tree with constant average terminal distortion, both for ultrametrics and spanning trees, appear in Part IV.

1.3 Preliminaries

Here we provide formal definitions for the notions of terminal distortion. Let (X, d_X) be a finite metric space, with $K \subseteq X$ a set of terminals. Throughout the thesis we assume $|K| \leq |X|/2$, as otherwise one may just use the standard results.

Definition 1. *Let (X, d_X) be a metric space, and let $K \subseteq X$ be a subset of terminals. For a target metric (Y, d_Y), an embedding $f : X \to Y$ has* terminal distortion α *if there exists $c > 0$, such that for all $v \in K$ and $u \in X$,*[2]

$$d_X(v, u) \leq c \cdot d_Y(f(v), f(u)) \leq \alpha \cdot d_X(v, u) \ .$$

We say that the embedding has strong terminal distortion (α, β) *if it has terminal distortion α, and in addition there exists $c' > 0$, such that for all $u, w \in X$,*

$$d_X(u, w) \leq c' \cdot d_Y(f(u), f(w)) \leq \beta \cdot d_X(u, w) \ .$$

We will also consider probabilistic embeddings.

Definition 2. *For a class of metrics $\mathcal{Y}$, a distribution $\mathcal{D}$ over embeddings $f_Y : X \to Y$ with $Y \in \mathcal{Y}$ has* expected terminal distortion α *if each f_Y is non-contractive (that is, for all $u, w \in X$ and $Y \in \text{supp}(\mathcal{D})$, it holds that $d_X(u, w) \leq d_Y(f_Y(u), f_Y(w))$), and for all $v \in K$ and $u \in X$,*

$$\mathbb{E}_{Y \sim \mathcal{D}}[d_Y(f_Y(v), f_Y(u))] \leq \alpha \cdot d_X(v, u) \ .$$

The distribution $\mathcal{D}$ has strong expected terminal distortion (α, β) *if it has* expected terminal distortion α*, and in addition for all $v, u \in X$,*

$$\mathbb{E}_{Y \sim \mathcal{D}}[d_Y(f_Y(v), f_Y(u))] \leq \beta \cdot d_X(v, u) \ .$$

Finally, we consider average distortion of a single embedding:

[2] In most of our results the embedding has a one-sided guarantee (that is, non-contractive or non-expansive) for all pairs.

Definition 3. *For $f : X \to Y$ a non-contractive embedding, the* average terminal distortion *of f is*[3]

$$\frac{1}{|K| \cdot |X|} \sum_{v \in K, u \in X} \frac{d_Y(f(v), f(u))}{d_X(v, u)} \ .$$

Let $G = (V, E, w)$ be a (weighted) graph. Throughout the thesis all graphs are undirected. For $U \subseteq V$, $G[U]$ is the induced graph on the vertices of U, with edge set $E(U)$. Let d_G be the shortest path metric on G. For a subgraph H of G, $w(H) = \sum_{e \in E(H)} w(e)$, and define the *lightness* of H to be $\Psi(H) = \frac{w(H)}{w(MST(G))}$, where $MST(G)$ is a minimum spanning tree of G.

For a metric space (X, d), let $\text{diam}(X) = \max_{y,z \in X}\{d(y, z)\}$. For any $x \in X$ and $r \geq 0$ let $B_{(X,d)}(x, r) = \{y \in X \mid d(x, y) \leq r\}$ (we often omit the subscript when the metric is clear from context) be the ball centered at x with *radius* r. Denote by $B^o(x, r) = \{u \in X : \ d(x, u) < r\}$ the open ball with radius r and center x. For $A \subseteq X$, $|A|_k$ denotes the number of terminals in A, i.e., $|A|_k = |A \cap K|$. A metric d' on X *dominates* d if for all $x, y \in X$, $d(x, y) \leq d'(x, y)$. For a point $x \in X$ and a subset $A \subseteq X$, $d(x, A) = \min_{a \in A}\{d(x, a)\}$.

For $1 \leq p < \infty$, ℓ_p^k is a normed vector space, over the vector space $\mathbb{R}^k$. The ℓ_p norm of the vector $\vec{x} = (x_1, x_2, \ldots, x_k)$ is $\|\vec{x}\|_p = \left(\sum_{i=1}^k |x_i|^p\right)^{\frac{1}{p}}$. The special case where $p = 2$ is called Hilbert space. Consider also ℓ_∞^k, a normed vector space over $\mathbb{R}^k$ with the norm $\|\vec{x}\|_\infty = \max_i \{|x_i|\}$. For two functions $f : X \to \mathbb{R}^a$ and $g : X \to \mathbb{R}^b$, $f \oplus g$ is a function from X into $\mathbb{R}^{a+b}$ such that the first a coordinates of $f \oplus g$ are identical to the coordinates of f, and the last b coordinates of $f \oplus g$ are identical to the coordinates of g.

We use the symbol $\tilde{O}$ as a variant of big O notation that ignores poly-logarithmic factors. Thus, $f(n) = \tilde{O}(g(n))$ is shorthand for $f(n) \leq g(n) \cdot \log^k(g(n))$ for some constant k.

An ultrametric (U, d) is a metric space satisfying a strong form of the triangle inequality, that is, for all $x, y, z \in U$, $d(x, z) \leq \max\{d(x, y), d(y, z)\}$. The following definition is known to be an equivalent one (see [BLMN05]).

Definition 4. *An ultrametric U is a metric space (U, d) whose elements are the leaves of a rooted labeled tree T. Each $z \in T$ is associated with a label $\Phi(z) \geq 0$ such that if $q \in T$ is a descendant of z then $\Phi(q) \leq \Phi(z)$ and $\Phi(q) = 0$ iff q is a leaf. The distance between leaves $z, q \in U$ is defined as $d_T(z, q) = \Phi(lca(z, q))$ where $lca(z, q)$ is the least common ancestor of z and q in T.*

[3]Note that pairs in $K \times K$ are counted twice, but this can only affect the average by a factor of 2. We also assume that if $u = v$ then $0/0 = 0$.

Part I

General Transformations

We first note that a transformation in the setting of embedding metrics into ultrametrics was already attained by [MN06]. In their work on Ramsey partitions, [MN06] showed that if a subset $Y \subseteq X$ of a metric space (X, d) is embedded into an ultrametric with distortion α, then this embedding can be extended to the whole of X, such that the distortion of any pair $y \in Y$ and $x \in X$ is at most 6α. This directly implies that any embedding result for a subset-closed family[4] of metrics into a single ultrametric with distortion $\alpha(\cdot)$ (that depends on the cardinality of the metric), can be translated to an embedding with terminal distortion $O(\alpha(k))$. Thus the following theorem is implicit from [MN06].

Theorem 1 ([MN06])**.** *Let $\mathcal{X}$ be a subset-closed family of metric spaces. Let $(X, d) \in \mathcal{X}$, and a set of terminals $K \subseteq X$ of size $|K| = k$. If there is a function $\alpha : \mathbb{N} \to \mathbb{R}$, such that every $(Z, d_Z) \in \mathcal{X}$ of size $|Z| = m$ embeds with non-contractive embedding ρ into an ultrametric with distortion $\alpha(k)$, then there is an non-contractive embedding $\tilde{\rho}$ of X into an ultrametric with* terminal *distortion at most $6\alpha(k)$.*

Actually, in their proof, [MN06] showed a somewhat stronger bound on the distortion, that $\hat{\rho}$ in fact satisfies:

$$\text{For all } t \in K \text{ and } x \in X, \;\; \tilde{\rho}(x, t) \leq 3 \cdot \max\{\rho(k_x, t), d_X(x, t)\} \;. \tag{1}$$

Where k_x is the closest point in K to x. In [FRT04], it was shown that for each metric space (X, d) of size n, there is a distribution $\mathcal{D}$ over non-contracting embeddings from X into ultrametrics with expected distortion $O(\log n)$. Thus, Theorem 1 reinforced by (1), combined with [FRT04], implies a probabilistic embedding of any metric space into ultrametrics with expected terminal distortion $O(\log k)$:

Theorem 2. *Let (X, d) be a metric spaces and $K \subseteq X$ a subset of terminals of size k, there is a distribution over embeddings of X into ultrametrics with expected* terminal *distortion $O(\log k)$.*

Proof. By [FRT04], there is a distribution $\mathcal{D}$ over non-contractive embeddings from K into ultrametrics with expected distortion $O(\log k)$. By Theorem 1, for each $\rho \in \text{supp}(\mathcal{D})$ there is an embedding $\tilde{\rho}$ from X into an ultrametric such that (1) holds. Note that $\tilde{\rho}$ is a non-contractive embedding. Set $\mathcal{D}'$ to be the distribution that picks the embedding $\tilde{\rho}$ with probability $\Pr_{\mathcal{D}}(\rho)$. Fix any $x \in X$ and $t \in K$, then

$$\begin{aligned}
\mathbb{E}_{\tilde{\rho}\sim\mathcal{D}'}[\tilde{\rho}(x,t)] &\overset{(1)}{\leq} \mathbb{E}_{\tilde{\rho}\sim\mathcal{D}'}[3 \cdot \max\{\rho(k_x,t), d(x,t)\}] \\
&\leq 3\left(\mathbb{E}_{\rho\sim\mathcal{D}}[\rho(k_x,t)] + d(x,t)\right) \\
&\leq 3\left(d(x,t) + O(\log k) \cdot d(k_x,t)\right) \\
&\leq 3 \cdot d(x,t) + O(\log k) \cdot d(x,t) = O(\log k) \cdot d(x,y) \;,
\end{aligned}$$

where the last inequality is due to $d(k_x, t) \leq d(k_x, x) + d(x, t) \leq 2d(x, t)$. □

[4] A family of metrics $\mathcal{F}$ is subset-closed if for every $(X, d_X) \in \mathcal{F}$, the metric d induced on any $Y \subseteq X$ is in $\mathcal{F}$ as well. All the families considered in this thesis are subset-closed.

In what follows we show more general theorems of a similar flavor, for embeddings into normed spaces and graph families. We start with a small warm up, for the case where we are interested in embedding between Euclidean spaces.

Theorem 3. *For a set X of n points in ℓ_2^d and a subset of terminals $K \subseteq X$ of size k, there is an embedding $f : \ell_2^d \to \ell_2^k$ with terminal distortion* 1.

Proof. Without loss of generality, we assume that $v_1 = \vec{0}$ (otherwise shift all the points). Consider $V = span\{K\}$, a vector space of dimension $r \leq k-1$. Let $u_1, u_2, \ldots, u_r$ be an orthonormal basis for V. Define a linear transformation $T : V \to \ell_2^r$ by $T(u_i) = e_i$. Obviously, T is an isometry. We extend T to $\hat{T} : \ell_2^d \to \ell_2^{r+1}$ in the following way: for $x \in \ell_2^d$, x can be uniquely represented as $x = x' + x_\perp$, where $x' \in V$ and $x_\perp \in V^\perp$. Define $\hat{T}(x) = T(x') \oplus \|x_\perp\|_2$. Hence for every $v \in K$ and $x \in X$:

$$\begin{aligned}\left\|\hat{T}(v) - \hat{T}(x)\right\|_2^2 &= \left\|T(v) \oplus 0 - T(x') \oplus \|x_\perp\|_2\right\|_2^2 \\ &= \left\|T(v) - T(x')\right\|_2^2 + \left\|\vec{0} - x_\perp\right\|_2^2 \\ &= \left\|v - x'\right\|_2^2 + \left\|\vec{0} - x_\perp\right\|_2^2 = \|v - x\|_2^2 \ .\end{aligned}$$

Since $r \leq k-1$, $\hat{T}$ is the desired terminal embedding f. □

We say that a family of graphs $\mathcal{G}$ is *leaf-closed*, if it is closed under adding leaves. That is, for any $G \in \mathcal{G}$ and $v \in V(G)$, the graph G' obtained by adding a new vertex u and connecting u to v by an edge, has $G' \in \mathcal{G}$. Note that many natural families of graphs are leaf-closed, e.g. trees, planar graphs, minor-free graphs, bipartite graphs, bounded tree-width graphs, etc.

Theorem 4. *Let $\mathcal{X}$ be a subset-closed family of metric spaces. Let $(X, d) \in \mathcal{X}$ and a set of terminals $K \subseteq X$ of size $|K| = k$. If there are functions $\alpha, \gamma : \mathbb{N} \to \mathbb{R}$, such that every $(Z, d_Z) \in \mathcal{X}$ of size $|Z| = m$ embeds into $\ell_p^{\gamma(m)}$ with distortion $\alpha(m)$, then there is an embedding of X into $\ell_p^{\gamma(k)+1}$ with* terminal *distortion $2^{(p-1)/p} \cdot ((2\alpha(k))^p + 1)^{1/p}$.*

Additionally, if $\mathcal{G}$ is a leaf-closed family of graphs, and any $(Z, d_Z) \in \mathcal{X}$ of size $|Z| = m$ (probabilistically) embeds into (a distribution over) $\mathcal{G}$ with distortion $\alpha(m)$, then there is a (probabilistic) embedding of X into (a distribution over) $\mathcal{G}$ with terminal *distortion $2\alpha(k) + 1$.*

Proof. We start by proving the first assertion. Since $\mathcal{X}$ is subset-closed, $K \in \mathcal{X}$, and by the assumption there exists an embedding $f : K \to \mathbb{R}^{\gamma(k)}$ with distortion $\alpha(k)$ under the ℓ_p norm. We assume w.l.o.g that f is non-contractive. For each $x \in X \setminus K$, let $k_x \in K$ be the nearest point to x in K (that is, $d(x, K) = d(x, k_x)$). Extend f to an embedding $\hat{f} : X \to \mathbb{R}^{\gamma(k)+1}$ by defining for $t \in K$, $\hat{f}(t) = (f(t), 0)$, and for $x \in X \setminus K$, $\hat{f}(x) = (f(k_x), d(x, k_x))$. Fix any $t \in K$ and $x \in X$. Note that by definition of k_x, $d(x, k_x) \leq d(x, t)$, and by the triangle inequality, $d(t, k_x) \leq d(t, x) + d(x, k_x) \leq 2d(t, x)$, so that,

$$\begin{aligned}\|\hat{f}(t) - \hat{f}(x)\|_p^p &= \|f(t) - f(k_x)\|_p^p + d(x, k_x)^p \\ &\leq (\alpha(k) \cdot d(t, k_x))^p + d(x, k_x)^p \\ &\leq (2\alpha(k) \cdot d(t, x))^p + d(t, x)^p \\ &= d(t, x)^p \cdot ((2\alpha(k))^p + 1) \ .\end{aligned}$$

On the other hand, since f does not contract distances,

$$\begin{aligned}\|\hat{f}(t)-\hat{f}(x)\|_p^p &= \|f(t)-f(k_x)\|_p^p+d(x,k_x)^p\\ &\geq d(t,k_x)^p+d(x,k_x)^p\\ &\geq (d(t,k_x)+d(x,k_x))^p/2^{p-1}\\ &\geq d(x,t)^p/2^{p-1}\ ,\end{aligned}$$

where the second inequality is by the power mean inequality. We conclude that the terminal distortion is at most $2^{(p-1)/p}\cdot((2\alpha(k))^p+1)^{1/p}$.

For the second assertion, we start by showing the deterministic version. As $K\in\mathcal{X}$, there is a non-contractive embedding f of K into $G\in\mathcal{G}$ with distortion at most $\alpha(k)$. As above, for each $x\in X$ define k_x as the nearest point in K to x (note that if $x\in K$ then $k_x=x$). And for each $x\in X\setminus K$, add to G a new vertex $f'(x)$ that is connected by an edge of length $d(x,k_x)$ to $f(k_x)$. The resulting graph $G'\in\mathcal{G}$, because it is a leaf-closed family. Note that f' is non-contractive, as for all $x,y\in X$ it holds that

$$\begin{aligned}d_{G'_f}(f'(x),f'(y)) &= d_{G'_f}(f'(x),f'(k_x))+d_{G'_f}(f'(k_x),f'(k_y))+d_{G'_f}(f'(k_y),f'(y))\\ &= d(x,k_x)+d_{G_f}(f(k_x),f(k_y))+d(k_y,y)\\ &\geq d(x,k_x)+d(k_x,k_y)+d(k_y,y)\geq d(x,y)\ .\end{aligned}$$

Fix any $x\in X$ and $t\in K$, then as above $d(t,k_x)\leq 2d(t,x)$, and so

$$d_{G'}(f'(t),f'(x))=d_G(f(t),f(k_x))+d_{G'}(f'(x),f'(k_x))\leq\alpha(k)\cdot d(t,k_x)+d(x,k_x)\leq(2\alpha(k)+1)\cdot d(t,x)\ ,$$

which concludes the proof.

The extension to probabilistic embedding is straightforward. As $K\in\mathcal{X}$, it probabilistically embeds into a distribution over $\mathcal{G}$ with distortion $\alpha(k)$. Let $\mathcal{D}$ be a such distribution, where each $f\in\text{supp}(\mathcal{D})$ is an embedding from K into $G_f\in\mathcal{G}$ that chosen with probability $\Pr_{\mathcal{D}}(f)$ such that the expected distortion of each pair in K is bounded by $\alpha(k)$. For each such an embedding f, let $G'_f\in\mathcal{G}$ be the extension of G_f as defined in the previous case (i.e. for each $x\in X\setminus K$, add a new vertex $f(x)$ that is connected by an edge of length $d(x,k_x)$ to $f(k_x)$). Let f' be the appropriate embedding from X into G'_f. Note that f' is non-contractive. Set $\mathcal{D}'$ to be the distribution that picks the embedding f' with probability $\Pr_{\mathcal{D}}(f)$. Fix any $x\in X$ and $t\in K$, then

$$\begin{aligned}E_{f'\sim\mathcal{D}'}[d_{G'_f}(f'(t),f'(x))] &= E_{f'\sim\mathcal{D}'}[d_{G_f}(f(t),f(k_x))+d_{G'_f}(f'(x),f'(k_x))]\\ &= E_{f\sim\mathcal{D}}[d_{G_f}(f(t),f(k_x))]+d(x,k_x)\\ &\leq \alpha(k)\cdot d(t,k_x)+d(x,k_x)\\ &\leq 2\alpha(k)\cdot d(t,x)+d(t,x)=(2\alpha(k)+1)\cdot d(t,x)\ ,\end{aligned}$$

as required.

□

Remark 1. *Note that for any* $p,\alpha\geq 1$, $2\alpha+1\leq 2^{(p-1)/p}\cdot((2\alpha)^p+1)^{1/p}\leq 4\alpha$.

Next, we study strong terminal embeddings into normed spaces. Fix any metric (X,d), a set of terminals $K\subseteq X$ and $1\leq p\leq\infty$. Let $f:K\to\ell_p$ be a non-expansive embedding. We say that

f is *Lipschitz extendable*, if there exists a non-expansive $\hat{f} : X \to \ell_p$ which is an extension of f (that is, the restriction of $\hat{f}$ to K is exactly f). It is not hard to verify that any Fréchet embedding is Lipschitz extendable (in our context, it will be convenient to call an embedding $f : K \to \ell_p^t$ *Fréchet*, if there are sets $A_1, \dots, A_t \subseteq X$ such that for all $i \in [t]$, $f_i(x) = \frac{d(x,A_i)}{t^{1/p}}$). For example, the embeddings of [Bou85, KLMN05, ALN08] are Fréchet. Fréchet embeddings are indeed Lipschitz extendable by the following extension, which maps each $y \in X \setminus K$ to $\hat{f}(y) = \left(\frac{d(y,A_i)}{t^{1/p}}\right)_{i\in[t]}$. The triangle inequality implies $\hat{f}$ is non-expansive.

Theorem 5. *Let $\mathcal{X}$ be a subset-closed family of metric spaces. Let $(X, d) \in \mathcal{X}$ of size $|X| = n$ and a set of terminals $K \subseteq X$ of size $|K| = k$. If any $(Z, d_Z) \in \mathcal{X}$ of size $|Z| = m$ embeds into $\ell_p^{\gamma(m)}$ with distortion $\alpha(m)$ by a Lipschitz extendable map, then there is an embedding of X into $\ell_p^{\gamma(n)+\gamma(k)+1}$ with* strong *terminal distortion $O(\alpha(k), \alpha(n))$.*

Proof of Theorem 5. Let $(X, d) \in \mathcal{X}$ be a metric on n points, $K \subseteq X$ of size $|K| = k$. There is a non-expansive embedding $g : X \to \ell_p^{\gamma(n)}$ with distortion at most $\alpha(n)$. Since $\mathcal{X}$ is subset-closed, there exists a Lipschitz extendable embedding $f : K \to \ell_p^{\gamma(k)}$, which is non-expansive and has distortion $\alpha(k)$. Let $\hat{f}$ be the extension of f to X, note that by definition of Lipschitz extendability, $\hat{f}$ is also non-expansive. Finally, let $h : X \to \mathbb{R}$ be defined by $h(x) = d(x, K)$. The embedding $F : X \to \ell_p^{\gamma(n)+\gamma(k)+1}$ is defined by the concatenation of these maps $F = g \oplus \hat{f} \oplus h$.

Since all the three maps $g, \hat{f}, h$ are non-expansive, it follows that for any $x, y \in X$,

$$\|F(x) - F(y)\|_p^p \le \|g(x) - g(y)\|_p^p + \|\hat{f}(x) - \hat{f}(y)\|_p^p + |h(x) - h(y)|^p \le 3d(x, y)^p \ ,$$

so F has expansion at most $3^{1/p}$ for all pairs. Also note that

$$\|F(x) - F(y)\|_p \ge \|g(x) - g(y)\|_p \ge \frac{d(x, y)}{\alpha(n)} \ ,$$

which implies the distortion bound for all pairs is satisfied. It remains to bound the contraction for all pairs containing a terminal. Let $t \in K$ and $x \in X$, and let $k_x \in K$ be such that $d(x, K) = d(x, k_x)$ (it could be that $k_x = x$). If it is the case that $d(x, t) \le 3\alpha(k) \cdot d(x, k_x)$ then by the single coordinate of h we get sufficient contribution for this pair:

$$\|F(t) - F(x)\|_p \ge |h(t) - h(x)| = h(x) = d(x, k_x) \ge \frac{d(x, t)}{3\alpha(k)} \ .$$

The other case is that $d(x, t) > 3\alpha(k) \cdot d(x, k_x)$, here we will get the contribution from $\hat{f}$. First observe that by the triangle inequality,

$$d(t, k_x) \ge d(t, x) - d(x, k_x) \ge d(t, x)(1 - 1/(3\alpha(k))) \ge 2d(t, x)/3 \ . \tag{2}$$

By another application of the triangle inequality, using that $\hat{f}$ is non-expansive, and that f has distortion $\alpha(k)$ on the terminals, we get the required bound on the contraction:

$$\begin{aligned}
\|F(t)-F(x)\|_p &\geq \|\hat{f}(t)-\hat{f}(x)\|_p \\
&\geq \|\hat{f}(t)-\hat{f}(k_x)\|_p - \|\hat{f}(k_x)-\hat{f}(x)\|_p \\
&\geq \|f(t)-f(k_x)\|_p - d(x,k_x) \\
&\geq \frac{d(t,k_x)}{\alpha(k)} - \frac{d(t,x)}{3\alpha(k)} \\
&\overset{(2)}{\geq} \frac{2d(t,x)}{3\alpha(k)} - \frac{d(t,x)}{3\alpha(k)} \\
&= \frac{d(t,x)}{3\alpha(k)} .
\end{aligned}$$

□

Remark 2. *In Theorem 4, if the assumed embedding of every metric in $\mathcal{X}$ is Lipschitz extendable, then the terminal embedding can be made non-expansive (with the same dimension and distortion bounds), by using the techniques of Theorem 5.*

Remark 3. *The results of Theorem 4 and Theorem 5 can hold also if $\mathcal{X}$ is a family of graphs, provided that the assumed embedding allows Steiner nodes. That is, if Z is a graph with terminal set K, $|K|=k$, then there is a (Lipschitz extendable) embedding of K to $\ell_p^{\gamma(k)}$ with distortion $\alpha(k)$. Note that it is not clear whether the metric on K can be induced from some graph in the family. In fact, for many graph families (e.g. planar graphs), the following question is open: given a graph Z in the family with terminals K, is there another graph in the family over the vertex set K, that preserves the shortest-path distances with respect to Z (up to some constant).*

We list some implications of Theorem 4 and Theorem 5:

Corollary 1. *Let (X,d) be a metric space, and $K \subseteq X$ a set of terminals of size $|K|=k$. Then for any $1 \leq p \leq \infty$,*

1. *(X,d) can be embedded to $\ell_p^{O(\log k)}$ with terminal distortion $O(\log k)$.*
2. *(X,d) can be embedded to $\ell_p^{O(\log n+\log^2 k)}$ with strong terminal distortion $(O(\log k), O(\log n))$.*
3. *If (X,d) is an ℓ_2 metric, it can be embedded to $\ell_2^{O(\log k)}$ with terminal distortion $O(1)$.*
4. *If (X,d) is a planar metric, it can be embedded to ℓ_p with strong terminal distortion $(O((\log k)^{\min\{1/2,1/p\}}), O((\log n)^{\min\{1/2,1/p\}}))$.*
5. *If (X,d) is a negative type metric, it can be embedded to ℓ_2 with strong terminal distortion $(\tilde{O}(\sqrt{\log k}), \tilde{O}(\sqrt{\log n}))$.*

The corollary follows from known embedding results. (1) and (2) are from [Bou85], with improved dimension due to [ABN11], (3) is from [JL84], (4) from [KLMN05], and (5) from [ALN08].

Part II
Terminal Distortion By a Single Tree

2 Introduction and preliminaries

This part is dedicated to questions regarding finding a single spanning tree of a graph, where the goal is to minimize the terminal distortion of the tree. In this part of the thesis, we consider only embeddings of graphs into their spanning trees. The notation of terminal distortion (which we will call terminal stretch) in this setting is defined as follows: For a subset $K \subseteq V$ of terminals, and a spanning tree T of G, the *terminal stretch* of T, denoted by $TermStretch_{G,K}(T)$, is defined by

$$TermStretch_{G,K}(T) = max_{v \in K, u \in V} \left\{ \frac{d_T(v,u)}{d_G(v,u)} \right\}.$$

We are giving a special treatment for the case of a metric space rather than a general graph. (Recall that a metric space can be viewed as a complete graph.) Any algorithm for graphs works also for metric spaces, and every lower bound for metric spaces is also a lower bound for graphs.

We will show in Theorem 6, that for a terminal set of size k, there is always a spanning tree with terminal stretch at most $2k-1$. Moreover, this is tight since by Theorem 9 there are graphs where no spanning tree has terminal stretch less then $2k-1$. Hence, to make the problem more involved, instead of only minimizing the terminal stretch, we are interested in finding a tradeoff between the terminal stretch and the lightness.

The single-terminal case ($k = 1$) was analyzed by Awerbuch et al. [BAP92] and Khuller et al. [KRY93]. They showed that for every weighted graph $G = (V, E, w)$, designated vertex $v \in V$, and a parameter $\alpha > 1$, there exists a spanning tree T for G of lightness at most $1 + \frac{2}{\alpha-1}$ and terminal stretch with respect to v at most α. This tradeoff between $TermStretch$ and lightness was shown to be tight in [KRY93]. The running time of the algorithm in [KRY93] is $O(m + n \cdot \log n)$, where $m = |E|$ and $n = |V|$. We will henceforth refer to a spanning tree T that provides terminal stretch α with respect to a designated vertex and *lightness* β (for some parameters α and β) as a *shallow-light-tree* with parameters α and β, or shortly, an (α, β)-SLT. A small modification of Khuller et al.'s algorithm produces for any subset $A \subset V$ a forest F with stretch α with respect to A and *lightness* $1 + \frac{2}{\alpha-1}$. (A forest F is said to provide stretch α with respect to A if for every vertex $u \in V$, $d_F(A, u) \leq \alpha \cdot d_G(A, u)$. Each vertex $v \in A$ ends up belonging to different connected component of F. (To obtain such a forest F one should add a new vertex v_A to the graph and connect it to each of the vertices of A with edges of weight zero. Then we compute an (α, β)-SLT with respect to v_A in the modified graph. Finally, we remove v_A from the SLT. The resulting forest is F.)

A tree $T = (V', E', w')$ is called a *Steiner tree* for a graph $G = (V, E, w)$ if (1) $V \subseteq V'$, (2) for any edge $e \in E \cap E'$, the edge has the same weight in both G and T, i.e. $w(e) = w'(e)$, and (3) for any pair of vertices $u, v \in V$ it holds that $d_T(u, v) \geq d_G(u, v)$. The *minimum Steiner tree* T of G, denoted $SMT(G)$, is a Steiner tree of G with minimum weight. It is well-known that for any graph G, $w(SMT(G)) \geq \frac{1}{2} MST(G)$. (See, e.g., [GP68], Section 10.)

In Section 3 we devise algorithms for building trees that achieve terminal stretch $2k-1$, and algorithms for building light trees (lightness $O\left(\frac{1}{\alpha-1}\right)$) with terminal stretch roughly $(2k-1) \cdot \alpha$. In the end of this section, we provide efficient implementations of the mentioned algorithms. In

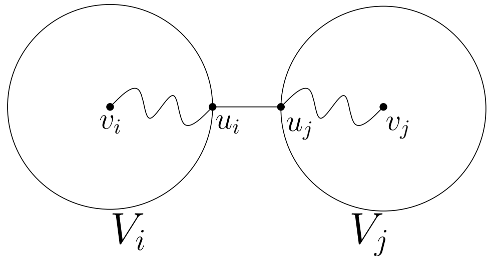


Figure 1: An illustration for the neighbors lemma: v_i and v_j are terminals. The edge $\{v_i, v_j\}$ belongs to the MST of the super-graph G'. The neighbors lemma asserts that there is an edge $\{u_i, u_j\}$ on the shortest path between v_i and v_j that crosses between V_i and V_j.

Section 4 we show that there are graphs and metrics such that the lowest terminal stretch that can be achieved for them by a spanning tree is $2k-1$. Then we show some lower bounds on the lightness of spanning trees with terminal stretch $(2k-1)\cdot\alpha$ for $\alpha > 1$. In Section 5 we show that the problem of finding a minimum weight tree with terminal stretch at most $(2k-1)\cdot\beta$ (for $1 \leq \beta \leq 1 + 1/(6k)$) is NP-Hard.

3 Algorithms

In all the algorithms we assume without loss of generality that all edge weights are different, and every two different paths have different length. If it is not the case then one can break ties in an arbitrary (but consistent) way.

3.1 k-terminal tree

Consider the $2k$-*cycle* $C_{2k} = (v_1, v_2, \ldots, v_{2k})$, for a positive integer parameter $k \geq 2$. Let $V'=\{v_2, v_4, \ldots, v_{2k}\}$ be the subset of terminals. It is easy to see that every spanning tree of C_{2k} has terminal stretch $2k-1$ with respect to V'. (See also Theorem 9.) Next we show that for every graph $G = (V, E, w)$, $w : E \to \mathbb{R}^+$, and subset $V' \subseteq V$ of k terminals (denote $V' = \{v_1, v_2, \ldots, v_k\}$) there exists a spanning tree T with terminal stretch $2k-1$ with respect to V'. In view of the aforementioned lower bound, this result is tight.

For every $i \in [k]$, define $V_i = \{u \in V \mid \text{for every } j \in [k] \setminus \{i\},\ d(v_i, u) < d(v_j, u)\}$ to be the subset of vertices which are closer to v_i than to any other terminal. Observe that $\{V_i\}_{i \leq k}$ is a partition of V, i.e., $V = \bigcup_{i=1}^{k} V_i$, and $V_i \cap V_j = \emptyset$ for every pair of different indexes $i, j \in [k]$. Let T_i be the shortest path tree (shortly, SPT) of V_i rooted in v_i. Now we define a super-graph $G' = \left(V', \binom{V'}{2}, w'\right)$, where $w'(v_i, v_j) = d_G(v_i, v_j)$. Let T' be the MST of G'.

Lemma 2. ***The neighbors lemma:*** *If $\{v_i, v_j\} \in T'$ then there exist $u_i \in V_i$, $u_j \in V_j$ such that $\{u_i, u_j\} \in E$ and the shortest path from v_i to v_j in G contains the edge $\{u_i, u_j\}$.* (See Figure 1 for an illustration.)

Proof. Suppose for contradiction that the shortest path $P_{i,j}$ from v_i to v_j contains a vertex $u \in V_r$, $r \neq i, j$. Then $d_G(v_i, v_r) \leq d_G(v_i, u) + d_G(u, v_r) < d_G(v_i, u) + d_G(u, v_j) = d_G(v_i, v_j)$. Similarly,

$d(v_j, v_r) < d(v_i, v_j)$. Hence $\{v_i, v_j\}$ is the heaviest edge in the triangle $\{v_i, v_j, v_r\}$ in G'. This is a contradiction to the assumption that $\{v_i, v_j\}$ belongs to the MST of G'. Hence all the vertices in $P_{i,j}$ belong to $V_i \cup V_j$, which implies the lemma. □

An edge $\{u_i, u_j\} \in E$ that satisfies the condition of Lemma 2 with respect to an edge $\{v_i, v_j\}$ of the MST T' of G' will be called the *representative edge* of $\{v_i, v_j\}$. (Since we assumed that all paths have distinct lengths, it follows that such an edge is unique.) Let $R \subseteq E$ be the set of *representative edges* of T', i.e.,

$$R = \left\{\{u_i, u_j\} \mid \{u_i, u_j\} \text{ is the } \textit{representative edge} \text{ of } \{v_i, v_j\}, \{v_i, v_j\} \in T'\right\}.$$

Define $T = \bigcup_{i=1}^{k} T_i \bigcup R$ to be the spanning tree that our algorithm returns.[5] In other words, the tree T contains the forest $T_1, T_2, \ldots, T_k$ of shortest path trees for $V_1, V_2, \ldots, V_k$ rooted at the terminals $v_1, v_2, \ldots, v_k$ respectively, and also the set R of representative edges of the MST T' of G' that connect between the connected components $V_1, V_2, \ldots, V_k$.

The next lemma shows that for every edge $\{v_i, v_j\}$ of the MST T' of G', the tree T preserves the distance between v_i and v_j.

Lemma 3. If $\{v_i, v_j\} \in T'$ then $d_T(v_i, v_j) = d_G(v_i, v_j)$.

Proof. Let $\{u_i, u_j\}$ be the representative edge of the edge $\{v_i, v_j\}$. Since $V_i = \{u \mid d_G(v_i, u) < d_G(v_j, u), \text{ for all } i \neq j\}$, it follows that for every $u \in V_i$, $d_{G(V_i)}(v_i, u) = d_G(v_i, u)$. Since T_i is an SPT of $G(V_i)$ rooted at v_i and $u_i \in V_i$, it follows that $d_{T_i}(v_i, u_i) = d_{G(V_i)}(v_i, u_i) = d_G(v_i, u_i)$. Similarly, $d_{T_j}(v_j, u_j) = d_G(v_j, u_j)$. Since $T_i, T_j \subseteq T$ and $\{u_i, u_j\} \in R \subseteq T$, it follows that

$$d_T(v_i, v_j) \leq d_T(v_i, u_i) + d_T(u_i, u_j) + d_T(u_j, v_j) = d_G(v_i, u_i) + d_G(u_i, u_j) + d_G(u_j, v_j) = d_G(v_i, v_j).$$

(The last inequality holds because, by definition of the representative edge, the edge $\{u_i, u_j\}$ belongs to the shortest path $P_{i,j}$ connecting the terminals v_i and v_j in G.) □

The next lemma shows that the tree T approximates distances between pairs of terminals up to a factor of $k-1$.

Lemma 4. For every pair of terminals $v_i, v_j \in V'$, $d_T(v_i, v_j) \leq (k-1) \cdot d_G(v_i, v_j)$.

Proof. Since $G' = \left(V', \binom{V'}{2}, w'\right)$ is a connected k-vertex graph, there exists a path π with at most $k-1$ edges from v_i to v_j in the MST T' of G'. Denote $\pi = \left(v_i = v^1 v^2 \ldots v^r = v_j\right)$, $r \leq k$. By Lemma 3, for every index s, $s \in [r-1]$, $d_{T'}\left(v^s, v^{s+1}\right) = d_G\left(v^s, v^{s+1}\right) \leq d_G(v_i, v_j)$ (because edges of π belong to the MST T'). Hence

$$d_T(v_i, v_j) \leq \sum_{s=1}^{r-1} d_T\left(v^s, v^{s+1}\right) \leq \sum_{s=1}^{r-1} d_G(v_i, v_j) = (r-1) \cdot d_G(v_i, v_j) \leq (k-1) \cdot d_G(v_i, v_j).$$

□

[5] To verify that T is a spanning tree of G note that $\{T_1, T_2, \ldots, T_k\}$ is a spanning forest of G. Also, T is acyclic as otherwise there would have been a cycle either in one of the trees T_i or in the MST T' of G'.

Finally, consider a terminal $v_i \in V'$ and a vertex $u \in V_j$, for some pair of indexes i and j. Observe that $d_T(v_j, u) \le d_T(v_i, u)$, and the inequality is strict for $i \ne j$. Hence

$$\begin{aligned} d_T(v_i, u) &\le d_T(v_i, v_j) + d_T(v_j, u) \le (k-1) \cdot d_G(v_i, v_j) + d_{T_j}(v_j, u) \\ &\le (k-1) \cdot (d_G(v_i, u) + d_G(u, v_j)) + d_G(v_j, u) \le (2k-1) \cdot d_G(v_i, u) . \end{aligned}$$

We proved the following theorem:

Theorem 6. *For a weighted graph $G = (V, E, w)$ with a non-negative weight function $w : E \to \mathbb{R}^+$, and a subset of terminals $V' \subseteq V$ of size k, there exists a spanning tree T with terminal stretch at most $2k - 1$.*

We will show in Section 3.4.1 that such a spanning tree T can be computed within $O(m + n \cdot \log n)$ time.

3.2 Light k-terminal trees for metric spaces

Next we devise a construction of *light* spanning trees with a small terminal stretch. We start with the case of metric spaces. Later we extend our results to general graphs. A finite metric space (X, d) can be viewed as a complete graph $G = \left(V, \binom{V}{2}, w\right)$. Here w is assumed to satisfy the triangle inequality. We are also given a set of terminals $V' \subset V$ of size k. We write $V'=\{v_1, v_1, \ldots, v_k\}$. The algorithm builds a tree with terminal stretch $((k-1) + k\alpha)$, for $\alpha > 1$, with lightness bounded by $3 + \frac{2}{\alpha - 1}$.

The algorithm starts by building a forest F which is an $\left(\alpha, 1 + \frac{2}{\alpha-1}\right)$-SLT from the vertex set V' (see the introduction). The forest F satisfies that for every vertex $u \in V$, it holds that $\frac{d_F(V',u)}{d_G(V',u)} \le \alpha$ and $\Psi(F) = \frac{w(F)}{w(MST_G)} \le 1 + \frac{2}{\alpha-1}$. Also, no two terminals belong to the same connected component of F. Let V_i be the connected component of v_i and $T_i \subseteq F$ be the edges of the forest F induced by V_i. (Observe that T_i is a spanning tree for V_i.) It follows that for every $u \in V_i,\ d_F(V', u) = d_{T_i}(v_i, u) \le \alpha \cdot d_G(V', u)$.

Now, let T' be the MST of $G(V')$. ($G(V')$ is the subgraph of G induced by V'.) Define $T = F \cup T' = \bigcup_{i=1}^{k} T_i \cup T'$. T is a spanning tree of G. Since T' is a minimum spanning tree for the $k-$point metric $G(V')$, it follows that for any pair of terminals $v_i, v_j \in V'$, $d_T(v_i, v_j) = d_{T'}(v_i, v_j) \le (k-1) \cdot d_G(v_i, v_j)$. Hence for any terminal $v_i \in V'$ and any vertex $u \in V_j$, it holds that

$$\begin{aligned} d_T(v_i, u) &\le d_T(v_i, v_j) + d_T(v_j, u) = d_{T'}(v_i, v_j) + d_{T_j}(v_j, u) \\ &\le (k-1) \cdot d_G(v_i, v_j) + \alpha \cdot d_G\left(V', u\right) \le (k-1) \cdot (d_G(v_i, u) + d_G(v_j, u)) + \alpha \cdot d_G(v_i, u) \\ &\underset{(1)}{\le} (k-1) \cdot (d_G(v_i, u) + \alpha \cdot d_G(v_i, u)) + \alpha \cdot d_G(v_i, u) = ((k-1) + k\alpha) \cdot d_G(v_i, u) \end{aligned}$$

(The inequality (1) holds because $d_G(v_j, u) \le d_F(v_j, u) = d_F(V', u) \le \alpha \cdot d_G(V', u) \le \alpha \cdot d_G(v_i, u)$. The first inequality is because F is a subgraph of G; the second equality is because $u \in V_j$; the third follows from the fact that F is $\left(\alpha, 1 + \frac{2}{\alpha-1}\right)$-SLT with respect to V', and the last inequality is because $v_i \in V'$.)

Next we argue that T is a light tree. Observe also that while T' is the MST of $G(V')$, the $MST\ \mathcal{T}$ of G is also a Steiner tree for V'. Thus $w(\mathcal{T}) \ge w(SMT(G(V')))$, where $SMT(G(V'))$

is the minimum Steiner tree of $G(V')$. Since $w(SMT(G(V'))) \geq \frac{1}{2}w(MST(G(V'))) = \frac{1}{2}w(T')$ (see the introduction), it follows that $w(\mathcal{T}) \geq \frac{1}{2}w(T')$, i.e., $w(T') \leq 2w(\mathcal{T})$. Hence $w(T) = w(T') + w(F) \leq \left(3 + \frac{2}{\alpha-1}\right) \cdot w(\mathcal{T})$. We conclude:

Theorem 7. *For any finite metric space (X, d), subset of terminals $X' \subseteq X$ of size k, and a parameter $\alpha \geq 1$, there exists a spanning tree T with terminal stretch at most $(k-1) + k\alpha$ and lightness at most $3 + \frac{2}{\alpha-1}$.*

Observe that for $\alpha = 1 + \epsilon$, $\epsilon > 0$ is a small parameter, we get terminal stretch $2k - 1 + k\epsilon$ with lightness $3 + \frac{2}{\epsilon}$. By setting $\epsilon' = k\epsilon$ we get terminal stretch $2k - 1 + \epsilon'$ and lightness $3 + \frac{2k}{\epsilon'}$. Note also that by setting $\alpha = 1$ we can get terminal stretch $2k - 1$ (with no guarantee on the lightness).

We will show in Section 3.4.2 that a spanning tree T that satisfies the assertion of Theorem 7 can be computed in $O(n^2)$ time.

3.3 Light k-terminal trees for general graphs

In this section we devise a construction of light k-terminal trees for general graphs. By doing so we extend Theorem 7 from metric spaces to general graphs. Consider a weighted graph $G = (V, E, w)$, $w : E \to \mathbb{R}^+$, a subset of terminals $V' \subseteq V$ of size k, and a parameter $\alpha \geq 1$. Our algorithm builds a spanning tree with terminal stretch $k\alpha + (k-1)\alpha^2$ and lightness $2\alpha + 1 + \frac{2}{\alpha-1}$. Qualitatively this tradeoff is similar to the tradeoff of Theorem 7. (The latter tradeoff applies only to metric spaces.) Specifically, for small $\epsilon > 0$, we get terminal stretch $2k - 1 + \epsilon$ and lightness $3 + \frac{6k}{\epsilon}$.

Similarly to the construction from Section 3.2, our more general construction starts by building a forest F (an $\left(\alpha, 1 + \frac{2}{\alpha-1}\right)$-SLT rooted at V') such that for every $u \in V$, $\frac{d_F(V',u)}{d_G(V',u)} \leq \alpha$ and $\Psi(F) \leq 1 + \frac{2}{\alpha-1}$. The vertex set V_i is defined to be the connected component of v_i in F and $T_i \subseteq F$ is the spanning tree of V_i. We also refer to V_i as the *component of* v_i. Recall that for every $u \in V_i$, $d_F(V', u) = d_{T_i}(v_i, u) \leq \alpha \cdot d_G(V', u)$. However, while in the metric case the ultimate tree T was the union of F with the MST of G', it is no longer the case in the current construction. Let $G' = (V', E', w')$ be the super-graph in which two terminals share an edge between them if and only if there is an edge between the components V_i to V_j in G. Formally, $E' = \{(v_i, v_j) : \exists u_i \in V_i, u_j \in V_j \text{ such that } \{u_i, u_j\} \in E\}$. The weight $w'(v_i, v_j)$ is defined to be the length of the shortest path between v_i and v_j which uses *exactly one edge* that does not belong to F. (In other words, among all the paths between v_i and v_j in G which use exactly one edge that does not belong to F, let P be the shortest one. Then $w'(v_i, v_j) = w(P)$.) Note also that $w'(v_i, v_j)$ is given by $w'(v_i, v_j) = \min_{e \in E}\left\{d_{F \cup \{e\}}(v_i, v_j)\right\}$. We call to the edge $e_{i,j} = \{u_i, u_j\}$ for which this minimum is achieved $(w'(v_i, v_j) = d_{F \cup \{e_{i,j}\}}(v_i, v_j))$ the *representative edge* of $\{v_i, v_j\}$. (By the assumption that different paths have different lengths, the representative edge is unique.) Observe that $\{v_i, v_j\} \in E'$ implies that $w'(v_i, v_j) < \infty$. Let T' be the MST of G'. Define $R = \left\{e_{i,j} \mid e_{i,j} \text{ is the representative edge of } e'_{i,j} = (v_i, v_j) \in T'\right\}$. Finally, set $T = F \cup R = \bigcup_{i=1}^{k} T_i \cup R$. Obviously, T is a spanning tree of G. Next, we show that T has bounded terminal stretch and lightness.

The next lemma shows that for every pair of terminals $v_i, v_j \in V'$, there is a path between them in G' in which all edges have weight (with respect to w') at most $\alpha \cdot d_G(v_i, v_j)$.

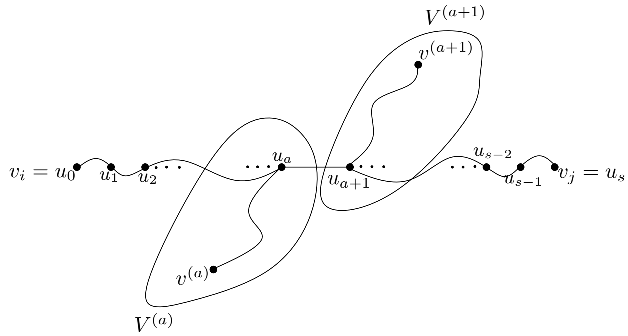


Figure 2: An illustration for the bottleneck lemma: v_i and v_j are terminals. The edge $\{u_a, u_{a+1}\}$ belongs to the shortest path from v_i to v_j in G. We conclude that for terminals $v^{(a)}, v^{(a+1)}$ such that $u_a \in V^{(a)}$ and $u_{a+1} \in V^{(a+1)}$ it holds that $w'\left(v^{(a)}, v^{(a+1)}\right) \leq \alpha \cdot d_G(v_i, v_j)$.

Lemma 5. ***The bottleneck lemma:*** *For every pair of terminals $v_i, v_j \in V'$, there exists a path $P : v_i = z_0 z_1 ... z_r = v_j$ in G' such that for every index $s = 0, 1, \ldots, r-1$, it holds that $(z_s, z_{s+1}) \in E'$ and $w'(z_s, z_{s+1}) \leq \alpha \cdot d_G(v_i, v_j)$. (See Figure 2 for illustration.)*

Proof. Let $P_{i,j} : v_i = u_0 u_1 ... u_s = v_j$ be the shortest path from v_i to v_j in G, i.e., $w(P_{i,j}) = d_G(v_i, v_j)$. For each a, let $V^{(a)}$ be the connected component of u_a, and $v^{(a)}$ is the terminal in $V^{(a)}$. Consider the path $P = v^{(0)}, v^{(1)}, ..., v^{(s)}$. (This path is not necessarily simple. In particular, it might contain self-loops.) For every index $a < s$,

$$\begin{aligned} w'\left(v^{(a)}, v^{(a+1)}\right) &\underset{(1)}{\leq} d_{F \cup \{\{u_a, u_{a+1}\}\}}\left(v^{(a)}, v^{(a+1)}\right) = d_F\left(v^{(a)}, u_a\right) + d_G(u_a, u_{a+1}) + d_F\left(u_{a+1}, v^{(a+1)}\right) \\ &\underset{(2)}{\leq} \alpha \cdot d_G(v_i, u_a) + d_G(u_a, u_{a+1}) + \alpha \cdot d_G(v_j, u_{a+1}) \\ &< \alpha \cdot (d_G(v_i, u_a) + d_G(u_a, u_{a+1}) + d_G(v_j, u_{a+1})) \underset{(3)}{=} \alpha \cdot d_G(v_i, v_j). \end{aligned}$$

Note that if for some index a it holds that $v^{(a)} = v^{(a+1)}$ then $w'\left(v^{(a)}, v^{(a+1)}\right) = 0$, and the inequality above holds trivially. Otherwise, (if $v^{(a)} \neq v^{(a+1)}$) inequality (1) follows by the definition of w' and by the assumptions that $\{u_a, u_{a+1}\} \in E$, $u_a \in V^{(a)}$, $u_{a+1} \in V^{(a+1)}$. Inequality (2) follows from the properties of the SLT tree T. (Specifically $d_F\left(v^{(a)}, u_a\right) = d_F(V', u_a) \leq \alpha \cdot d_G(V', u_a) \leq \alpha \cdot d_G(v_i, u_a)$.) Equality (3) follows because the edge $\{u_a, u_{a+1}\}$ is on the shortest path from v_i to v_j in G.

In particular, one can remove cycles from P and obtain a simple path with the desired properties. We get a simple path $P' = v^{(0)}, v^{(1)} \ldots, v^{(r)}$ such that for every index $q < r$ we have $w'\left(v^{(q)}, v^{(q+1)}\right) \leq \alpha \cdot d_G(v_i, v_j)$, as required. □

Recall that T' is an MST of G', and $T = F \cup R$ is the tree constructed by our algorithm. The next corollary follow from Lemma 5.

Corollary 6. *For $\{v_i, v_j\} \in T'$, we have $w'(v_i, v_j) = d_T(v_i, v_j) \leq \alpha \cdot d_G(v_i, v_j)$.*

Proof. By Lemma 5, $w'(v_i, v_j) \leq \alpha \cdot d_G(v_i, v_j)$. (Indeed, otherwise the edge $\{v_i, v_j\}$ is the strictly heaviest edge in a cycle in G', contradicting the assumption that it belongs to the MST of G'.) Since $\{v_i, v_j\} \in E'$ and the representative edge of $\{v_i, v_j\}$ was taken into T, it follows that $w'(v_i, v_j) = d_T(v_i, v_j)$. □

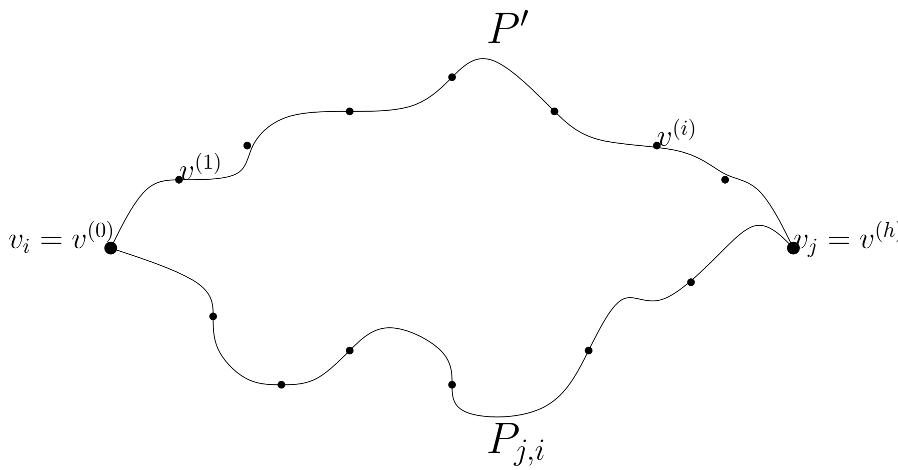


Figure 3: The two paths P' and $P_{j,i}$ considered in the proof of Lemma 7. The path P' is contained in T', while all edges of $P_{j,i}$ have weight at most $\alpha \cdot d_G(v_i, v_j)$.

Next we consider a general pair of terminals $v_i, v_j \in V'$.

Lemma 7. *For $v_i, v_j \in V'$, we have $d_T(v_i, v_j) \le d_{T'}(v_i, v_j) \le \alpha \cdot (k-1) \cdot d_G(v_i, v_j)$.*

Proof. Let $P' : v_i = v^{(0)}v^{(1)} \dots v^{(h)} = v_j$ be the (unique) path in T' between v_i and v_j. Since T' is a spanning tree of the k-vertex graph G', it follows that $h \le k-1$. Observe also that for every index $a \in [h-1]$, by Corollary 6 we have $w'\left(v^{(a)}, v^{(a+1)}\right) = d_T\left(v^{(a)}, v^{(a+1)}\right)$. Also, we next argue that $w'\left(v^{(a)}, v^{(a+1)}\right) \le \alpha \cdot d_G(v_i, v_j)$. Indeed, suppose for contradiction that $w'\left(v^{(a)}, v^{(a+1)}\right) > \alpha \cdot d_G(v_i, v_j)$. Let $P_{j,i}$ be a path between v_j and v_i in G' such that all its edges have weight at most $\alpha \cdot d_G(v_i, v_j)$. The existence of this path is guaranteed by Lemma 5. In particular, since $w'\left(v^{(a)}, v^{(a+1)}\right) > \alpha \cdot d_G(v_i, v_j)$, it follows that $\left\{v^{(a)}, v^{(a+1)}\right\} \notin P_{j,i}$. Consider the cycle $P' \circ P_{j,i}$ in G'. It is not necessarily a simple cycle, but since $\left\{v^{(a)}, v^{(a+1)}\right\} \notin P_{j,i}$, the edge $\left\{v^{(a)}, v^{(a+1)}\right\}$ belongs to a simple cycle C contained in $P' \circ P_{j,i}$. The heaviest edge of C clearly does not belong to $P_{j,i}$, because the edge $\left\{v^{(a)}, v^{(a+1)}\right\}$ is heavier that each of the edges in $P_{j,i}$. Hence the heaviest edge belongs to P', but $P' \subseteq T'$. This is a contradiction to the assumption that T' is an MST of G'. (See Figure 3 for an illustration.) Hence $d_T\left(v^{(a)}, v^{(a+1)}\right) = w'\left(v^{(a)}, v^{(a+1)}\right) \le \alpha \cdot d_G(v_i, v_j)$. Finally,

$$\begin{aligned} d_T(v_i, v_j) &\le \sum_{a=0}^{h-1} d_T\left(v^{(a)}, v^{(a+1)}\right) = \sum_{a=0}^{h-1} w'\left(v^{(a)}, v^{(a+1)}\right) \\ &\le \sum_{a=0}^{h-1} \alpha \cdot d_G(v_i, v_j) \le h \cdot \alpha \cdot d_G(v_i, v_j) \le \alpha \cdot (k-1) \cdot d_G(v_i, v_j). \end{aligned}$$

□

Next, we show that T provides a relatively small stretch for every terminal vertex pair (v', u), $v' \in V', u \in V$.

Lemma 8. *$TermStretch_{G,V'}(T) \le k \cdot \alpha + (k-1)\alpha^2$.*

Proof. Consider a terminal $v_i \in V'$ and a vertex $u \in V_j$. Observe that $d_T(v_i, u) \leq d_T(v_i, v_j) + d_T(v_j, u)$. By Lemma 7, $d_T(v_i, v_j) \leq \alpha \cdot (k-1) \cdot d_G(v_i, v_j)$. Also, since $F \subseteq T$, and F is an $\left(\alpha, 1 + \frac{2}{\alpha-1}\right) - SLT$ rooted at V', it follows that $d_T(v_j, u) \leq \alpha \cdot d_G(v_i, u)$. Hence

$$\begin{aligned} d_T(v_i, u) &\leq d_T(v_i, v_j) + d_T(v_j, u) \leq \alpha \cdot (k-1) \cdot d_G(v_i, v_j) + \alpha \cdot d_G(v_i, u) \\ &\leq \alpha \cdot (k-1) \cdot (d_G(v_i, u) + d_G(u, v_j)) + \alpha \cdot d_G(v_i, u) \\ &\leq \alpha \cdot (k-1) \cdot (d_G(v_i, u) + \alpha \cdot d_G(v_i, u)) + \alpha \cdot d_G(v_i, u) = \left(k \cdot \alpha + (k-1)\alpha^2\right) \cdot d_G(v_i, u). \end{aligned}$$

(The last inequality is because $d_G(v_j, u) \leq d_F(v_j, u) = d_F(V', u) \leq \alpha \cdot d_G(V', u) \leq \alpha \cdot d_G(v_i, u)$.) Hence $TermStretch_{G,V'}(T) = max_{v \in V', u \in V}\left\{\frac{d_T(v,u)}{d_G(v,u)}\right\} \leq k \cdot \alpha + (k-1)\alpha^2$. □

Finally, in the next lemma we analyze the lightness of the tree T.

Lemma 9. *The lightness of T is bounded by $\Psi(T) \leq 2\alpha + 1 + \frac{2}{\alpha-1}$.*

Proof. The main challenge is to bound $w(R)$. (Recall that R is the set of the *representative edges* of T'.) Consider an edge $\{v_i, v_j\} \in T'$, and let $\{u_i, u_j\}$ be its representative edge. Then $d_G(u_i, u_j) \leq w'(v_i, v_j)$. Also, since $\{v_i, v_j\} \in T' \subseteq E'$, it follows that $w'(v_i, v_j) = d_{G'}(v_i, v_j)$. Hence $d_G(u_i, u_j) \leq d_{G'}(v_i, v_j)$. Therefore, $w(R) \leq w'(T')$. Next we provide an upper bound for the weight $w'(T')$ of the MST T' of G'. Define the graph $\widetilde{G} = \left(V', \binom{V'}{2}, \tilde{w} = d_G|_{V'}\right)$, where $d_G|_{V'}$ is the shortest path metric in G restricted to V'. Let $\widetilde{T}$ be the MST of $\widetilde{G}$. We build a new tree $\hat{T}$ by the following algorithm:

Algorithm 1

---

1. $\hat{T} \leftarrow \widetilde{T}$
2. for $\{v_i, v_j\} = \tilde{e} \in \tilde{T}$
   (a) Let $P_{\tilde{e}}$ be a path from v_i to v_j which consists of edges in E' such that for each edge e in $P_{\tilde{e}}$, $w'(e) \leq \alpha \cdot d_G(v_i, v_j) = \alpha \cdot \tilde{w}(\tilde{e})$. (By Lemma 5, such a path exists.)
   (b) Let $e' \in P_{\tilde{e}}$ be an edge such that $\left(\hat{T} \setminus \{\tilde{e}\}\right) \cup \{e'\}$ is connected.
   (c) $\hat{T} \leftarrow \left(\hat{T} \setminus \{\tilde{e}\}\right) \cup \{e'\}$.

In each step in the loop we replace an edge $\tilde{e} = \{v_i, v_j\}$ from $\widetilde{T}$ by an edge e' from G' of weight $w'(e) \leq \alpha \cdot \tilde{w}(\tilde{e})$. Hence the resulting tree $\hat{T}$ is a spanning tree of G', and $w'\left(\hat{T}\right) \leq \alpha \cdot \tilde{w}\left(\widetilde{T}\right)$. Since T' is the MST of G', it follows that $w'(T') \leq w'\left(\hat{T}\right)$. The MST of G is a Steiner tree for $\tilde{G}$. Hence $\tilde{w}\left(SMT\left(\tilde{G}\right)\right) \leq w(MST(G))$. Also, (by [GP68], sec. 10), $\tilde{w}\left(MST\left(\tilde{G}\right)\right) = \tilde{w}\left(\tilde{T}\right) \leq 2 \cdot \tilde{w}\left(SMT\left(\tilde{G}\right)\right) \leq 2 \cdot w(MST(G))$. Hence

$$w(R) \leq w'(T') \leq w'\left(\hat{T}\right) \leq \alpha \cdot \tilde{w}\left(\widetilde{T}\right) \leq 2 \cdot \alpha \cdot w(MST(G)).$$

Since $w(F) \leq \left(1 + \frac{2}{\alpha-1}\right) \cdot w(MST(G))$, we conclude that

$$w(T) = w(R \cup F) = w(R) + w(F) \leq \left(2\alpha + 1 + \frac{2}{\alpha - 1}\right) \cdot w(MST(G)).$$

□

Lemma 8 and Lemma 9 imply the following theorem:

Theorem 8. *For a weighted graph $G = (V, E, w)$, $w : E \to \mathbb{R}_+$, and a subset of terminals $V' \subseteq V$ of size k, there exists a spanning tree T of G with terminal stretch $k \cdot \alpha + (k-1)\alpha^2$ and lightness $2\alpha + 1 + \frac{2}{\alpha-1}$.*

Remark 4. *The bound $2\alpha + 1 + \frac{2}{\alpha-1}$ is minimized by setting $\alpha = 2$. So there is no point in using the algorithm with $\alpha > 2$, as for such values of α both the terminal-stretch and lightness decrease by choosing the parameter $\alpha = 2$ (instead of $\alpha > 2$). Hence if one allows terminal-stretch at least $6k - 4$, Theorem 8 provides a tree with lightness at most 7 and this stretch.*

Remark 5. *Note that when substituting $\alpha = 1$ in Theorem 8 we obtain a $k-$terminal tree with stretch exactly $2k - 1$. Hence Theorem 8 generalizes Theorem 6.*

Remark 6. *We will argue in Section 3.4.3 that a spanning tree T as above can be constructed in $O(m + n \log n)$ time.*

Remark 7. *In the regime $\alpha = 1+\epsilon$, for a small parameter $\epsilon > 0$, we obtain stretch $2k-1+O(\epsilon \cdot k) = (2k-1)(1 + O(\epsilon))$ and lightness $O\left(\frac{1}{\epsilon}\right)$.*

3.4 Implementations

In this section we present implementations of the algorithms presented in Sections 3.1, 3.2 and 3.3. In each implementation the description will use definitions from the respective algorithm. Throughout n is the number of vertices, m is the number of edges and k is the number of terminals.

3.4.1 Implementation of the k-terminal-tree algorithm

In this section we describe an implementation for the algorithm from Section 3.1, which constructs a k-terminal tree with terminal stretch at most $2k-1$ for an arbitrary weighted graph $G = (V, E, w)$ and a set $V' = \{v_1, v_2, \ldots, v_k\}$ of terminals.

Recall that V_i is the subset of vertices which are closer to v_i than to any other terminal, and T_i is the shortest path tree of V_i rooted at v_i. Now define a super-graph $G' = \left(V', \binom{V'}{2}, w'\right)$, where $w'(v_i, v_j) = d_G(v_i, v_j)$. Let T' be the MST of G'. An edge $\{u_i, u_j\} \in E$ with $u_i \in V_i$, $u_j \in V_j$ and such that the shortest path from v_i to v_j in G goes through $\{u_i, u_j\}$ is called the representative edge of $\{v_i, v_j\}$. By Lemma 2, each edge in T' has a representative edge. The set R is defined as the set of all the representative edges of T'. The algorithm returns the tree $T = \bigcup_{i=1}^{k} T_i \bigcup R$.

The implementation differs from the algorithm described above as it uses a different auxiliary graph G'' instead of G'. G'' can be computed faster than G', while G'' and G' will have the same MST and hence the same set of representative edges R. Therefore the resulting tree will be the same. We start by adding a new vertex x, and connecting it to all the terminals with edges of zero weight. Next, we run Dijkstra's algorithm ([CLRS09] chapter 24) from x and remove all the edges which contain x from the resulting tree. As a result we obtain a forest $F = \{T_1, T_2, \ldots, T_k\}$ of k shortest paths trees rooted at the terminals $v_1, v_2, \ldots, v_k$, respectively. For each index $i \in [k]$, T_i is a shortest path tree in G rooted at v_i for the set V_i. The running time of this step is the time required for invoking Dijkstra's algorithm in a graph with $n+1$ vertices and $m+k$ edges, i.e., $O(m + n \log n)$. At this point the algorithm knows for every vertex $u \in V$ the index $i \in [k]$ such that $u \in V_i$, and the distance $d_G(v_i, u)$ between u and its closest terminal v_i. We construct the auxiliary graph $G'' = \left(V', \binom{V'}{2}, w''\right)$ by setting $w''(v_i, v_j) = min_{e \in E}\left\{d_{T_i \cup T_j \cup \{e\}}(v_i, v_j)\right\}$. (There may be edges of infinite weight.) To compute G'' we initialize all weights as ∞. Then for each edge $e = \{u, x\}$ we check if it connects two different components (V_i, V_j). If it is the case then we check if $d_{T_i}(v_i, u) + w(u, x) + d_{T_j}(x, v_j)$ is smaller then the currently recorded weight $w''(v_i, v_j)$. If this condition holds too then we set $w''(v_i, v_j) = d_{T_i}(v_i, u) + w(u, x) + d_{T_j}(x, v_j)$, and the edge e is recorded as the representative edge for the super-edge $\{v_i, v_j\}$. Hence computing G'' requires $O(m)$ time. At this point the algorithm knows for each super-edge $\{v_i, v_j\}$ of finite weight, its representative edge. Now we compute the MST T'' of G''. This step requires $O(k \log k + m)$ time. (Each edge in G'' has a unique representative edge. Hence there are at most m edges in G''.) For every pair $\tilde{e} = \{v_i, v_j\} \in \binom{V'}{2}$ of terminals, $w''(v_i, v_j) = min_{e \in E}\left\{d_{T_i \cup T_j \cup \{e\}}(v_i, v_j)\right\} \geq d_G(v_i, v_j) = w'(v_i, v_j)$. By Lemma 2 for each edge $\tilde{e} = \{v_i, v_j\}$ of the MST of G' it holds that $w''(v_i, v_j) = w'(v_i, v_j)$. Hence G' and G'' have the same MST $T' = T''$ with the same representative edges set R. Hence we return $T = F \cup R$. The overall running time is $O(m + n \log n)$.

3.4.2 Implementation of the algorithm for metric spaces

In this section we describe an implementation for the algorithm from Section 3.2, which given a finite metric space (X, d) with subset $X' \subseteq X$ of size k and a parameter $\alpha \geq 1$ constructs a spanning tree T with terminal stretch at most $(k-1) + k\alpha$ and lightness at most $3 + \frac{2}{\alpha - 1}$. Recall that the algorithm starts with computing an $\left(\alpha, 1 + \frac{2}{\alpha-1}\right)$-SLT F with respect to the set X'. Then it computes the MST T' of $G' = \left(V', \binom{V'}{2}, d\right)$ and returns $F \cup T'$.

The implementation is straightforward. Computing the SLT on a complete graph requires $O\left(n^2\right)$ time. (See [KRY93].) The computation of the MST requires $O\left(k^2\right)$ time. Hence the total running time is $O\left(n^2\right)$.

3.4.3 Implementation of algorithm for general graphs

In this section we describe an implementation for the algorithm from Section 3.3, which given a weighted graph $G = (V, E, w)$, a subset $V' \subseteq V$ of terminals $\{v_1, v_2, \ldots, v_k\}$ and a parameter $\alpha \geq 1$ constructs a spanning tree T with terminal stretch at most $k \cdot \alpha + (k-1)\alpha^2$ and lightness $2\alpha + 1 + \frac{2}{\alpha-1}$. Recall that the algorithm starts with computing an $\left(\alpha, 1 + \frac{2}{\alpha-1}\right)$-SLT F with respect to V'. For each $i \in [k]$, let V_i denote the connected component of v_i in F, and let $T_i \subseteq F$ stand for $F[V_i]$, i.e., the set of edges of the forest F induced by the vertex set V_i. The algorithm constructs the graph $G' = \left(V', \binom{V'}{2}, w'\right)$ by setting $w'(v_i, v_j) = min_{e \in E}\left\{d_{T_i \cup T_j \cup \{e\}}(v_i, v_j)\right\}$. An edge $e = \{u_i, u_j\} \in E$ with $u_i \in V_i$, $u_j \in V_j$ and such that $w'(v_i, v_j) = d_{T_i \cup T_j \cup \{e\}}(v_i, v_j)$ is called the representative edge of $\{v_i, v_j\}$. Then the algorithm computes the MST T' of G'. Let R be the set of all the representative edges of T'. The algorithm returns the tree $T = F \bigcup R$.

The implementation starts by invoking the $\left(\alpha, 1 + \frac{2}{\alpha - 1}\right)$-SLT algorithm with respect to the set of terminals. This step requires $O(m + n\log n)$ time. The set R is computed in the same way as described in Section 3.4.1. (Note that the auxiliary graph described there is the actual graph used in the current algorithm.) This step requires $O(m + n\log n)$ time. Hence the total time is $O(m + n\log n)$.

4 Lower bounds

In this section we provide lower bounds on the terminal stretch and lightness of k-terminal trees. For each lower bound we start with showing a lower bound for graphs and proceed to showing a lower bound for metric spaces. (Observe that the latter is more general.) The lower bounds exhibit similar tradeoffs, while the analysis is significantly simpler for graphs than for metric spaces.

4.1 A lower bound for terminal stretch

The following simple lower bound was mentioned in the beginning of Section 3.1.

Theorem 9. *For any k there is a weighted graph G with $n = 2k$ vertices and k terminals such that any spanning tree has terminal stretch at least $2k - 1$.*

Proof. Consider the cycle graph C_{2k} such that the terminals and the non-terminal vertices alternate (As usual V' stands for the set of terminals.) Any spanning tree T of C_{2k} is obtained by removing a single edge $\{v_i, v_{i+1}\}$. Observe that either v_i or v_{i+1} is a terminal. Hence

$$TermStretch_{C_{2k},V'}(T) = max_{v\in V', u\in V}\frac{d_T(v,u)}{d_{C_{2k}}(v,u)} \geq \frac{d_{C_{2k}\setminus\{v_i,v_{i+1}\}}(v_i,v_{i+1})}{d_{C_{2k}}(v_i,v_{i+1})} = 2k - 1.$$

□

Remark 8. *The same result for any $n > 2k$ follows by adding additional $n - 2k$ non-terminal vertices to the graph, and connecting them all to an arbitrary vertex of C_{2k}. In the resulting graph every spanning tree has terminal stretch at least $2k - 1$.*

Next we extend Theorem 9 to metric spaces.

Theorem 10. *For any k there is a metric space (M, d) with $n = 2k$ vertices and k terminals such that any spanning tree for (M, d) has terminal stretch at least $2k - 1$.*

Proof. Let M be the metric generated by the cycle graph C_{2k}, where there are k terminals and the terminals and the non-terminal vertices alternate. In [Gup01], Lemma 7.1, Gupta showed that the stretch of every spanning tree of M is at least $n - 1 = 2k - 1$. Moreover, the maximal stretch is achieved by an original edge e of C_{2k}. One of the two endpoints of e is a terminal. Hence the terminal stretch of any spanning tree for M is at least $2k - 1$. □

Remark 9. *One can extend Theorem 10 to $n > 2k$ by adding additional $n - 2k$ non-terminal vertices to M at a large distance from all the terminals. For a spanning tree T of M, if some shortest path between a terminal v and a vertex u which is belongs to C_{2k} uses at least one of the new vertices then the terminal stretch is too high. Hence if the terminal stretch is small, the tree restricted to C_{2k} is connected. Hence by Theorem 10 it follows that the terminal stretch is at least $2k - 1$.*

4.2 A lower bound on the lightness in graphs

In this section we prove a lower bound on the tradeoff between terminal stretch and lightness of k-terminal trees.

Theorem 11. *For any positive integer parameters k, n and $\epsilon > 0$ such that $k \leq \frac{\epsilon}{2}n$, there exists a graph G with $N = (n+1)\,k$ vertices, such that any spanning tree T for G with terminal stretch at most $(2k-1)\left(1+\frac{\epsilon}{k^2}\right)$ has lightness at least $\Omega\left(\frac{1}{\epsilon}\right)$.*

Proof. Consider the following graph G on $N = k \cdot (n+1)$ vertices. There are k terminals $V' = \{v_0, v_1, ..., v_{k-1}\}$ in G. For every index $i \in [0, k-1]$, there is a n-vertex path P_i. All edges in these paths have unit weight. Also, for each index $i \in [0, k-1]$, both v_i and $v_{i+1 \pmod{k}}$ are connected to every vertex in P_i by edges of weight w, where $w > 1$ is a parameter that will be determined later. (To simplify the notation we will henceforward write v_{i+1} instead $v_{i+1 \pmod{k}}$. Generally, all the arithmetic operations on indexes of vertices $v_0, ..., v_{k-1}$ and paths $P_0, ..., P_{k-1}$ are performed modulo k.) See Figure 4 for an illustration.

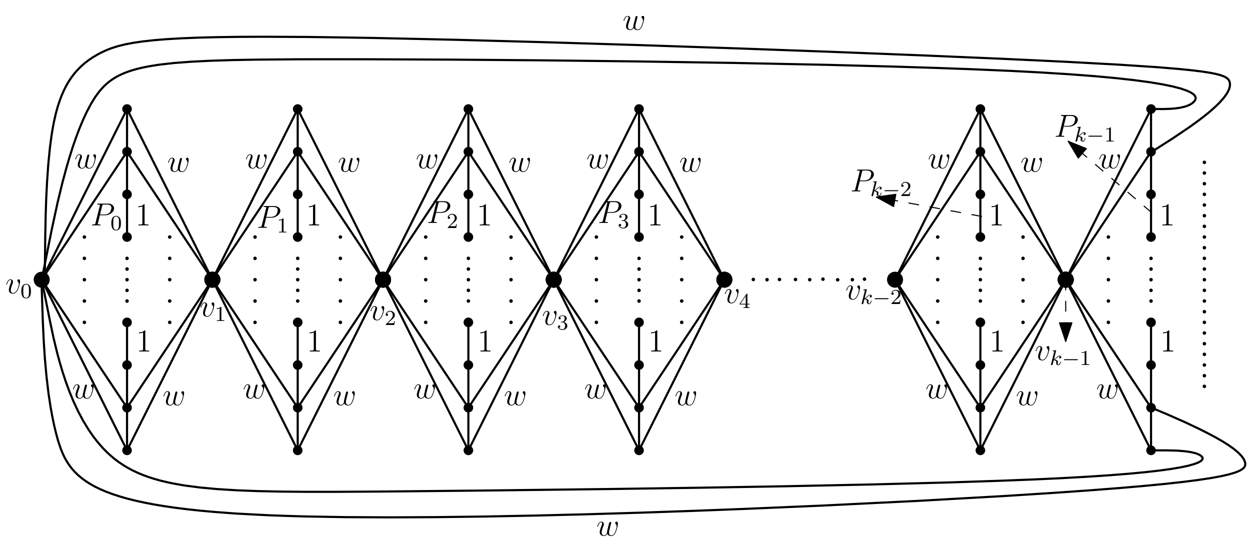


Figure 4: An illustration of the graph used in the proof of Theorem 11. The k terminals are depicted by the big dots. The vertices v_i and v_{i+1} are connected to each vertex of an n-vertex path P_i by edges of weight w.

Each spanning tree of G contains $N - 1 = k + kn - 1$ edges. There are $k \cdot (n-1)$ edges of unit weight, and all the other edges have weight w. Hence the weight of the MST is at least $k\,(n-1)\cdot 1 + (2k-1)\cdot w$. It is easy to verify that there actually exists a spanning tree of that weight. We will show that for any $\beta < \frac{1}{w(2k-1)}$, every tree with terminal stretch at most $(2k-1)\,(1+\beta)$ has weight at least $\Omega(nw)$.

Let T be a spanning tree for G with terminal stretch at most $(2k-1)\,(1+\beta)$, for some $\beta > 0$. There exists an index i such that the path between v_i and v_{i+1} in T does not use vertices from the path P_i. (Otherwise there is a cycle in T that passes through $v_0, \ldots, v_{k-1}$.) Without loss of generality assume that $i = 0$. Therefore, $d_T\,(v_0, v_1) \geq (k-1)\cdot 2w$.

Claim 10. *For every vertex u in P_0, if $\beta < \frac{1}{(2k-1)w}$ then either (v_0, u) or (v_1, u) is an edge of T.*

Proof. Without loss of generality the shortest path from v_0 to u goes through v_1. Assume for contradiction that the edge $(v_{1,}u)$ does not belong to T. Then

$$\begin{aligned} TermStretch_{G,V'}(T) &\geq \frac{d_T(v_0,u)}{d_G(v_0,u)} = \frac{d_T(v_0,v_1) + d_T(v_1,u)}{d_G(v_0,u)} \geq \frac{(k-1)\,2w + w + 1}{w} \\ &= 2k - 1 + \frac{1}{w} = (2k-1)\left(1 + \frac{1}{(2k-1)\,w}\right) > (2k-1)(1+\beta), \end{aligned}$$

contradiction. □

By Claim 10, $\beta < \frac{1}{(2k-1)w}$ implies that $w(T) \geq n \cdot w$. Hence for every k-terminal tree T with terminal stretch at most $(2k-1)(1+\beta)$, with $\beta < \frac{1}{(2k-1)w}$, it holds that $\Psi(T) = \frac{w(T)}{w(MST)} \geq \frac{nw}{k(n-1)+(2k-1)w}$. We set $w = \frac{k}{\epsilon}$. Then the condition $\beta < \frac{1}{(2k-1)w}$ translates to $\beta < \frac{\epsilon}{(2k-1)k}$. This condition implies that

$$\Psi(T) = \frac{n\frac{k}{\epsilon}}{k(n-1) + (2k-1)\frac{k}{\epsilon}} = \frac{nk}{\epsilon k(n-1) + (2k-1)\,k} \geq \frac{nk}{\epsilon nk + 2k^2} = \frac{n}{\epsilon n + 2k}.$$

As $k \leq \frac{\epsilon}{2}n$ we obtain $\Psi(T) \geq \frac{1}{2\epsilon}$. □

Our algorithm from Theorem 8 guarantees stretch $(2k-1)(1+O(\epsilon))$ and lightness $O\left(\frac{1}{\epsilon}\right)$. In the graph G from the above proof lightness smaller than $\frac{1}{\epsilon}$ implies stretch at least $(2k-1)\left(1+\frac{\epsilon}{(2k-1)k}\right)$. Hence our bounds are tight for $k = O(1)$, but generally there is a gap of $O\left(k^2\right)$ between the upper and lower bounds.

4.3 A lower bound on the lightness in metric spaces

The metric space case is similar to the graph case. We will use the metric closure of the graph from the previous section. The main difficulty is however to show that the non-graph edges do not help at all.

For a positive integer parameter k and a $k-$sequence $n_0, n_1, \ldots, n_{k-1}$ of positive integer numbers, we define a graph $G_{k,n_0,n_1,\ldots,n_{k-1}}$. The graph has $n = k + \sum_i n_i$ vertices and k terminals. The k terminals are $V' = \{v_0, \ldots, v_{k-1}\}$. For every index $i \in [0, k-1]$, there is an n_i-vertex path P_i. (We will use P_i to denote both the path, and the set of vertices in the path.) All edges in these paths have unit weight. Also, for each $i \in [0, k-1]$, both v_i and v_{i+1} (the index arithmetic is modulo k) are connected to every vertex in P_i by edges of weight w, for a parameter $w > 1$. Observe that the graph G from Section 4.2 satisfies $G = G_{k,n_0,n_1,\ldots,n_{k-1}}$ with $n_0 = n_1 = \cdots = n_{k-1} = n$. Let $\overline{G}_{k,n_0,\ldots,n_{k-1}}$ denote the metric closure of $G_{k,n_0,\ldots,n_{k-1}}$. We also write $G = G_{k,n_0,\ldots,n_{k-1}}$ and $\overline{G} = \overline{G}_{k,n_0,\ldots,n_{k-1}}$.

Lemma 11. *For any spanning tree T of $\overline{G}_{k,n_0,\ldots,n_{k-1}}$ there exists an index i such that for any vertex $u \in P_i$ either $\frac{d_T(v_i,u)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_i,u)} \geq 2k-1$ or $\frac{d_T(v_{i+1},u)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_i,u)} \geq 2k-1$.*

Remark 10. *Observe that for a graph spanning tree T this lemma follows directly from the observation that there exists an index i such that the path in T from v_i to v_{i+1} does not contain vertices of P_i. Indeed, for this index i and a vertex $u \in P_i$, either $(v_i, u) \notin T$ or $(v_{i+1}, u) \notin T$. In the former case $\frac{d_T(v_i,u)}{d_{G_{k,n_0,\ldots,n_{k-1}}}(v_i,u)} \geq 2k-1$ and in the latter $\frac{d_T(v_{i+1},u)}{d_{G_{k,n_0,\ldots,n_{k-1}}}(v_{i+1},u)} \geq 2k-1$. The lemma proves this statement in a much greater generality, specifically, for T being a spanning tree of the metric closure $\overline{G}_{k,n_0,\ldots,n_{k-1}}$ of $G_{k,n_0,\ldots,n_{k-1}}$.*

Proof. For a spanning tree T of $\overline{G}_{k,n_0,\ldots,n_{k-1}}$ and an index $i \in [0, k-1]$, denote by $t_{T,i}$ the number of vertices u in P_i such that $\frac{d_T(v_i,u)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_i,u)} \geq 2k-1$ or $\frac{d_T(v_{i+1},u)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_{i+1},u)} \geq 2k-1$. We will show that for any spanning tree T there exists an index i such that $t_{T,i} = n_i$. For a tree T, define $\gamma(T) = \min_i \{n_i - t_{T,i}\}$. Observe that $\gamma(T) \geq 0$. It suffices to prove that for every tree T, $\gamma(T) = 0$. Also, let $\mu = \max_T \{\gamma(T)\}$, where the maximum is taken over all spanning trees of $\overline{G}_{k,n_0,\ldots,n_{k-1}}$. It is enough to show that $\mu = 0$, i.e., that for every spanning tree T, $\gamma(T) = 0$.

For each vertex, we define the right and the left *hemisphere* with respect to this vertex. Consider a supergraph, where we replace each path P_i by a supernode p_i. We obtain the $2k-$cycle C_{2k}. The *right hemisphere* of v_i consists of all the vertices between v_i to its antipodal vertex (in the path $p_i, v_{i+1}, p_{i+1}, \ldots$), while the *left hemisphere* of v_i consists of all the vertices in the other shortest path from v_i to its antipodal vertex ($p_{i-1}, v_{i-1}, p_{i-2}, \ldots$). Formally, for a terminal v_i, if k is even, let $R(v_i) = \bigcup \left\{P_i, \{v_{i+1}\}, P_{i+1}, \ldots, \left\{v_{i+\frac{k}{2}}\right\}\right\}$ and $L(v_i) = \bigcup \left\{\left\{v_{i+\frac{k}{2}}\right\}, P_{i+\frac{k}{2}}, \left\{v_{i+\frac{k}{2}+1}\right\}, \ldots, P_{i-1}\right\}$ denote the right and the left hemispheres with respect to v_i, respectively. Similarly, if k is odd then $R(v_i) = \bigcup \left\{P_i, \{v_{i+1}\}, \ldots, P_{i+\frac{k-1}{2}}\right\}$ and $L(v_i) = \bigcup \left\{P_{i+\frac{k-1}{2}}, \left\{v_{i+\frac{k+1}{2}}\right\}, \ldots, P_{i-1}\right\}$. In addition we define hemispheres for non-terminal vertices. For an index i, all the vertices in P_i have the same hemispheres. For a vertex $u \in P_i$, if k is even, then $R(u) = \bigcup \left\{\{v_{i+1}\}, P_{i+1}, \{v_{i+2}\}, \ldots, P_{i+\frac{k}{2}}\right\}$ and $L(u) = \bigcup \left\{P_{i+\frac{k}{2}}, \left\{v_{i+\frac{k}{2}+1}\right\}, P_{i+\frac{k}{2}+1}, \ldots, \{v_i\}\right\}$. If k is odd then $R(u) = \bigcup \left\{\{v_{i+1}\}, P_{i+1}, \{v_{i+2}\}, \ldots, \left\{v_{i+\frac{k+1}{2}}\right\}\right\}$ and $L(u) = \bigcup \left\{\left\{v_{i+\frac{k+1}{2}}\right\}, P_{i+\frac{k+1}{2}}, \left\{v_{i+\frac{k+1}{2}+1}\right\}, \ldots, \{v_i\}\right\}$. (See Figure 5 for an illustration.)

For a vertex $u \in V$ and an index $i \in [0, k-1]$, we say that all the edges from u to P_i, i.e., $\{\{u, z\} \mid z \in P_i\}$, are of the *same type*. In addition, the edge from u to v_i has a unique type. (Note that each non-terminal vertex might be incident to edges of $2k$ different types, while a terminal vertex might be incident to edges of $2k-1$ different types.) An edge $e = \{u, z\}$ such that u, z are in the same path P_i called a *path-internal edge*. For illustration of these definitions, see Figure 6. For a simple path π in $\overline{G}_{k,n_0,\ldots,n_{k-1}}$, we say that π is a *one-sided* path if for any internal vertex x in π, the two edges $(x, y_1), (x, y_2)$ which are incident on x in π, connect x to different hemispheres, i.e., y_1 is in the left hemisphere with respect to x, and y_2 is in the right one, or vice versa.

The proof (that $\mu = 0$) is by induction on $\sum_{i=1}^{k} n_i = A$. The base case where $\sum_{i=1}^{k} n_i = k$ (i.e., for all $i \in [0, k-1]$, $n_i = 1$) follows by Theorem 10.
During the induction step, in order to use the induction hypothesis we will delete some vertex from $\overline{G}_{k,n_0,\ldots,n_{k-1}}$. By deleting a vertex u from a path P_i we get a new graph $\overline{G}_{k,n_0,\ldots,n_{i-1},n_i-1,n_{i+1}n_{k-1}}$. The weight of all the edges, other than edges whose both endpoints are incident in P_i, remain the same. While for vertices w_1, w_2 in P_i, if the shortest path between them in the graph without the complimentary edges goes trough u, the weight of the edges between them decreases by 1, otherwise (the shortest path does not use u), the weight of their common edge remain the same.

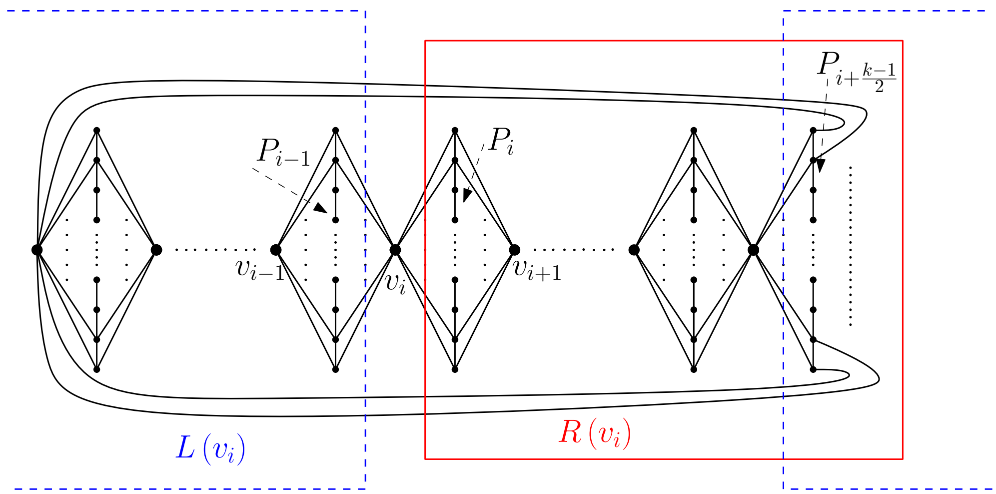


Figure 5: An illustration of the partition of the graph to the right and the left hemispheres with respect to a terminal v_i. This example is for odd k. Note that the path $P_{i+\frac{k-1}{2}}$ is both in the right and the left hemispheres.

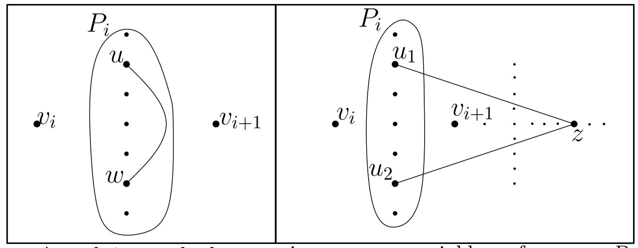
The induction step: assume that the claim is true for A and we will prove it for $A+1$. Let T be some spanning tree of $\overline{G}_{k,n_0,\ldots,n_{k-1}}$ with minimal weight among all the trees with $\gamma(T)=\mu$. By our assumption $\sum_{i=1}^{k} n_i = A+1$.

Claim 12. *For any vertex u (either a terminal or a non-terminal one), if there exist two edges $\{u,a\}$,$\{u,b\}$ in T that connect u to two vertices $a,b \in R(u)$, then the two vertices a and b belong to the same path P_i. The same is true for the left hemisphere $L(u)$ of u as well.*

Proof. Suppose for contradiction that there exist two edges $\{u,a\}$,$\{u,b\}$ in T as above (i.e., with $a,b \in R(u)$) and such that these two edges have different type. (In other words, $d_{\overline{G}}(u,a) \neq d_{\overline{G}}(u,b)$.) Without loss of generality $d_{\overline{G}}(u,a) < d_{\overline{G}}(u,b)$. We construct a new tree T' by replacing $\{u,b\}$ by $\{a,b\}$. This change decreases the weight of the tree. Note that any path π in T that uses the edge $\{u,b\}$ can be replaced by a similar path that uses the edges $\{u,a\}$ and $\{a,b\}$ instead of $\{u,b\}$. Clearly, the length of the path does not change. Hence for every index i, the value $t_{T,i}$ does not increase. Hence $\gamma(T) = \min_i \{n - t_{T,i}\}$ does not decrease, i.e., $\gamma(T') \geq \gamma(T)$. Since T is a tree with $\gamma(T) = \mu = \max_{T''}\{\gamma(T'')\}$, it follows that $\gamma(T') = \gamma(T) = \mu$. This is a contradiction to the minimality of the weight of T among trees with $\gamma()$ value equal to μ. Obviously, the same argument also applies for $a,b \in L(u)$. □

We now continue proving Lemma 11. The rest of the proof splits into a number of cases which are characterized by existence of certain edges in the tree T.

- Case 1: There is a path-internal edge in T. Assume that there is a path-internal edge between vertices in a path P_j, for some index $j \in [0,k-1]$. Note that if all the vertices which have path-internal edges in P_j have more than one path-internal edge incident on them in T, then the graph induced by those vertices contains a cycle. However, this graph is a subgraph of T,



A path-internal edge z is a common neighbor of $u_1, u_2 \in P_i$.

Figure 6: An illustration of the special edges that we use in the proof of Lemma 11.

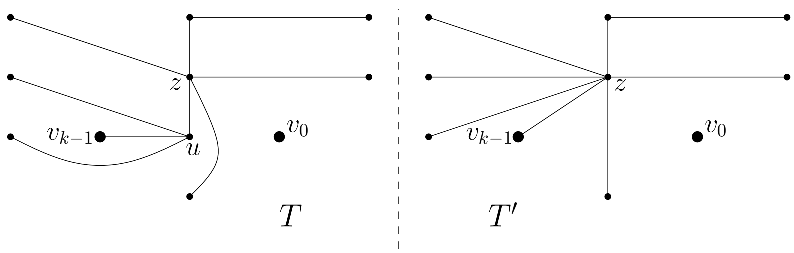


Figure 7: An illustration of deleting a path-internal edge $\{u, z\}$ in the tree T.

contradiction. Hence there is a vertex $u \in P_j$ such that u has exactly one path-internal edge $\{u, z\}$ incident on u in T. Assume without loss of generality that $j = 0$. Delete the vertex u from the graph and from the tree, and replace each edge $\{u, a\}$ incident on u in T, by an edge $\{z, a\}$. (Note that u has only one path-internal edge in T, $\{u, z\}$. Therefore vertices a as above do not belong to $P_j = P_0$, and this guarantees that the weight of each new edge $\{z, a\}$) is equal to the weight of its source $\{u, a\}$). Note also that no multiple edges between the same pair of vertices are introduced by this operation because T is acyclic. Denote the resulting tree by T'. Note that T' is a spanning tree for $\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}$. (See Figure 7 for an illustration.) (T' is obviously connected and we reduce the number of edges by exactly 1, hence it is a spanning tree of $\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}$.) Note that for every two vertices a, b in $\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}$, we have $d_{T'}(a, b) \leq d_T(a, b)$. (To see this consider the shortest path π from a to b in T. If $u \notin V(\pi)$ then π belong to T' as well, i.e., $d_{T'}(a, b) \leq d_T(a, b)$. Otherwise, if $\{u, z\} \in E(\pi)$, then denote $\pi = (a = w_0, w_1, \ldots, w_r = u, w_{r+1} = z, \ldots, w_h = b)$. Observe that all the edges of the path $\pi' = (a = w_0, w_1, \ldots, w_{r-1}, w_{r+1} = z, \ldots, w_h = b)$ belong to T', proving the claim. Finally, if $\{u, z\} \notin E(\pi)$ but $u \in V(\pi)$, then write $\pi = (a = w_0, \ldots, w_{r-1}, w_r = u, w_{r+1}, \ldots, w_h = b)$ and observe that all the edges of the path $\pi' = (a = w_0, \ldots, w_{r-1}, w_r = z, w_{r+1}, \ldots, w_h = b)$ belong to T'. The claim follows as the total length of π' is no greater than that of π.) Denote $n'_0 = n_0 - 1$, $n'_i = n_i$, for every $i \in [1, k-1]$. By the induction hypothesis, there is an index i such that $t_{T',i} = n'_i$. If $i \neq 0$ then we show that $t_{T,i} = n_i$. Indeed, in this case $n'_i = n_i$. Also, for every vertex $q \in P_i$, it holds that $d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_i, q) = d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_{i+1}, q) = d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_i, q) = d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_{i+1}, q) = w$, while $d_T(v_i, q) \geq d_{T'}(v_i, q)$ and

$d_T(v_{i+1},q) \geq d_{T'}(v_{i+1},q)$. Hence $\frac{d_{T'}(v_i,q)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_i,q)} \geq 2k-1$ implies

$$\frac{d_T(v_i,q)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_i,q)} \geq \frac{d_{T'}(v_i,q)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_i,q)} \geq 2k-1.$$

Analogously,

$$\frac{d_T(v_{i+1},q)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_{i+1},q)} \geq \frac{d_{T'}(v_{i+1},q)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_{i+1},q)},$$

so if the latter fraction is greater or equal than $2k-1$, so is the fraction $\frac{d_T(v_{i+1},q)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_{i+1},q)}$. Since

$$\max\left\{\frac{d_{T'}(v_i,q)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_i,q)}, \frac{d_{T'}(v_{i+1},q)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_{i+1},q)}\right\} \geq 2k-1,$$

it follows that

$$\max\left\{\frac{d_T(v_i,q)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_i,q)}, \frac{d_T(v_{i+1},q)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_{i+1},q)}\right\} \geq 2k-1.$$

Hence $t_{T,i} \geq t_{T',i} = n_i$. Since, by definition, $t_{T,i} \leq n_i$, it follows that $t_{T,i} = n_i$, as required. The case $i = 0$ is similar. In this case $t_{T',0} = n'_0 = n_0 - 1$. For a vertex $a \in P_0 \backslash \{u,z\}$, $d_T(v_0,a) \geq d_{T'}(v_0,a)$, and $d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_0,a) = d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_0,a) = w$. Hence

$$\frac{d_T(v_0,a)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_0,a)} \geq \frac{d_{T'}(v_0,a)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_0,a)}.$$

Observe also that by the same argument

$$\frac{d_T(v_1,a)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_1,a)} \geq \frac{d_{T'}(v_1,a)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_1,a)}.$$

Hence

$$\begin{aligned}&\max\left\{\frac{d_T(v_0,a)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_0,a)}, \frac{d_T(v_1,a)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_1,a)}\right\}\\ \geq\ &\max\left\{\frac{d_{T'}(v_0,a)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_0,a)}, \frac{d_{T'}(v_1,a)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_1,a)}\right\}.\end{aligned}$$

Since each vertex $a \in P_0 \backslash \{u,z\}$ contributes 1 to $t_{T',0}$, it means that

$$\max\left\{\frac{d_{T'}(v_0,a)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_0,a)}, \frac{d_{T'}(v_1,a)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_1,a)}\right\} \geq 2k-1.$$

Hence

$$\max\left\{\frac{d_T(v_0,a)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_0,a)}, \frac{d_T(v_1,a)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_1,a)}\right\} \geq 2k-1,$$

i.e., the vertex a contributes 1 also to $t_{T,0}$. Moreover, $d_T(v_0,u), d_T(v_0,z) \geq d_{T'}(v_0,z)$, and analogously $d_T(v_1,u), d_T(v_1,z) \geq d_{T'}(v_1,z)$. Since z contributes 1 to $t_{T',0}$ it follows that u and z each contribute 1 to $t_{T,0}$ and so $t_{T,0} = n_0$, completing the proof of case 1.

- Case 2: There are no path-internal edges in T, and there is a path P_i, with two vertices $u_1, u_2 \in P_i$ which have a common neighbor $z \notin P_i$. (Observe that there can be at most one common neighbor of u_1 and u_2 in T.) Without loss of generality, $i = 0$. Similarly to the previous case, delete u_2 and replace any edge $\{u_2, a\}$ by the edge $\{u_1, a\}$. Denote the resulting tree by T'. Note that T' is a spanning tree for $\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}$. ($T'$ is obviously connected and we reduce the number of edges by exactly 1.) Note also that for each two vertices a, b in $\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}$, it holds that $d_{T'}(a,b) \leq d_T(a,b)$. To see this, consider the shortest path π from a to b in T. If the vertex u_2 does not belong to π then the claim is trivial. Otherwise, write $\pi = (a = w_0, \ldots, w_{r-1}, w_r = u_2, w_{r+1}, \ldots, w_h = b)$. Let $\pi' = (a = w_0, \ldots, w_{r-1}, w_r = u_1, w_{r+1}, \ldots, w_h = b)$. Note that π' is a path from a to b in T', and the length of π' in T' is equal to that of π in T. Hence $d_{T'}(a,b) \leq d_T(a,b)$. Therefore, for every $q \in P_i$,
$$\begin{aligned} &\max\left\{\frac{d_T(v_i,q)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_i,q)}, \frac{d_T(v_{i+1},q)}{d_{\overline{G}_{k,n_0,\ldots,n_{k-1}}}(v_{i+1},q)}\right\} \\ \geq\ &\max\left\{\frac{d_{T'}(v_i,q)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_i,q)}, \frac{d_{T'}(v_{i+1},q)}{d_{\overline{G}_{k,n_0-1,n_1,\ldots,n_{k-1}}}(v_{i+1},q)}\right\}. \end{aligned}$$
Denote $n'_0 = n_0 - 1$, $n'_i = n_i$, for every $i \in [1, k-1]$. By the induction hypothesis, there is a path P'_i such that $t_{T',i} = n'_i$. If $i \neq 0$ then obviously $t_{T,i} = t_{T',i} = n'_i$. Otherwise, $i = 0$. Every vertex $a \in P_0 \setminus \{u_2\}$ contributes 1 to $t_{T',0}$, and hence contributes 1 also to $t_{T,0}$. Also, $d_T(u_2, v_0) \geq d_{T'}(u_1, v_0)$ and $d_T(u_2, v_{k-1}) \geq d_{T'}(u_1, v_{k-1})$. Since u_1 contributes 1 to $t_{T',0}$, it follows that u_2 contributes 1 to $t_{T',0}$ as well. Hence $t_{T,0} = t_{T',0} + 1 = n'_0 + 1 = n_0$, as required.

- Case 3: There are no path-internal edges, and for every index $i \in [0, \ldots, k-1]$, no two vertices in P_i have a common neighbor in T. Hence for every vertex v, the set of edges incident on v in T may contain at most one edge connecting it to its left hemisphere, and at most one edge connecting it to its right hemisphere. Indeed, by Claim 12, v cannot be incident on two edges of different type connecting it to its left (or right) hemisphere. Also, if v is incident on two edges of the same type, then the other endpoints of these edges belong to the same path P_i and have a common neighbor v. This is a contradiction. Hence any path π in T is one-sided. There exist two terminals v_i, v_{i+1} such that there is no path between them of length $2w$. (Otherwise, there is a cycle.) Hence for all $u \in P_i$, either $\{v_i, u\} \notin T$ or $\{v_{i+1}, u\} \notin T$. Without loss of generality, $\{u, v_i\} \notin T$. Then the shortest one-sided path between u and v_i is of length at least $(2k-1)w$ (since it must visit each terminal), implying that the stretch between u and v_i in T is at least $2k-1$. Hence $t_{T,i} = n_i$, as required.

□

Now we are ready to prove the main theorem of this section. It states that with respect to the tradeoff between the terminal stretch and the lightness, a metric spanning tree for $G_{k,n,n,\ldots,n}$ can do no better than a graph spanning tree for the same graph.

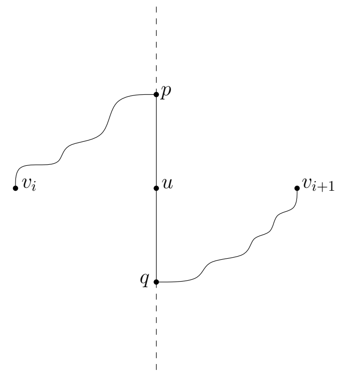


Figure 8: A visualization of the contradiction in the proof of Theorem 12. We assume that u has no incident edge of weight w. From the assumption that the distance from u to v_i is $(2k-1)\,w$ we conclude that the distance from q to v_i is larger than $(2k-1)\,w$, which is a contradiction.

Theorem 12. *For any positive integer parameters k, n and $\epsilon > 0$ such that $k \le \frac{\epsilon}{2}n$, any spanning tree T of $\overline{G}_{k,n,n,\dots,n}$, where $w = \frac{k}{\epsilon}$, with terminal stretch at most $(2k-1)\left(1+\frac{\epsilon}{2k^2}\right)$ has lightness at least $\Omega\left(\frac{1}{\epsilon}\right)$.*

Proof. Fix a spanning tree T of $\overline{G}_{k,n,n,\dots,n}$. By Lemma 11, there is a path P_i such that for any $u \in P_i$, either the pair $\{v_i, u\}$ or the pair $\{v_{i+1}, u\}$ incurs stretch at least $2k-1$. We will show that each $u \in P_i$ is incident at a not path-internal edge in T. (Each such an edge has weight at least w.) Before proving this assertion we make a few observations.

- Each $u \in P_i$ is at distance at most $(2k-1)\cdot w$ from both v_i and v_{i+1}. Otherwise, the respective distance (assume without loss of generality that it is between u and v_i) is at least $(2k-1)\,w+1$, and so

$$\begin{aligned}\frac{d_T(u,v_i)}{d_{\overline{G}_{k,n,\dots,n}}(u,v_i)} &= \frac{(2k-1)\,w+1}{w} = (2k-1)\left(1+\frac{1}{(2k-1)\,w}\right)\\ &= (2k-1)\left(1+\frac{\epsilon}{(2k-1)\,k}\right) > (2k-1)\left(1+\frac{\epsilon}{2k^2}\right).\end{aligned}$$

  This contradicts the assumption that T has terminal stretch at most $(2k-1)\left(1+\frac{\epsilon}{2k^2}\right)$.

- Each $u \in P_i$ is at distance exactly $(2k-1)\cdot w$ from either v_i or v_{i+1}, this follows by the assumption that u is of distance at least $(2k-1)\,w$ from v_i or v_{i+1} and the previous observation.

- Given any tree T, and two vertices u_1, u_2, such that u_1 and u_2 are neighbors in T, it holds that for any vertex $v \notin \{u_1, u_2\}$, the path from u_1 to v passes through u_2 or the path from u_2 to v passes through u_1.

Assume by contradiction, that there is $u \in P_i$ with no incident edge of weight at least w. Let p be the first vertex on the (unique) path from u to v_i in T, and q be the first vertex on the path from u to v_{i+1} in T. Both p and q belong to P_i. Without loss of generality, $d_T(v_i,u) = (2k-1)\,w$.

Hence $d_T(v_i, p) < d_T(v_i, u) = (2k-1)w$, which implies $d_T(v_{i+1}, p) = (2k-1)w$. The path in T from p to v_{i+1} passes through u, as otherwise $p = q$, and so $d_T(v_{i+1}, u) > d_T(v_{i+1}, q) = (2k-1)w$, contradiction. Hence $p \neq q$. Necessarily the path in T from q to v_i passes through u, which implies $d_T(v_i, q) > d_T(v_i, u) = (2k-1)w$, contradiction.

We conclude that $w(T) > n \cdot w$. Hence,

$$\Psi(T) = \frac{w(T)}{w(MST)} > \frac{n \cdot w}{(2k-1)w + (n-1)k} = \frac{n \cdot \frac{k}{\epsilon}}{(2k-1)\frac{k}{\epsilon} + (n-1)k} > \frac{n}{\epsilon n + 2k}.$$

As $k \leq \frac{\epsilon}{2}n$ we obtain $\Psi(T) \geq \frac{1}{2\epsilon}$, as required. □

5 NP-completeness of finding minimum weight spanning tree with terminal stretch $(2k-1)\alpha$

In this section we discuss the hardness of finding an MST with small terminal stretch. The main result presented in this section (Theorem 15) is that for small enough β, a graph G and a subset of terminals K, it is NP-hard to decide whether there is an MST with terminal stretch at most $(2k-1)\beta$. The proof is by a reduction from a problem presented in [KRY93]. Its definition is provided below.

Definition 5. *Given a weighted graph $G = (V, E, w)$, and a subset of terminals $K \subseteq V$, a spanning tree T of G with TermStretch at most α (with respect to K) and lightness at most β is called an $(\alpha, \beta) - k - LTT$ (LTT stands for light terminal tree). Given a weighted graph $G = (V, E, w)$, and a subset of terminals $K \subseteq V$, the decision problem of determining whether there is an $(\alpha, \beta) - k - LTT$ in G is called the* (α, β)-k-LTT-problem.

Definition 6. *Given a weighted graph $G = (V, E, w)$ and a root vertex $x \in V$, a tree T with stretch at most α with respect to x (i.e., for all $y \in V$, $d_T(x, y) \leq \alpha \cdot d_T(x, y)$) and lightness at most β is called an $(\alpha, \beta) - SLT$. Given a weighted graph $G = (V, E, w)$, and a root vertex $x \in V$, the decision problem of determining whether there is an $(\alpha, \beta) - SLT$ in G is called the* (α, β)-SLT-problem.

Note that an $(\alpha, \beta) - SLT$ is an $(\alpha, \beta) - k - LTT$ with $k = 1$. Define $\delta(G) = min_{\{u,w\} \in E} d_G(u, w)$ to be the minimal weight of an edge in the graph G. Define $D(G) = max_{\{u,w\} \in E} d_G(u, w)$ to be the maximal weight of an edge in the graph G.

Theorem 13. *(From [KRY93]) For a fixed $1 < \alpha \leq 2$, the $(\alpha, 1) - SLT$ problem is NP-complete. Moreover, restricted to graphs G with $\delta(G) = 1$ and $D(G) = 5.5$, the $(\alpha, 1) - SLT$ problem is still NP-complete.*

Khuller et al. in [KRY93] proved that the $(\alpha, 1) - SLT$ problem in NP-complete by a reduction from 3-SAT. They actually showed also that for every $\beta < 1 + \frac{2}{\alpha - 1}$, the $(\alpha, \beta) - SLT$ problem is NP-complete as well. However, we will only need the case $\beta = 1$. We start with a basic version of this problem, i.e., by showing that determining if there exists an MST with terminal stretch at most $2k - 1$ is NP-complete.

Theorem 14. $(2k-1, 1) - k - LTT$ *problem is NP-complete.*

Proof. The proof is by a reduction from $(\alpha, 1) - SLT$, which is NP-complete by Theorem 13. We are given a graph $G = (V, E, w)$, a vertex $r \in V$ and a constant $1 < \alpha < 2$. We construct a new graph $\hat{G}$ which consists of k copies of G (denoted $G_1, G_2, \ldots, G_k$ where the vertices of G_i are $\{u_i \mid u \in V\}$) and k additional terminals. In addition to the edges of the copies of G, there is a cycle $C_{2k} = \{v_1, r_1, v_2, r_2, \ldots, v_k, r_k\}$, where $\{v_1, \ldots, v_k\}$ are the terminals, and $\{r_1, \ldots, r_k\}$ are the k copies of the vertex r. All cycle edges have weight $x > D(G)$. In addition, from any terminal v_i, there is an edge to every vertex s in the ith and the $(i-1)$st copies of G. These edges have weight $x + \frac{\alpha \cdot d_G(r,s)}{2k-1}$. (See Figure 9 for an illustration.)

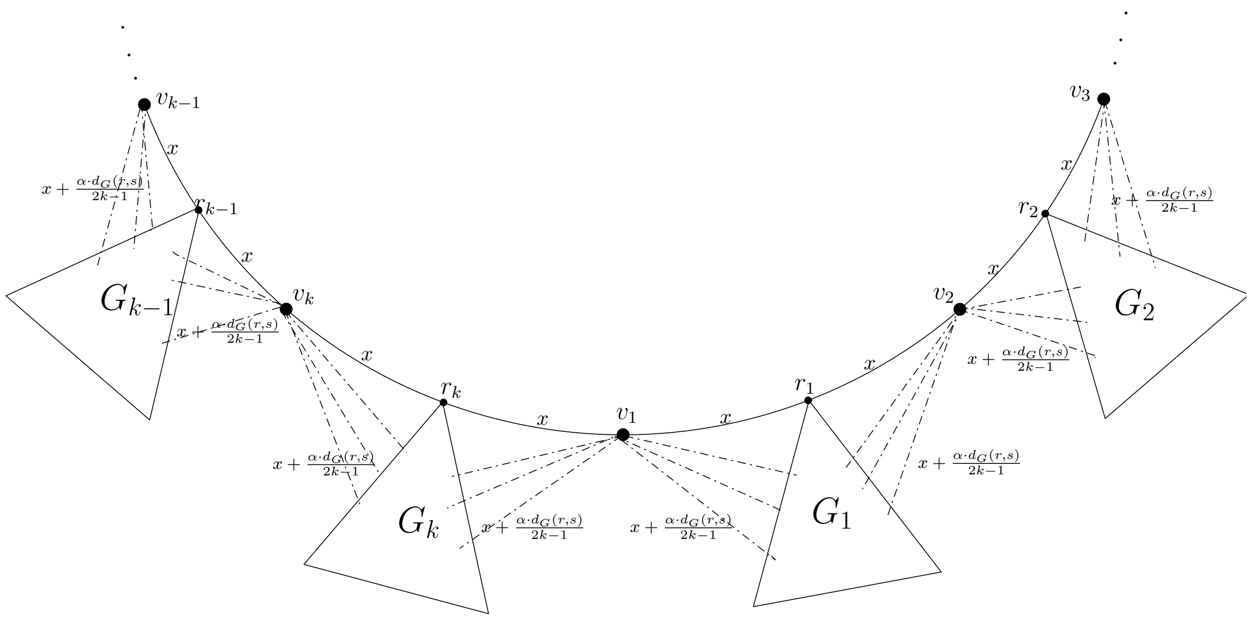


Figure 9: An illustration of the graph $\hat{G}$ constructed in the proof of Theorem 14. There are k copies of G, where the copy i is rooted at the vertex r_i. There are k additional terminals. The roots and the terminals placed on a cycle, and the two sets alternate. Each terminal has edges to all vertices that belong to one of his two neighboring copies.

Now we analyze the reduction. In other words, we show that there is an MST T of $G = (V, E, w)$ with stretch α from r if and only if there is an MST $\hat{T}$ of $\hat{G}$ with terminal stretch $2k-1$ with respect to $\{v_1, \ldots, v_k\}$.

- Assume that T is an MST of G with stretch α with respect to r. We construct a tree $\hat{T}$ for $\hat{G}$. $\hat{T}$ consists of k copies of T and all the edges of C_{2k} except of $\{r_k, v_1\}$. $\hat{T}$ is an MST. (It is obviously a spanning tree, and it contains no edge which is the heaviest in some cycle.) Obviously the greatest stretch is achieved on edges connecting v_1 to vertices of G_k. Hence it is sufficient to show that the stretch of those edges is bounded by $2k-1$. For all s_k in G_k,

$$\frac{d_{\hat{T}}(v_1, s_k)}{d_{\hat{G}}(v_1, s_k)} = \frac{(2k-1)\,x + d_T(r,s)}{x + \frac{\alpha \cdot d_G(r,s)}{2k-1}} \leq \frac{(2k-1)\,x + \alpha \cdot d_G(r,s)}{x + \frac{\alpha \cdot d_G(r,s)}{2k-1}} = 2k-1 \ ,$$

as required.

- Assume that $\hat{T}$ is an MST of $\hat{G}$ with terminal stretch $2k-1$. For all i, the edges connecting v_i to any vertex in G_i or G_{i+1} except for r_i, r_{i+1} cannot belong to $\hat{T}$, because each such an edge is the heaviest edge in a cycle. (Recall that $x > D(G)$). Hence exactly $2k-1$ edges from C_{2k} have to be part of $\hat{T}$, as otherwise $\hat{T}$ is either not connected (if it contains less than $2k-1$ edges of C_{2k}) or contains a cycle (if it contains exactly $2k$ edges of C_{2k}). Without loss of generality $\{r_k, v_1\} \notin \hat{T}$. Let T be the subgraph of $\hat{T}$ on the edges of G_k. Necessarily T is a spanning tree of G_k. It remains to show that T has stretch α with respect to r_k. (This will complete the proof because G_k and G are isomorphic.) For all $s \in V$,

$$\begin{aligned}\frac{d_T(r_k, s_k)}{d_{G_k}(r_k, s_k)} &= \frac{d_{\hat{T}}(v_1, s_k) - d_{\hat{T}}(v_1, r_k)}{d_G(r, s)} \le \frac{(2k-1)\cdot d_{\hat{G}}(v_1, s_k) - (2k-1)\,x}{d_G(r, s)} \\ &= \frac{(2k-1)\cdot\left(x + \frac{\alpha\cdot d_G(r,s)}{2k-1}\right) - (2k-1)\,x}{d_G(r, s)} = \frac{\alpha\cdot d_G(r, s)}{d_G(r, s)} = \alpha\ .\end{aligned}$$

□

Next we generalize Theorem 14 by relaxing the condition on terminal stretch. Specifically, we show that determining if there exist an MST with terminal stretch at most $2k-1+\frac{2k-1}{6k}$ is NP-complete as well.

Theorem 15. *For $k \ge 2$ and $1 \le \beta \le 1 + \frac{1}{6k}$, $((2k-1)\beta, 1) - k - LTT$ is NP-complete.*

Proof. The proof is (also) by a reduction from the $(\alpha, 1) - SLT$ problem for $\alpha = 2$. Here we restrict the $(\alpha, 1) - SLT$ problem to graphs G where $\delta(G) = 1$ and $D(G) = 5.5$. (Recall that by Theorem 13, this problem is NP-complete.) We are given a graph $G = (V, E, w)$, vertex r. Set also $\alpha = 2$. Let $\hat{G}$ be the graph constructed in the proof of Theorem 14, with the following small modification. The weight of every edge from a terminal v_i to a vertex s in one of the two neighboring copies of G (other than the roots r_i and r_{i+1}) is set to $\frac{x}{\beta} + \frac{\alpha}{\beta}\cdot\frac{d_G(r,s)}{2k-1}$ instead of $x + \frac{\alpha\cdot d_G(r,s)}{2k-1}$. (The weight of the edges between v_i and r_i, r_{i+1} is still x.)

Fix $x = 6$, recall that $\alpha = 2$. Note that for $\beta \in \left[1, 1 + \frac{1}{6k}\right]$, the following two properties hold:

1. For every $s \in V \setminus \{r\}$, $x < \frac{x}{\beta} + \frac{\alpha}{\beta}\cdot\frac{d_G(r,s)}{2k-1}$. This inequality holds because

   $$\frac{x}{\beta} + \frac{\alpha}{\beta}\cdot\frac{d_G(r,s)}{2k-1} = \frac{6}{\beta} + \frac{2}{\beta}\cdot\frac{d_G(r,s)}{2k-1} \ge \frac{1}{1+\frac{1}{6k}}\left(6 + \frac{2}{2k-1}\right) > 6 = x$$

   (The first inequality is because $d_G(r,s) \ge 1 = \delta(G)$.)

2. For every $s \in V \setminus \{r\}$, $x + d_G(r,s) > \frac{x}{\beta} + \frac{\alpha}{\beta}\cdot\frac{d_G(r,s)}{2k-1}$. This inequality follows since $\beta \ge 1$ and $1 > \frac{\alpha}{\beta(2k-1)}$.

Note that the first property combined with the fact that $x > D(G)$ implies that any MST of $\hat{G}$ does not contain edges incident on terminals except for edges of C_{2k}. We argue that there is an $(\alpha, 1) - SLT$ in G if and only if there is an $((2k-1)\beta, 1) - k - LTT$ in $\hat{G}$.

- Assume that T is an $(\alpha, 1) - SLT$ in G, i.e., there exists an MST T of G with stretch at most α with respect to r. We construct a tree $\hat{T}$ which consists of k copies of T and all the edges of

C_{2k} except for $\{r_k, v_1\}$. $\hat{T}$ is an MST of $\hat{G}$. (It is obviously a spanning tree, and it contains no edge which is the heaviest in some cycle.) Similarly to the proof of Theorem 14, the greatest stretch is achieved by edges that connect v_1 with vertices of G_k. Hence it is sufficient to show that the stretch of those edges is bounded by $(2k-1)\,\beta$. By the second property, for every $s \in V \setminus \{r\}$, it holds that $d_{\hat{G}}(v_1, s_k) = \frac{x}{\beta} + \frac{\alpha}{\beta} \cdot \frac{d_G(r,s)}{2k-1}$. Hence for every vertex s_k in G_k,

$$\frac{d_{\hat{T}}(v_1, s_k)}{d_{\hat{G}}(v_1, s_k)} = \frac{(2k-1)\,x + d_T(r,s)}{\frac{x}{\beta} + \frac{\alpha}{\beta} \cdot \frac{d_G(r,s)}{2k-1}} \leq \beta \cdot \frac{(2k-1)\,x + \alpha \cdot d_G(r,s)}{x + \alpha \cdot \frac{d_G(r,s)}{2k-1}} = \beta \cdot (2k-1)\ .$$

Hence $\hat{T}$ is an $((2k-1)\,\beta, 1) - k - LTT$ of $\hat{G}$, as required.

- Assume that $\hat{T}$ is a $((2k-1)\,\beta, 1) - k - LTT$ in $\hat{G}$. Hence there is an MST $\hat{T}$ of $\hat{G}$ with terminal stretch at most $\beta \cdot (2k-1)$. We saw that the edges from v_i to any vertex s in G_i or G_{i+1} except for r_i, r_{i+1} cannot belong to $\hat{T}$. In addition, exactly $2k-1$ edges from C_{2k} have to belong to $\hat{T}$. Without loss of generality, $\{r_k, v_1\} \notin \hat{T}$. Let T be the subgraph of $\hat{T}$ induced by the edges of G_k. Necessarily T is an MST of G. (It is obviously a spanning tree, and it does not contain any edge which is the heaviest in some cycle.) Hence for all $s \in V \setminus \{r\}$,

$$\begin{aligned}\frac{d_T(r,s)}{d_G(r,s)} &= \frac{d_{\hat{T}}(v_1, s_k) - d_{\hat{T}}(v_1, r_k)}{d_{\hat{G}}(r_k, s_k)} \leq \frac{(2k-1)\,\beta \cdot d_{\hat{G}}(v_1, s_k) - (2k-1)\,x}{d_G(r,s)} \\ &= \frac{(2k-1)\,\beta \cdot \left(\frac{x}{\beta} + \frac{\alpha}{\beta} \cdot \frac{d_G(r,s)}{2k-1}\right) - (2k-1)\,x}{d_G(r,s)} = \frac{\alpha \cdot d_G(r,s)}{d_G(r,s)} = \alpha\ ,\end{aligned}$$

  hence T is a $(\alpha, 1) - SLT$ of G as required.

□

Remark 11. *We conclude that given a weighted graph $G = (V, E, w)$, a subset of terminals $K \subseteq V$ of size k and a constant $\beta \leq 1 + \frac{1}{6k}$ it is NP-hard to find a spanning tree of G of minimal weight among all the spanning trees with $TermStretch$ at most $(2k-1)\,\beta$.*

We leave as an open problem the question whether $((2k-1)\,\alpha, \beta) - k - LTT$ is NP-complete for more general values of α and β.

Part III
Probabilistic Embedding

This part is dedicated to probabilistic embeddings, i.e, a distribution over embeddings into trees, and the goal is to bound the expected distortion. In Section 6 we present an embedding of any metric space into a distribution over ultrametrics with strong terminal distortion. This result is obtained via a delicate modification of [FRT04]. In Section 7 we show an embedding of an arbitrary graph into a distribution over its spanning trees with strong terminal distortion. This result is based on the petal decomposition technique of [AN12].

6 Probabilistic Embedding into Ultrametrics with Strong Terminal Distortion

Let (X, d) be a metric on n points, and $K \subseteq X$ the set of terminals of size $|K| = k$. Theorem 4 combined with [FRT04] gives an embedding into a distribution over ultrametrics with expected terminal distortion $O(\log k)$. However, this does not give any guarantee on the distortion of two non-terminal points.[6] In order to obtain *strong* terminal distortion, we modify the FRT algorithm, so that it gives "preference" to the terminals. In particular, the heart of the FRT algorithm is a construction of a random partitioning scheme, based on [CKR04], whose purpose is to decompose the metric to bounded diameter pieces, such that the probability of separating pairs is "sufficiently small". The partition is created by choosing a random permutation of X, a random radius in some specified interval, and then creating clusters as balls of the chosen radius, in the order given by the permutation. To obtain improved expected distortion guarantees for the terminals, we enforce them to be the first k points of the permutation. We show that this restriction improves dramatically the expected distortion of *any pair* that contains a terminal, while all the other pairs suffer only a factor of 2 in the expected distortion bound.

Theorem 16. *Given a metric space (X, d) of size $|X| = n$ and a subset of terminals $K \subseteq X$ of size $|K| = k$, there exists a distribution over embeddings of X into ultrametrics with strong terminal distortion $(O(\log k), O(\log n))$.*

Proof. Let Δ be the diameter of X. We assume w.l.o.g that the minimal distance in X is 1, and let δ be the minimal integer so that $\Delta \le 2^\delta$. We shall create a hierarchical laminar partition, where for each $i \in \{0, 1, \dots, \delta\}$, the clusters of level i have diameter at most 2^i, and each of them is contained in some level $i + 1$ cluster. Recall Definition 4. The ultrametric is built in the natural manner, the root corresponds to the level δ cluster which is X, and each cluster in level i corresponds to an inner node of the ultrametric with label 2^i, whose children correspond to the level $i - 1$ clusters contained in it. The leaves correspond to singletons, that is, to the elements of X. Clearly, the ultrametric will dominate (X, d).

In order to define the partition, we sample a uniformly random permutation, such that the terminals are its preface, and sample some number $\beta \in [1, 2]$. In each step, each vertex x chooses the vertex u with minimal value according to π among the vertices of distance at most $\beta_i = \beta \cdot 2^{i-2}$ from x, and joins the cluster of u. Note that a vertex may not belong to the cluster associated with

[6] In fact, it is not hard to verify that this approach may lead to unbounded distortion.

it, and some clusters may be empty (which we can discard). The description of the hierarchical partition appears in Algorithm 2.

Algorithm 2 `Modified FRT`(X, K)

Choose a random permutation π of X such that $\pi^{-1}(K) = [k]$.
Choose $\beta \in [1, 2]$ randomly by the distribution with the following probability density function $p(x) = \frac{1}{x \ln 2}$.
Let $D_\delta = \{X\}$; Assign the label 2^δ to D_δ; $i \leftarrow \delta - 1$.
while D_{i+1} has non-singleton clusters **do**
    Set $\beta_i \leftarrow \beta \cdot 2^{i-2}$.
    for $l = 1, \ldots, n$ **do**
        for every cluster S in D_{i+1} **do**
            Create a new cluster in D_i, consisting of all unassigned vertices in S closer than β_i to $\pi(l)$.
            Assign the label 2^i to D_i;
        end for
    end for
    $i \leftarrow i - 1$.
end while

Let T denote the ultrametric created by the hierarchical partition of Algorithm 2, and $d_T(u, v)$ the distance between u to v in T. Consider the clustering step at some level i, where clusters in D_{i+1} are picked for partitioning. In each iteration l, all unassigned vertices z such that $d(z, \pi(l)) \leq \beta_i$, assign themselves to the cluster of $\pi(l)$. Fix an arbitrary pair $\{v, u\}$. We say that center w *settles* the pair $\{v, u\}$ at level i, if it is the first center so that at least one of u and v gets assigned to its cluster. Note that exactly one center w settles any pair $\{v, u\}$ at any particular level. Further, we say that a center w *cuts* the pair $\{v, u\}$ at level i, if it settles them at this level, and exactly one of u and v is assigned to the cluster of w at level i. Whenever w cuts a pair $\{v, u\}$ at level i, $d_T(v, u)$ is set to be $2^{i+1} \leq 8\beta_i$. We charge this length to the vertex w and define $d_T^w(v, u)$ to be $\sum_i \mathbf{1}(w \text{ cuts } \{v, u\} \text{ at level } i) \cdot 8\beta_i$ (where $\mathbf{1}(\cdot)$ denotes an indicator function). Clearly, $d_T(v, u) \leq \sum_{w \in X} d_T^w(v, u)$.

We now arrange the vertices of K in non-decreasing order of their distance from the pair $\{v, u\}$ (breaking ties arbitrarily). Consider the sth terminal w_s in this sequence. We now upper bound the expected value of $d_T^{w_s}(v, u)$. W.l.o.g assume that $d_X(w_s, v) \leq d_X(w_s, u)$. For a center w_s to cut $\{v, u\}$, it must be the case that:

1. $d_X(w_s, v) \leq \beta_i < d_X(w_s, u)$ for some i.
2. w_s settles $\{v, u\}$ at level i.

For any i, let $g_i(y) = \begin{cases} 2^{i-2} & \text{if } y < 2^{i-2} \\ y & \text{if } y \in [2^{i-2}, 2^{i-1}) \\ 2^{i-1} & \text{if } y \geq 2^{i-1} \end{cases}$, for $1 \leq a \leq b$ it holds that $[a, b) = \bigcup_{i \geq 2} [g_i(a), g_i(b))$ (where $[x, x) = \emptyset$). Moreover, it holds that

$$\Pr[\beta_i \in [a, b)] = \Pr\left[\beta \in \left[\frac{a}{2^{i-1}}, \frac{b}{2^{i-1}}\right)\right] = \int_{g_i(a)/2^{i-2}}^{g_i(b)/2^{i-2}} \frac{1}{x \ln 2} dx = \int_{g_i(a)}^{g_i(b)} \frac{1}{x \ln 2} dx \ .$$

Conditioning on $\beta_i \in [d_X(w_s,u), d_X(w_s,v))$, any one of $w_1,\dots w_s$ can settle $\{v,u\}$, and the probability that w_s is first in the permutation among $w_1,\dots w_s$ is $\frac{1}{s}$. Thus we obtain,

$$\begin{aligned}\mathbb{E}[d_T^{w_s}(v,u)] &\le \sum_{i\ge 2}\int_{g_i(d_X(w_s,v))}^{g_i(d_X(w_s,u))} 8x\cdot\frac{1}{s}\cdot\frac{dx}{x\ln 2} \\ &= \frac{8}{s\cdot\ln 2}\int_{d(w_s,v)}^{d_X(w_s,u)} dx = \frac{8}{s\cdot\ln 2}(d_X(w_s,u)-d_X(w_s,v)) \le \frac{16}{s}\cdot d_X(v,u)\ .\end{aligned} \tag{3}$$

The crucial observation is that if $y\in X\setminus K$, and at least one of v,u is a terminal, w.l.o.g $v\in K$, then y cannot settle $\{v,u\}$. The reason is that v always appears before y in π, so v will surely be assigned to a cluster when it is the turn of y to create a cluster. This leads to the conclusion that for all $v\in K$ and $u\in X$

$$\mathbb{E}[d_T(v,u)] \le \sum_{s=1}^{k}\mathbb{E}[d_T^{w_s}(v,u)] \overset{(3)}{\le} 16 d_X(v,u)\sum_{s=1}^{k}\frac{1}{s} = O(\log k)\cdot d_X(v,u)\ .$$

It remains to bound the expected distortion of non-terminal pairs, so fix some $v,u\in X\setminus K$. Arrange the vertices $y_1,\dots,y_{n-k}$ of $X\setminus K$ according to their distance from $\{v,u\}$, in non-decreasing order. A similar reasoning as above gives the bound of (3) for the sth vertex y_s in this ordering to cut $\{v,u\}$ (recall that the permutation over the non-terminals is uniformly distributed). In fact, the bound is even slightly loose, because we disregard the possibility that a terminal will settle $\{v,u\}$. We conclude that

$$\begin{aligned}\mathbb{E}[d_T(v,u)] &\le \sum_{s=1}^{k}\mathbb{E}[d_T^{w_s}(v,u)] + \sum_{s=1}^{n-k}\mathbb{E}[d_T^{y_s}(v,u)] \\ &\le 16 d_X(v,u)\left(\sum_{s=1}^{k}\frac{1}{s}+\sum_{s=1}^{n-k}\frac{1}{s}\right) \\ &= O(\log k+\log n)\cdot d_X(v,u) \\ &= O(\log n)\cdot d_X(v,u)\ .\end{aligned}$$

□

7 Strong Terminal Embedding of Graphs into a Distribution of Spanning Trees

In this section we consider strong terminal embedding of graphs into a distribution over their spanning trees. We will show $(\tilde{O}(\log k), \tilde{O}(\log n))$-strong terminal distortion as promised in the introduction.

Theorem 17. *Given a weighted graph $G=(V,E,w)$ on n vertices, and a subset of k terminals $K\subseteq V$, there exists a distribution over spanning trees of G with strong terminal distortion $(\tilde{O}(\log k), \tilde{O}(\log n))$.*

We will use the framework of [AN12], in particular their petal decomposition structure in order to obtain the distribution over spanning trees. Roughly speaking, it is an iterative method to build a spanning tree. In each level, the current graph is partitioned into smaller diameter pieces, called petals, and a single central piece, which are then connected by edges in a tree structure. Each of the petals is a ball in a certain metric. The main advantage of this framework, is that it produces a spanning tree whose diameter is proportional to the diameter of the graph, but allows very large freedom for the choice of radii of the petals. Specifically, if the graph diameter is Δ, each radius can be chosen in an interval of length $\approx \Delta$. Intuitively, if a pair is separated by the petal decomposition, then its distance in the tree will be $O(\Delta)$. So we would like a method to choose the radii, that will give the appropriate bounds both on pairs containing a terminal, and for all other pairs, simultaneously. We note that in the star-partition framework of [EEST08] (used also in [ABN08]), the tree radius increases introduces a factor to the distortion that inherently depends on n, which does not seem to allow terminal distortion that depends on k alone.

In order to "take care" of pairs containing a terminal, we will need to somehow give the terminals a "preference" in the decomposition, as we did in Section 6. Since the [CKR04, FRT04] style partitioning scheme cannot work in a graph setting (because it does not produce strong diameter clusters, and the cluster's center may not be contained in it), we turn to the partitions based on truncated exponential distribution as in [Bar96, Bar04, ABN08]. To implement the "terminal preference" idea, we first build petals with the terminals as centers, and only then build petals for the remaining points. There are few technical subtleties needed to assure that pairs containing a terminal suffer small expected distortion. By a careful choice for the petal's center and the interval from which to choose the radius, we guarantee that in a level where no terminals are separated, none of the relevant balls around terminals are in danger of being cut. Using that the radius of clusters decreases geometrically, we conclude that each ball is "at risk" in at most $O(\log k)$ levels.

7.1 Petal decomposition description

Algorithm 3 $T =$ `hierarchical-petal-decomposition`$(G[X], x_0, t, K)$

if $|X| = 1$ **then**
  return $G[X]$.
end if
Let $(X_0, \dots, X_s, (y_1, x_1), \dots, (y_s, x_s), t_0, \dots, t_s) =$ `petal-decomposition`$(G[X], x_0, t, K)$;
for each $j \in [0, \dots, s]$ **do**
  $T_j =$ `hierarchical-petal-decomposition`$(G[X_j], x_j, t_j, K \cap X_j)$;
end for
Let T be the tree formed by connecting $T_0, \dots, T_s$ using the edges $\{y_1, x_1\}, \dots, \{y_s, x_s\}$;

In this section we present the petal decomposition algorithm, and quote some of its properties. We do not provide full proofs of these properties, these can be found in [AN12].

We start by executing the recursive procedure `hierarchical-petal-decomposition` with parameters (G, x_0, x_0, K), for an arbitrary vertex x_0 (which is our initial center and target).

Algorithm 4 $(X_0, \ldots, X_s, \{y_1, x_1\}, \ldots, \{y_s, x_s\}, t_0, \ldots, t_s) =$ `petal-decomposition`$(G[X], x_0, t, K)$

Let $\Delta = \Delta_{X,x_0}$; Let $Y_0 = X$; Set $j = 1$;
if $d_X(x_0, t) \geq 5\Delta/8$ **then**
    Let $(X_1, x_1) =$ `create-petal`$(X, t, x_0, [d_X(x_0, t) - 5\Delta/8, d_X(x_0, t) - \Delta/2], K)$;
    Let $Y_1 = Y_0 \setminus X_1$;
    Let y_1 be the neighbor of x_1 on $P_{x_0 t}$ (the one closer to x_0);
    Set $t_0 = y_1$, $t_1 = t$; $j = 2$;
else
    set $t_0 = t$.
end if
Creating the terminal petals:
while $K \cap Y_{j-1} \setminus B_X(x_0, 3\Delta/4) \neq \emptyset$ **do**
    Let $t_j \in Y_{j-1} \cap K$ be an arbitrary terminal satisfying $d_X(x_0, t_j) > 3\Delta/4$;
    Let $(X_j, x_j) =$ `create-petal`$(Y_{j-1}, t_j, x_0, [0, \Delta/8], K)$;
    $Y_j = Y_{j-1} \setminus X_j$;
    For each edge $e \in P_{x_j t_j}$, set its weight to be $w(e)/2$;
    Let y_j be the neighbor of x_j on $P_{x_0 t_j}$ (the one closer to x_0);
    Let $j = j + 1$;
end while
Creating the non-terminal petals:
while $Y_{j-1} \setminus B_X(x_0, 7\Delta/8) \neq \emptyset$ **do**
    Let $t_j \in Y_{j-1}$ be an arbitrary vertex satisfying $d_X(x_0, t_j) > 7\Delta/8$;
    Let $(X_j, x_j) =$ `create-petal`$(Y_{j-1}, t_j, x_0, [0, \Delta/32], \emptyset)$;
    $Y_j = Y_{j-1} \setminus X_j$;
    For each edge $e \in P_{x_j t_j}$, set its weight to be $w(e)/2$;
    Let y_j be the neighbor of x_j on $P_{x_0 t_j}$ (the one closer to x_0);
    Let $j = j + 1$;
end while
Let $s = j - 1$;
Creating the stigma X_0:
Let $X_0 = Y_s$;

`hierarchical-petal-decomposition` calls the algorithm `petal-decomposition`, and gets back a partition of V to sets $X_0, X_1, \ldots, X_s$, a set of edges $(y_1, x_1), \ldots, (y_s, x_s)$ and a set of vertices $t_0, \ldots, t_s$. The sets $X_1, \ldots, X_s$ are called petals, they are balls in a certain metric called a "cone metric" which is defined at Definition 7. The set X_0 is called a "stigma", it is the remainder of G after we "tear off" the petals. For each index $i \geq 1$, x_i is the center of the petal X_i, while t_i is its target. The algorithm proceeds recursively by executing `hierarchical-petal-decomposition` on each of the petals and on the stigma with the appropriate center and target. Each of the recursive calls generates a tree T_i. We connect the different trees we get from the different recursive calls using the edges $(y_1, x_1), \ldots, (y_s, x_s)$, to construct a tree T, which is in particular the output of our algorithm. Our version of `hierarchical-petal-decomposition` differs from the version of [AN12] only by sending additional data to `petal-decomposition` about the identity of the terminals.

The `petal-decomposition` procedure is the main procedure in our algorithm. It is responsible

Algorithm 5 $(W, x) = \texttt{create-petal}(X, Y, t, x_0, [lo, hi], K)$

$W_r = \bigcup_{p \in P_{x_0 t}:\ d_Y(p,t) \le r} B_{(Y, \rho(Y, x_0, p))}(p, (r - d_Y(p,t))/2)$;
if $K \neq \emptyset$ **then**
  $[lo, hi] = \texttt{reform-interval}(Y, t, x_0, [lo, hi], K)$;
  $L_k = \lceil \log \log k \rceil$; $R = lo - hi$;
  Let $1 \le q \le L_k$ be the minimal integer satisfying $\left|W_{lo+qR/L_k}\right|_k \le \frac{2|X|_k}{2^{\log^{1-q/L_k} k}}$;
  Set $a = lo + (q-1)R/L_k$, $b' = a + R/2L_k$, $\chi = \frac{|X|_k + 1}{|W_a|_k}$, $\hat{\chi} = \max\{\chi, e\}$;
  Set $\lambda = (2 \ln \hat{\chi})/(b' - a) = (4 L_k \ln \hat{\chi})/R = (160 L_k \ln \hat{\chi})/\Delta$;
else
  $L_n = \lceil \log \log n \rceil$; $R = lo - hi$;
  Let $1 \le q \le L_n$ be the minimal integer satisfying $\left|W_{lo+qR/L_n}\right| \le \frac{2|X|_n}{2^{\log^{1-q/L_n} n}}$;
  Set $a = lo + (q-1)R/L_n$, $b' = a + R/2L_n$, $\chi = \frac{|X|+1}{|W_a|}$, $\hat{\chi} = \max\{\chi, e\}$;
  Set $\lambda = (2 \ln \hat{\chi})/(b' - a) = (4 L_n \ln \hat{\chi})/R = (2^7 L_n \ln \hat{\chi})/\Delta$;
end if
Sample $r \in [a, b']$ according to the truncated exponential distribution with parameter λ, which has the following density function: $f(r) = \frac{\lambda \cdot e^{-\lambda r}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}}$;
Let $r' \le r$ be the maximum value such that there exists a point $p_{r'}$ of distance r' from t on $P_{x_0 t}$;
return $(W_r, p_{r'})$;

to partition a given graph G into petals and stigma, and in addition to provide to each new petal a center and a target, and also a set of edges which will be used to connect between the different petals and the stigma. The procedure creates petals sequentially, by calling to the `create-petal` procedure, and providing it with the remaining graph, the center x_0, a target chosen for this particular petal, the overall set of terminals and interval from which the `create-petal` procedure is supposed to pick a radius. The procedure starts by (possibly) creating a special first petal X_1, around the target t. Then it proceeds by creating petals X_i, using terminal as targets, and when no potential terminals remain, it proceeds by creating petals around regular vertices, until no vertices remain outside of a certain ball around the center x_0. The remaining un-assigned vertices become the stigma X_0. For each new petal X_i, after its creation, we divide by 2 the weight of the edges between the root of the petal x_i and its target t_i. This technical change used in the analysis of [AN12] to bound the radius of the created tree. We say more about this issue and its implications at Fact 3 and the discussion following it. Our Version of `petal-decomposition` differs from the version of [AN12] in:

1. While in the original algorithm the targets were chosen arbitrarily, in our version we first choose terminals as targets, and only when there are no terminals left sufficiently far from x_0, we choose the other vertices as targets. The purpose of this ordering, is to ensure that only the terminal petals may cut certain balls around the other terminals.

2. The segments sent to `create-petal` procedure. In the various calls to `create-petal`, we sent different segments than in the original version. However, our intervals are contained in the intervals sent in the original version. The purpose of this change is to ensure that petals which are created in line 22 are far away from any non-covered terminal.

Algorithm 6 $[lo', hi'] = \texttt{reform-interval}(Y, t, x_0, [lo, hi], K)$

```
Let W_r = ⋃_{p∈P_{x_0 t}: d_Y(p,t)≤r} B_{(Y,ρ(Y,x_0,p))}(p, (r − d_Y(p,t))/2);
Set α ← hi − lo;
if K ⊆ W_{lo+3α/5} then
    return [lo + 4α/5, hi];
end if
if K ∩ W_{lo+2α/5} = ∅ then
    return [lo, lo + α/5];
end if
return [lo + 2α/5, lo + 3α/5];
```

3. We stop constructing petals at different radii for terminals and non-terminals. When there are no more terminals outside $B_X(x_0, 3\Delta/4)$ we start handling the non-terminals, but only those outside $B_X(x_0, 7\Delta/8)$. This again guarantees that non-clustered terminals will not be affected by non-terminals. In the original version there is no such transition. In particular, this change implies that the upper bound on the radius of the stigma bounded by $\frac{7}{8}\Delta$, rather than by $\frac{3}{4}\Delta$ as in the original version.

The `create-petal` procedure, receives from the `petal-decomposition` procedure as an input the global set X, the current cluster Y (the remaining graph after deleting the previously created petals), the center x_0, target t for this specific petal, set K of terminals (in X) and an interval $[lo, hi]$ from which `create-petal` is supposed to choose radius for the new petal. `create-petal` starts by specifying a sub interval of $[lo, hi]$ by calling to `reform-interval`, which is explained below. Then `create-petal` samples a radius according to a truncated exponential distribution with parameter λ. This parameter is essentially the density change of terminals (respectively, all vertices) between a large ball and a small ball around the target. This procedure can be viewed as the probabilistic equivalent of `petal-decomposition` from [AN12]. The original procedure finds deterministically a radius that separates a small fraction of the edges to get small distortion in average, while we sample a radius to obtain small distortion simultaneously for every pair, but only in expectation. However, our procedure picks a radius from the given interval, and this is all what we need to inherit the properties we are interested in from the original algorithm.

The `reform-interval` procedure does not appear at all in the original version. It gets an interval $[lo, hi]$ it addition to the current cluster Y, the center x_0, the target t of the current petal and a set of terminals K (all the terminals in X). It returns a sub interval of $[lo, hi]$ of size $(hi - lo)/5$, a change which is not significant for our analysis. The purpose of this is to ensure that certain balls around terminals may be cut only at levels when some terminals are separated (note that there can be at most k such levels). The features of this subinterval are formulated in Claim 14.

Next, we quote some properties of the petal decomposition algorithm proved in [AN12]. Their proofs remain valid even after our modification, because the algorithm allows for arbitrary choice of targets after the first petal is built, and allows any radius in the appropriate interval. For a subset $X \subseteq V$, let Δ_{X,x_0} denote the radius of $G[X]$ with respect to x_0 (the maximal distance from x_0 in $G[X]$). We often omit the subscript x_0 if it is clear from the context.

Fact 1. *For a graph G, a vertex set X, and some vertices $x_0, t \in X$,* `petal-decomposition` $(G[X], x_0, t, K)$ *returns* $(X_0, \ldots, X_s, \{y_1, x_1\}, \ldots, \{y_s, x_s\}, t_0, \ldots, t_s)$ *such that for each* $j \leq s$, $\Delta_{X_j, x_j} \leq 7\Delta_{X, x_0}/8$.

Fact 2. *The* ***`hierarchical-petal-decomposition`*** *algorithm returns a tree.*

Fact 3. *Every edge* $e \in E$ *can have its weight multiplied by* $\frac{1}{2}$ *at most once throughout the execution of the algorithm.* [7]

Fact 4. *For a graph* G *and tree a* T *created by the* ***`hierarchical-petal-decomposition`****,*

$$\Delta_T \leq 4\Delta_G.$$

Fact 5. *Let* $1 \leq j \leq s$ *be an integer and* $z \in Y_j$,[8] *then* $P_{x_0 z}(X) \subseteq G[Y_j]$.

From Fact 4 we can deduce that if the radius of G was Δ (with respect to some vertex x_0) then the diameter of the tree created by the hierarchical petal decomposition is bounded by 8Δ. In lines 15 and 24 of the `petal-decomposition` procedure we divide the weight of some edges by 2. By Fact 3 it can happen to any edge at most once, so in the analysis of the distortion we will ignore this factor (for simplicity). We now formally define the cone-metric, which is the metric used for creating petals, and some of its properties.

Definition 7 (Cone metric)**.** *Given a graph* $G = (V, E, w)$*, a subset* $X \subseteq V$ *and points* $x, y \in X$*, define the* cone-metric $\rho = \rho(X, x, y) : X^2 \to \mathbb{R}^+$ *as* $\rho(u, v) = |(d_X(x, u) - d_X(y, u)) - (d_X(x, v) - d_X(y, v))|$.

This is actually a pseudo metric. For a cone metric $\rho = \rho(X, x, y)$,

$$\rho(y, u) = |(d_X(x, u) - d_X(y, u)) - (d_X(x, y) - d_X(y, y))| = d_X(x, y) + d_X(y, u) - d_X(x, u)$$

is the difference between the shortest path from x to u in $G[X]$ to the shortest path from x and u in $G[X]$ that goes through y. Therefore, the ball $B_{(X,\rho)}(y, r)$ in the cone metric $\rho = \rho(X, x, y)$ centered in y with radius r, contains all such vertices u. Denote by $P_{xy}(G)$ the shortest[9] path between x and y in G. If Y is a subset of the vertices of G, we denote by $P_{xy}(Y)$ the shortest path from x to y in $G[Y]$.

A petal is a union of balls in the cone metric. In the `create-petal` algorithm, while working in a subgraph $G[Y]$ with two specified vertices: a center x_0 and a target t, we define $W_r(Y, x_0, t) = \bigcup_{p \in P_{x_0 t}:\ d_Y(p,t) \leq r} B_{(Y, \rho(Y, x_0, p))}(p, \frac{r - d_Y(p,t)}{2})$ which is union of balls in the cone metric, where any vertex p in the shortest path from x_0 to t of distance at most r from t is a center of a ball with radius $\frac{r - d_Y(p,t)}{2}$. We will often omit the parameters and write just W_r if Y, x_0 and t are clear from the context. The next claim, which is implicit in [AN12], states that W_r is monotone (in r) and that W_{r+4l} contains the ball with radius l around W_r (where the ball is taken in the shortest path metric in $G[Y]$).

Claim 13. *For* $W_r(Y, x_0, t) = \bigcup_{p \in P_{x_0 t}:\ d_Y(p,t) \leq r} B_{(Y, \rho(Y, x_0, p))}(p, (r - d_Y(p, t))/2)$

1. W_r *is monotone in* r*, i.e., for* $r \leq r'$ *it holds that* $W_r \subseteq W_{r'}$.
2. *For every* $y \in W_r$ *and* l*, the ball* $B_Y(y, l)$ *contained in* W_{r+4l}*, i.e.,* $\forall z,\ d_Y(y, z) \leq l \Rightarrow z \in W_{r+4l}$

[7] This multiplication may happen only for edges in the special paths P_{x_j, t_j} for a petal X_j which is not the first special petal. Note that after the creation of X_j, t_j will be the target for the first special petal in the decomposition of X_j which will ensure that the weights of the edges in P_{x_j, t_j} won't decrease again.

[8] Y_j is different in the `petal-decomposition` procedure.

[9] For simplicity, we assume that for each pair of vertices there is a unique shortest path in G.

7.2 Analysis

The analysis will proceed as follows: in Claim 14 we formulate the impact of the `reform-interval` procedure in limiting the occasions in which certain balls around the terminal may be cut. Then we will have a discussion, during which we give names and definitions to the different events used in the analysis, and bound the expected stretch of a pair of vertices using these definitions. Then we formulate the main lemma, Lemma 15. This lemma bounds the probability of a ball to be cut during a single execution of `petal-decomposition` procedure. The proof of Lemma 15 uses the properties of the truncated exponential distribution, which we use to sample the radii at `create-petal` procedure. Our main effort will be invested in the proof of this lemma. Using Lemma 15, we will prove Claim 16, a claim that bounds the sum of the probabilities of cutting a certain ball during an $O(\log\log\log k)$ ($O(\log\log\log n)$ for non terminals) fraction of the levels of `hierarchical-petal-decomposition`. Eventually, combining Claim 16 on different fractions of the levels, we are able to prove Theorem 17.

In what follows, fix a cluster X with center x_0, a target t, and the set of terminals K contained in X. Let $(X_0,\dots,X_s,\{y_1,x_1\},\dots,\{y_s,x_s\},t_0,\dots,t_s)$ be the result of applying `petal-decomposition`$(G[X],x_0,t,K)$. The following claim will be useful for bounding the number of levels in which a relevant ball around a terminal is in danger of being cut.

Claim 14. *The following assertions hold true:*

- *If X_1 is a petal created at line 3 of `petal-decomposition` and $K\cap X_1=\emptyset$, then for all $v\in K$, $B_X(v,\Delta/160)\cap X_1=\emptyset$.*
- *If X_j is a petal created at lines 3 or 13 of `petal-decomposition` and $K\subseteq X_j$, then for all $v\in K$, $B_X(v,\Delta/160)\subseteq X_j$.*
- *If X_j is a petal created at line 22 of `petal-decomposition`, then for all $v\in K\cap Y_{j-1}$, $B_X(v,\Delta/16)\cap X_j=\emptyset$.*

Proof. Let W_r be as defined in line 1 of `create-petal` when forming the petal X_j. For the first assertion, note that the condition of line 6 in `reform-interval` must be satisfied (otherwise W_r necessarily contains some terminal), no terminals are in $W_{lo+2\alpha/5}$, and thus hi' is set to $lo+\alpha/5$ (we use α as in `reform-interval`, the size of the interval sent by `create-petal`.) Now, for every $u\in X_1\subseteq W_{hi'}$ and $v\in K$, $d_X(u,v)\geq\alpha/20=\Delta/160$, as otherwise, Claim 13 implies that $v\in W_{hi'+4\cdot\alpha/20}=W_{lo+2\alpha/5}$, contradiction.

For the second assertion, observe that $j\in\{1,2\}$, because X_2 certainly contains a terminal since we are at the "creating terminal petals" stage. As $K\subseteq X_j$, it must be that the condition of line 3 in `reform-interval` is satisfied, and in line 4 we set lo' to be $lo+4\alpha/5$. We conclude that if $u\in B_X(v,\Delta/160)$ for some $v\in K$, then $d_{Y_{j-1}}(u,v)\leq\Delta/160$ (because by the first assertion, this ball around v was not cut by the first petal). By Claim 13, $u\in W_{lo+3\alpha/5+\Delta/40}=W_{lo+4\alpha/5}\subseteq X_j$.

Finally we prove the third assertion. By the termination condition in line 11 of `petal-decomposition`, it must be that $v\in B_X(x_0,3\Delta/4)$. By Fact 5, all distances in Y_{j-1} to x_0 remain the same as in X. Note that the parameter r of the petal $X_j=W_r$ is at most $\Delta/32$, and that $d_X(x_0,t_j)\geq 7\Delta/8$. By the definitions of a petal and of the cone-metric, for any $p\in P_{x_0t}$ with $d_X(p,t)\leq\Delta/32$ we have that if $u\in B_{(Y_{j-1},\rho(Y_{j-1},x_0,p))}(p,(r-d_{Y_{j-1}}(p,t))/2)$ then

$$\begin{aligned} d_X(x_0,u) &\geq d_X(x_0,p)+d_{Y_{j-1}}(p,u)-(r-d_{Y_{j-1}}(p,t))/2>d_X(x_0,p)-r\\ &\geq d_X(x_0,t)-d_X(p,t)-\Delta/32\geq 7\Delta/8-2\Delta/32\geq 13\Delta/16\ . \end{aligned}$$

By the triangle inequality we obtain that $d_X(v,u) \geq d_X(x_0,u) - d_X(x_0,v) > \Delta/16$, which concludes the proof.

□

By Claim 14, given a terminal v in cluster X of radius Δ, and a ball $B = B_X(v,l)$ of radius at most $l \leq \Delta/160$, the ball B might be cut in the petal decomposition of X only if the terminal v is separated from some other terminal. `petal-decomposition` picks terminal targets arbitrary but deterministically. Therefore we will be more specific, and decide that `petal-decomposition` pick terminal targets by increasing order of the remaining terminals. If the first petal is built in line 3, then the first target is given, and by the `reform-interval` procedure, it is deterministically determined whether the first petal will contain all, part, or none of the terminals. If there are terminals remaining, and the first special petal (if exists) contains none of the terminals, then the second target is once again picked deterministically, and again `reform-interval` deterministically determines whether the petal will contain all or only part of the terminals. In any case, the issue of whether all the terminals will be in single petal or not, is resolved deterministically. Hence, given a cluster X, root x_0 and target t, it is known if there is some danger that B may be cut. Moreover, there are at most k levels in which some terminals are separated, and hence involving a danger of cutting balls around terminals. (If $l > \Delta/160$ then we do not care whether B is cut or not, because the stretch induced on v and points outside of B will be a constant.)

7.3 Expected stretch

When dealing with weighted graphs, it could be that a tiny ball participates in many recursive levels, and thus is "threatened" many times. In order to avoid such situations, we shall change the algorithm slightly, and when performing a petal decomposition on a cluster X of radius Δ, we shall contract for each vertex $u \in X$, the ball of radius Δ/n^3 around u. In addition, if terminals are separated (which is determined deterministically) then for each terminal $v \in K$, we contract the ball of radius Δ/k^3 around v. These contractions are done sequentially; when a vertex contracts a ball of radius l it becomes a "supernode" of all the vertices within distance l, and for any edge leaving a vertex in the ball to some vertex z, we will have a corresponding edge with the same weight from the supernode to z. These contractions yield a cluster X', on which we run the `petal-decomposition` algorithm (note that X' may contain supernodes - vertices that correspond to several original vertices, and that X' induces a multi-graph). After the partition of X' to $X_0, \ldots, X_s$ is determined, we expand back the contracted balls, so that each vertex belongs to the cluster of its supernode.

This guarantees that a ball of radius l around each terminal can participate only at partitions of radii in the range $[l, k^3 \cdot l]$, and by Fact 1, the radii decrease by a factor of at least 7/8 every level, so there are at most $16 \log k$ levels in which each such ball participates. Similarly, a ball of radius l' around some vertex $u \in X$, can participate only at partitions of radii in the range $[l', n^3 \cdot l']$, so there are at most $16 \log n$ levels in which each such ball participates. This contraction, while saving small balls from being cut, may have an effect on the radius of the tree when we expand back the vertices. We claim that this will be a minor increase. To see this, note that in a particular level, expanding back the balls around terminals can increase distances by at most $2\Delta/k^2$ (because every contracted ball has diameter at most $2\Delta/k^3$, and there can be at most k contracted balls). Similarly, expanding back all the other contracted balls may increase distances by at most an additional term of $2\Delta/n^2$. Now, since there are at most k iterations in which terminals are separated (only in such levels the balls around them are contracted), even if the radius is increased by a factor of $1+2/k^2$ in every one of them, the total increase is at most $(1+2/k^2)^k < e^{2/k}$. Similarly, there are at most n iterations all in all (since $\emptyset \subsetneq X_0 \subsetneq X$), and even if the radius is increased by a factor of $1+2/n^2$ in every one of them, the total increase is at most $(1+2/n^2)^n < e^{2/n}$. Henceforth we shall ignore this minor increase, as it affects the stretch of every pair by at most a factor of $e^{2/k+2/n}$.

We will show that the distribution generated by the `hierarchical-petal-decomposition`(G, x_0, x_0, K) algorithm has strong terminal distortion $(\tilde{O}(\log k), \tilde{O}(\log n))$ (where x_0 is an arbitrary vertex). The hierarchical partition naturally corresponds to a laminar family of clusters, that are arranged in a hierarchical tree structure denoted by $\mathcal{T}$ (with V as the root, and each cluster X which is partitioned by `petal-decomposition` to $X_0, \ldots, X_s$, has them as its children in $\mathcal{T}$). The *level* of a cluster is its distance in $\mathcal{T}$ to the root. Note that there might be several trees that correspond to a single hierarchical partition (this depends on the precise edges connecting the different petals). For a pair of vertices x, y and hierarchical tree $\mathcal{T}$, let $d_{\mathcal{T}}(x,y)$ be the maximal distance between x and y in a tree T that corresponds to the hierarchical tree $\mathcal{T}$. Note that if x is separated from y in the hierarchical tree $\mathcal{T}$ in cluster X of radius Δ_X, then by Fact 4, $d_{\mathcal{T}}(x,y) \le 8\Delta_X$.

Fix any two points $x, y \in V$ and denote by $B_x = B_G(x, d(x,y))$. For $j \in [s]$, we say that B_x is *cut* by X_j if $B_x \cap X_j \notin \{\emptyset, B_x\}$. Further, B_x is cut in X if there exists some petal X_j that cuts B_x. For $i \ge 0$, let $S_{X,i}$ be the event that $B_x \subseteq X$ and X is a cluster in level i. When calling `petal-decomposition` on the cluster X, a sequence of clusters $X_1, \ldots, X_s, X_0$ is generated, with Y_j as defined in the algorithm. For $Y \subseteq X$ and integer $j \ge 1$, let $S_{X,i,Y,j}$ denote the event that $S_{X,i}$ holds, $B_x \subseteq Y$, and $Y = Y_{j-1}$. Define the following events

$$\mathcal{C}_{X,i,j} = \{B_x \cap X_j \notin \{\emptyset, B_x\}\} \qquad \text{“}B_x \text{ cut in level } i \text{ at iteration } j \text{ (when creating petal } X_j \text{ from } X)\text{”},$$
$$\mathcal{F}_{X,i,j} = \{B_x \cap X_j = \emptyset\} \qquad \text{“}B_x \text{ is not cut in level } i \text{ iteration } j\text{”}.$$

Denote by $\mathcal{F}_{X,i,(<j)}$ the event $\bigwedge_{0<h<j} \mathcal{F}_{X,i,h}$, by $\bar{\mathcal{F}}_{X,i,j} = \{B_x \cap X_j \neq \emptyset\}$ the event that B_x is either cut or contained in X_j, and let

$$\mathcal{E}_{X,i,j} = \{\mathcal{C}_{X,i,j} \wedge \mathcal{F}_{X,i,(<j)}\} \qquad \text{“}B_x \text{ is cut in level } i \text{ iteration } j \text{ but not in previous iterations”}.$$

Finally, define $\mathcal{E}_{X,i} = \left\{\mathcal{S}_{X,i} \wedge \bigcup_j \mathcal{E}_{X,i,j}\right\}$, as the event that B_x is cut for the first time in level i. With our new notations, we can bound the expected distance, as follows:

$$\begin{aligned}
\mathbb{E}[d_T(x,y)] &\le \sum_{\mathcal{T}} \Pr[\mathcal{T}] \cdot d_{\mathcal{T}}(x,y) \\
&= \sum_{\mathcal{T}} \sum_{i \in \mathbb{N}} \sum_{X \subseteq V} \Pr[\mathcal{E}_{X,i}] \cdot \Pr[\mathcal{T} \mid \mathcal{E}_{X,i}] \cdot d_{\mathcal{T}}(x,y) \\
&= \sum_{X,i} \Pr[\mathcal{E}_{X,i}] \sum_{\mathcal{T}} \Pr[\mathcal{T} \mid \mathcal{E}_{X,i}] \cdot d_{\mathcal{T}}(x,y) \\
&\le 8 \sum_{X,i} \Pr[\mathcal{E}_{X,i}] \cdot \Delta_X \cdot \sum_{\mathcal{T}} \Pr[\mathcal{T} \mid \mathcal{E}_{X,i}] \\
&= 8 \sum_{X,i} \Pr[\mathcal{E}_{X,i}] \cdot \Delta_X \ . \qquad (4)
\end{aligned}$$

Recall the constants defined in `create-petal`: $L_k = \lceil \log\log k \rceil$ and $L_n = \lceil \log\log n \rceil$. For $Y \subseteq X \subseteq V$ define $\varphi_k(X,Y,x) = \max\left\{\frac{|X|_k}{|B_Y(x,\Delta_X/(2^8 L_k))|_k}, e\right\}$, and similarly $\varphi_n(X,Y,x) = \max\left\{\frac{|X|}{|B_Y(x,\Delta_X/(2^8 L_n))|}, e\right\}$. The following lemma bounds the probability that a ball around a terminal or an arbitrary vertex, is cut. We defer its proof to Section 7.3.1.

Lemma 15. *For any integer i, and a cluster X with radius Δ_X (from some root vertex), it holds that if x is a terminal ($x \in K$) then (where x, y are as defined earlier)*

$$\Pr[\mathcal{E}_{X,i}]\Delta_X \leq 2^{14} d(x,y) L_k \sum_{j \leq k+1} \sum_{Y \subseteq X} \Pr\left[\mathcal{S}_{X,i,Y,j} \wedge \overline{\mathcal{F}}_{X,i,j}\right] \ln\left(\varphi_k\left(X,Y,x\right)\right) . \tag{5}$$

Moreover, without dependence wether x is a terminal, it holds that

$$\begin{aligned} \Pr[\mathcal{E}_{X,i}]\Delta_X \;\leq\;& 2^{14} d(x,y) L_k \sum_{j \leq k+1} \sum_{Y \subseteq X} \Pr\left[\mathcal{S}_{X,i,Y,j} \wedge \overline{\mathcal{F}}_{X,i,j}\right] \ln\left(\varphi_k\left(X,Y,x\right)\right) \\ &+ 2^{14} d(x,y) L_n \sum_{j} \sum_{Y \subseteq X} \Pr\left[\mathcal{S}_{X,i,Y,j} \wedge \overline{\mathcal{F}}_{X,i,j}\right] \ln\left(\varphi_n\left(X,Y,x\right)\right). \end{aligned} \tag{6}$$

The algorithm returns us a hierarchical tree $\mathcal{T}$. We refer to the base level (where there are only one cluster that contains all the vertices) as level 0, to the next level as level 1 and so on. For a hierarchical tree $\mathcal{T}$, let $A_{\mathcal{T}} \subseteq \mathbb{N}$ be the set of levels i, in which B_x is included in a cluster X with radius Δ_X, such that $d(x,y)$ is larger than Δ_X/k^3. Note that if x is a terminal, due to the contractions, if $d(x,y) \leq \Delta_X/k^3$, then the ball B_x could not be cut, hence $\mathcal{E}_{X,i} = \emptyset$. By Fact 1, $A_{\mathcal{T}} \subseteq \mathbb{N}$ is of size at most $\log_{8/7}(2 \cdot k^3) + 1 < 22 \log k$. (The radius has to be in the range $[d(u,v)/2, k^3 d(u,v)]$, once the radius is less than $d(u,v)/2$, B_x will surely be cut.) For any level i, denote by $A_{\mathcal{T},i} = \{i' \in A_{\mathcal{T}} \ :\ i' \geq i\}$. Similarly, let $Q_{\mathcal{T}}$ be the set of levels i, in which B_x is included in a cluster X with diameter at most $\Delta_X \leq n^3 \cdot d(x,y)$. For every vertex x due to the contraction, if $d(x,y) \leq \Delta_X/n^3$, $\mathcal{E}_{X,i} = \emptyset$. By Fact 1, $Q_{\mathcal{T}} \subseteq \mathbb{N}$ is of size at most $22 \log n$. For any level i, denote by $Q_{\mathcal{T},i} = \{i' \in Q_{\mathcal{T}} \ :\ i' \geq i\}$.

Define $l_k = \left\lceil \log_{8/7}(2^9 L_k) \right\rceil$, and $l_n = \left\lceil \log_{8/7}(2^9 L_n) \right\rceil$. Consider the partition of the indices of all the levels in the hierarchical tree to the sets $I_m = \{i \in \mathbb{N} \ :\ i = m \pmod{l_k}\}$ for $m \in \{0, 1, \ldots, l_k - 1\}$. Similarly, for $g \in \{0, 1, \ldots, l_n - 1\}$, $J_g = \{i \in \mathbb{N} \ :\ i = g \pmod{l_n}\}$. The next lemma, combined with Lemma 15, is used to bound the cut probability.

Claim 16. *For every positive integers m, g and t, it holds that:*

$$\sum_{i \in I_m:\ i \geq t} \sum_{X,j,Y} \Pr\left[\mathcal{S}_{X,i,Y,j} \wedge \overline{\mathcal{F}}_{X,i,j}\right] \ln\left(\varphi_k\left(X,Y,x\right)\right) \leq \sum_{X,\mathcal{T}} \Pr[\mathcal{S}_{X,t} \wedge \mathcal{T}] \left(\ln |X|_k + |A_{\mathcal{T},t}|\right) \tag{7}$$

$$\sum_{i \in J_g:\ i \geq t} \sum_{X,j,Y} \Pr\left[\mathcal{S}_{X,i,Y,j} \wedge \overline{\mathcal{F}}_{X,i,j}\right] \ln\left(\varphi_n\left(X,Y,x\right)\right) \leq \sum_{X,\mathcal{T}} \Pr[\mathcal{S}_{X,t} \wedge \mathcal{T}] \left(\ln |X| + |Q_{\mathcal{T},t}|\right) \tag{8}$$

Proof. Fix any such m, we will show (7) by (reverse) induction on $t \in I_m$.

For the base case, note that when t is sufficiently large, x and y must have been separated. E.g. if $t = \log_{8/7}(2\mathrm{diam}(G)/d(x,y))$, at levels $i \geq t$ the radius of any cluster will be less then $d(x,y)/2$

(using that the radius drops by at least 7/8 at every level). We get that $\Pr[\mathcal{E}_{X,i}] = 0$ for any possible X and $i \geq t$. Hence $\sum_{i\in I_m:\ i\geq t}\sum_{X,j,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]=0$ and (7) holds.

Assume (7) holds for $t+l_k \in I_m$, and prove for t. Observe that if Z is a cluster in level $t+l_k$ such that event $\mathcal{S}_{Z,t+l_k}$ holds, then there must be a cluster X at level t (the ancestor of Z) such that $\mathcal{S}_{X,t}$ holds, and there are also Y and j such that $Y = Y_{j-1}$, event $\bar{\mathcal{F}}_{X,i,j}$ holds, and $Z \subseteq Y \subseteq X$ (X_j is the ancestor of Z at level $t+1$). Also note that $\Delta_X \geq (8/7)^{l_k}\Delta_Z = 2^9 L_k\cdot\Delta_Z$, therefore $Z \subseteq B_Z(x,2\Delta_Z)\subseteq B_Y(x,\Delta_X/(2^8L_k))$. It follows that

$$
\begin{aligned}
&\sum_{Z,\mathcal{T}}\Pr[\mathcal{S}_{Z,t+l_k}\wedge\mathcal{T}]\left(\ln|Z|_k+|A_{\mathcal{T},t+l_k}|\right)\\
&\quad=\sum_{\mathcal{T},X,Y,Z,j\ :\ Z\subseteq Y\subseteq X}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\cdot\Pr[\mathcal{S}_{Z,t+l_k}\mid\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\cdot\left(\ln|Z|_k+|A_{\mathcal{T},t+l_k}|\right)\\
&\quad\leq\sum_{\mathcal{T},X,Y,Z,j\ :\ Z\subseteq Y\subseteq X}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\cdot\\
&\qquad\qquad\cdot\Pr[\mathcal{S}_{Z,t+l_k}\mid\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\left(\ln|B_Y(x,\frac{\Delta_X}{2^8L_k})|_k+|A_{\mathcal{T},t+l_k}|\right)\\
&\quad\leq\sum_{\mathcal{T},X,Y,j:Y\subseteq X}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\left(\ln|B_Y(x,\frac{\Delta_X}{2^8L_k})|_k+|A_{\mathcal{T},t+l_k}|\right)\ .
\end{aligned}\tag{9}
$$

Where the last inequality is by changing the order of summation and the fact that $\sum_{Z\subseteq Y}\Pr[\mathcal{S}_{Z,t+l_k}\mid\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\leq 1$. Now, if in the hierarchical tree $\mathcal{T}$, $t\in A_{\mathcal{T}}$, then $|A_{\mathcal{T},t}|\geq|A_{\mathcal{T},t+l_k}|+1$, so that

$$
\begin{aligned}
&\sum_{i\in I_m,\ i\geq t}\sum_{X,j,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_k\left(X,Y,x\right)\right)\\
&\quad=\sum_{i\in I_m,\ i\geq t+l_k}\sum_{X,j,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_k\left(X,Y,x\right)\right)\\
&\qquad+\sum_{X,j,Y}\Pr\left[\mathcal{S}_{X,t,Y,j}\wedge\overline{\mathcal{F}}_{X,t,j}\right]\ln\left(\varphi_k\left(X,Y,x\right)\right)\\
&\quad\overset{(7)}{\leq}\sum_{X,\mathcal{T}}\Pr[\mathcal{S}_{X,t}\wedge\mathcal{T}]\left(\ln|X|_k+|A_{\mathcal{T},t+l_k}|\right)+\sum_{X,j,Y_{j-1}}\Pr\left[\mathcal{S}_{X,t,Y,j}\wedge\overline{\mathcal{F}}_{X,t,j}\right]\ln\left(\varphi_k\left(X,Y_{j-1},x\right)\right)\\
&\quad\overset{(9)}{\leq}\sum_{\mathcal{T},X,Y,j:Y\subseteq X}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\left(\ln|B_Y(x,\frac{\Delta_X}{2^8L_k})|_k+|A_{\mathcal{T},t+l_k}|\right)\\
&\qquad+\sum_{\mathcal{T},X,Y,j}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\ln\varphi_k(X,Y,x)\\
&\quad\leq\sum_{\mathcal{T},X,Y,j}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\left(\ln|X|_k+|A_{\mathcal{T},t}|\right)\ .
\end{aligned}\tag{10}
$$

In the last inequality we used that if $\varphi_k(X,Y,x)=\frac{|X|_k}{|B_Y(x,\Delta_X/2^8L_k)|_k}$, then

$$\ln|B_Y(x,\Delta_X/2^8L_k)|_k+|A_{\mathcal{T},t+l}|+\ln\varphi_k(X,Y,x)=\ln|X|_k+|A_{\mathcal{T},t+l}|\leq\ln|X|_k+|A_{\mathcal{T},t}|\ ,$$

and if $\varphi_k(X,Y,x) = e$, then

$$\begin{aligned}\ln|B_Y(x,\Delta_X/2^8L)|_k + |A_{\mathcal{T},t+l}| + \ln\varphi_k(X,Y,x) &= \ln|B_Y(x,\Delta_X/2^8L|_k + |A_{\mathcal{T},t+l}| + 1 \\ &\leq \ln|X|_k + |A_t| \ .\end{aligned}$$

In addition note that

$$\begin{aligned}&\sum_{\mathcal{T},X,Y,j} \Pr[\mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j} \wedge \mathcal{T}]\,(\ln|X|_k + |A_{\mathcal{T},t}|) \\ &= \sum_{\mathcal{T},X,Y,j} \Pr[\mathcal{S}_{X,t} \wedge \mathcal{T}] \cdot \Pr[\mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j} \mid \mathcal{S}_{X,t} \wedge \mathcal{T}]\,(\ln|X|_k + |A_{\mathcal{T},t}|) \\ &= \sum_{\mathcal{T},X} \Pr[\mathcal{S}_{X,t} \wedge \mathcal{T}]\,(\ln|X|_k + |A_{\mathcal{T},t}|) \sum_{Y,j} \Pr[\mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j} \mid \mathcal{S}_{X,t} \wedge \mathcal{T}] \\ &= \sum_{\mathcal{T},X} \Pr[\mathcal{S}_{X,t} \wedge \mathcal{T}]\,(\ln|X|_k + |A_{\mathcal{T},t}|) \qquad (11)\end{aligned}$$

where the last equality follows by $\sum_{Y,j} \Pr[\mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j} \mid \mathcal{S}_{X,t} \wedge \mathcal{T}] = 1$, which holds because for every different Y,j, the events $\mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j}$ are disjoint. This concludes the proof of (7).

The proof of (8) is fully symmetric. Fix some g, we will show (8) by (reverse) induction on $t \in J_g$. The base case is trivial (for every X and $t \geq \log_{8/7}(2\text{diam}(G)/d(x,y))$, $\Pr[\mathcal{S}_{X,i}] = 0$ and therefore $\sum_{i\in J_g:\ i\geq t} \sum_{X,j,Y} \Pr\left[\mathcal{S}_{X,i,Y,j} \wedge \overline{\mathcal{F}}_{X,i,j}\right] = 0$.) Assume (8) holds for $t + l_n \in J_g$, and prove for t. For cluster Z such that event $\mathcal{S}_{Z,t+l_n}$ holds, there are unique clusters X,Y and an index j, such that the events $\mathcal{S}_{X,t}, \bar{\mathcal{F}}_{X,i,j}$ holds, $Y = X_j$, and $Z \subseteq B_Z(x, 2\Delta_Z) \subseteq B_Y(x, \Delta_X/(2^8 L_n))$. Similarly to equation (9) we get

$$\begin{aligned}&\sum_{Z,\mathcal{T}} \Pr[\mathcal{S}_{Z,t+l_n} \wedge \mathcal{T}]\,(\ln|Z| + |Q_{\mathcal{T},t+l_n}|) \\ &= \sum_{\mathcal{T},X,Y,Z,j\ :\ Z\subseteq Y\subseteq X} \Pr[\mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j} \wedge \mathcal{T}] \cdot \Pr[\mathcal{S}_{Z,t+l_n} \mid \mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j} \wedge \mathcal{T}] \cdot (\ln|Z| + |Q_{\mathcal{T},t+l_n}|) \\ &\leq \sum_{\mathcal{T},X,Y,j:Y\subseteq X} \Pr[\mathcal{S}_{X,t,Y,j} \wedge \bar{\mathcal{F}}_{X,t,j} \wedge \mathcal{T}] \left(\ln|B_Y(x, \frac{\Delta_X}{2^8 L_n})| + |Q_{\mathcal{T},t+l_n}|\right) \ . \qquad (12)\end{aligned}$$

Now we make the induction step:

$$
\begin{aligned}
&\sum_{i\in J_g,\ i\geq t}\sum_{X}\sum_{j,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_n\left(X,Y,x\right)\right)\\
&=\sum_{i\in J_g,\ i\geq t+l_n}\sum_{X}\sum_{j,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_n\left(X,Y,x\right)\right)\\
&\quad+\sum_{X}\sum_{j,Y}\Pr\left[\mathcal{S}_{X,t,Y,j}\wedge\overline{\mathcal{F}}_{X,t,j}\right]\ln\left(\varphi_n\left(X,Y,x\right)\right)\\
&\overset{(8)}{\leq}\sum_{X,\mathcal{T}}\Pr[\mathcal{S}_{X,t+l_n}\wedge\mathcal{T}]\left(\ln|X|+|Q_{\mathcal{T},t+l_n}|\right)\\
&\quad+\sum_{X}\sum_{j,Y}\Pr\left[\mathcal{S}_{X,t,Y,j}\wedge\overline{\mathcal{F}}_{X,t,j}\right]\ln\left(\varphi_n\left(X,Y,x\right)\right)\\
&\overset{(12)}{\leq}\sum_{\mathcal{T},X,Y,j}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\left(\ln|B_Y(x,\frac{\Delta_X}{2^8L_k})|+|Q_{\mathcal{T},t+l_n}|\right)\\
&\quad+\sum_{\mathcal{T},X,Y,j}\Pr\left[\mathcal{S}_{X,t,Y,j}\wedge\overline{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}\right]\ln\left(\varphi_n\left(X,Y_{j-1},x\right)\right)\\
&\leq\sum_{\mathcal{T},X,Y,j}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\left(\ln|X|+|Q_{\mathcal{T},t}|\right)\ .
\end{aligned}
$$

By the same calculation as in equation (11) we have $\sum_{\mathcal{T},X,Y,j}\Pr[\mathcal{S}_{X,t,Y,j}\wedge\bar{\mathcal{F}}_{X,t,j}\wedge\mathcal{T}]\left(\ln|X|+|Q_{\mathcal{T},t}|\right)=\sum_{X,\mathcal{T}}\Pr[\mathcal{S}_{X,t}\wedge\mathcal{T}]\left(\ln|X|+|Q_{\mathcal{T},t}|\right)$, Which concludes the proof of (8). □

Proof of Theorem 17. The proof of Theorem 17 follows from Lemma 15, Claim 16 and equation (4). We start by proving the assertion bounding the expected stretch over pairs containng a terminal of Theorem 17. Recall that for every hierarchical tree $\mathcal{T}$, $|A_{\mathcal{T}}|\leq 22\log k$. Note that $|X|_k\leq k$.

$$
\begin{aligned}
\mathbb{E}[d_T(x,y)] &\overset{(4)}{\leq} 8\sum_{X,i}\Pr[\mathcal{E}_{X,i}]\cdot\Delta_X\\
&= 8\sum_{m=0}^{l_k-1}\sum_{i\in I_m}\sum_{X}\Pr[\mathcal{E}_{X,i}]\cdot\Delta_X\\
&\overset{(5)}{\leq} 2^{17}L_k\cdot d(x,y)\sum_{m=0}^{l_k-1}\sum_{i\in I_m}\sum_{X,j\leq k+1,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_k\left(X,Y,x\right)\right)\\
&\overset{(7)}{\leq} 2^{17}d(x,y)L_k\sum_{m=0}^{l_k-1}\sum_{X,\mathcal{T}}\Pr[\mathcal{S}_{X,0}\wedge\mathcal{T}]\left(\ln|X|_k+|A_{\mathcal{T},0}|\right)\\
&\leq 2^{17}L_k\cdot d(x,y)\left(\ln k+22\log k\right)l_k\sum_{\mathcal{T}}\Pr[\mathcal{S}_{V,0}\wedge\mathcal{T}]\\
&\leq 2^{29}\log k\log\log k\log\log\log k\cdot d(x,y)=\tilde{O}\left(\log k\right)d(x,y)\ .
\end{aligned}
$$

Similarly, for the assertion bounding the expected stretch over all pairs of Theorem 17, recall that for every hierarchical tree $\mathcal{T}$, $|Q_{\mathcal{T}}| \le 22\log n$. Note that $|X| \le n$.

$$\begin{aligned}
\mathbb{E}[d_T(x,y)] &\overset{(4)}{\le} 8\sum_{X,i}\Pr[\mathcal{E}_{X,i}]\cdot\Delta_X \\
&\overset{(6)}{\le} 8\sum_{X,i}2^{14}d(x,y)L_k\sum_{j\le k}\sum_{Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_k\left(X,Y,x\right)\right) \\
&\quad + 8\sum_{X,i}2^{14}d(x,y)L_n\sum_{j}\sum_{Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_n\left(X,Y,x\right)\right) \\
&\le 2^{17}L_k\cdot d(x,y)\sum_{m=0}^{l_k-1}\sum_{i\in I_m}\sum_{X,j\le k,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_k\left(X,Y,x\right)\right) \\
&\quad + 2^{17}d(x,y)L_n\sum_{g=0}^{l_n-1}\sum_{i\in J_g}\sum_{X,j,Y}\Pr\left[\mathcal{S}_{X,i,Y,j}\wedge\overline{\mathcal{F}}_{X,i,j}\right]\ln\left(\varphi_n\left(X,Y,x\right)\right) \\
&\overset{(7)\wedge(8)}{\le} 2^{17}d(x,y)L_k\sum_{m=0}^{l_k-1}\sum_{i\in I_m}\sum_{X,\mathcal{T}}\Pr[\mathcal{S}_{X,0}\wedge\mathcal{T}]\left(\ln|X|_k+|A_{\mathcal{T},0}|\right) \\
&\quad + 2^{17}d(x,y)L_n\sum_{g=0}^{l_n-1}\sum_{i\in J_g}\sum_{X,\mathcal{T}}\Pr[\mathcal{S}_{X,0}\wedge\mathcal{T}]\left(\ln|X|+|Q_{\mathcal{T},0}|\right) \\
&\le \left(2^{29}\log k\log\log k\log\log\log k+2^{29}\log n\log\log n\log\log\log n\right)d(x,y) \\
&= \tilde{O}\left(\log n\right)d(x,y)\ .
\end{aligned}$$

□

7.3.1 Proof of Lemma 15

Let X be the vertex set of the graph in the current level i of the petal decomposition, with arbitrary center x_0, target t and radius Δ (with respect to x_0). Recall the vertices $x, y \in X$, and the ball $B_x = B_G(x, d(x,y))$. Set $\gamma = d(x,y)/\Delta$. As X and i are fixed, and all the probabilistic events in the statement of the lemma are contained in $\mathcal{S}_{X,i}$, we will implicitly condition all the probabilistic events during the proof on $\mathcal{S}_{X,i}$ (i.e. our sample space is restricted to $\mathcal{S}_{X,i}$). We shall also omit X and i from the subscript of the probabilistic events (i.e., we will write $\mathcal{C}_j, \mathcal{F}_j, \mathcal{F}_{(<j)}, \mathcal{E}_j, \mathcal{E}$ and we will write $\mathcal{Z}_{Y,j}$ instead of $\mathcal{S}_{X,i,Y,j}$).

The petal decomposition algorithm returns a partition $(X_0, \dots, X_s, (y_1, x_1), \dots, (y_s, x_s), t_0, \dots, t_s)$. We make a small change in the numbering of the created petals: let $t_1, \dots, t_{k+1}$ be the terminal targets chosen in lines 6 or 12 of `petal-decomposition`, and $t_{k+2}, \dots, t_{n+1}$ be the non-terminal targets chosen in line 21. Observe that in that notation we always have exactly $n+1$ petals, while there might be an index j such that $X_j = \emptyset$, in that case we say that X_j is an imaginary petal. Note that there are at most $k+1$ terminal targets because there are just k terminals (in addition to the first special petal whose target is not necessarily a terminal).

Recall the definitions from `create-petal` procedure, some of them depend on the type of petal (terminal or non-terminal petal), and some of them are actually random variables (which depend on the previously created petals). We will write these with an index j to clarify which petal they correspond to. In the terminal petal case ($j \le k+1$), $L_k = \lceil \log\log k \rceil$, $R_j = hi - lo = \Delta/40$ (after reformation), $a_j = lo + (q_j - 1)R_j/L_k$, $b'_j = a_j + R_j/(2L_k)$, $b_j = a_j + R_j/L_k$, where q_j is chosen in such a way that

$$\frac{2|X|_k}{2^{\log^{1-(q-1)/L_k} k}} \le |W_{a_j}|_k \le |W_{b_j}|_k \le \frac{2|X|_k}{2^{\log^{1-q/L_k} k}} \,. \tag{13}$$

Also, $\hat{\chi}_j = \max\{\frac{|X|_k+1}{|W_a|_k}, e\}$, $\lambda_j = \frac{2\ln\hat{\chi}_j}{b'_j - a_j} = \frac{4L_k \ln\hat{\chi}_j}{R_j} = \frac{160 L_k \ln\hat{\chi}_j}{\Delta}$. The radius r_j of the petal $X_j = W_{r_j}$, is chosen from $[a_j, b'_j]$ with density function $f(r) = \frac{\lambda_j \cdot e^{-\lambda_j r}}{e^{-\lambda_j \cdot a_j} - e^{-\lambda_j \cdot b'_j}}$. Analogously, in the non-terminal case ($j > k+1$): $L_n = \lceil \log\log n \rceil$, $R_j = hi - lo = \Delta/32$ (no reformation), $a_j = lo + (q_j - 1)R_j/L_n$, $b'_j = a_j + R_j/(2L_n)$, $b_j = a_j + R_j/L_n$, where q_j is chosen in such a way that $\frac{2|X|}{2^{\log^{1-(q_j-1)/L_n} n}} \le |W_{a_j}| \le |W_{b_j}| \le \frac{2|X|}{2^{\log^{1-q_j/L_n} n}}$. Also, $\hat{\chi}_j = \max\{\frac{|X|+1}{|W_a|}, e\}$, $\lambda_j = \frac{2\ln\hat{\chi}_j}{b'_j - a_j} = \frac{4L_n \ln\hat{\chi}_j}{R_j} = \frac{128 L_n \ln\hat{\chi}_j}{\Delta}$. The radius r_j of the petal $X_j = W_{r_j}$, is chosen from $[a_j, b'_j]$ with density function $f(r) = \frac{\lambda_j \cdot e^{-\lambda_j r}}{e^{-\lambda_j \cdot a_j} - e^{-\lambda_j \cdot b'_j}}$.

Let $\delta_j = e^{-8\lambda_j \gamma \Delta}$. Towards the proof, we assume that $\gamma \le \frac{1}{2^{10} L_k}$ as otherwise the assertions of the lemma are trivial (as $\Pr[\mathcal{E}_{X,i}] \le \sum_j \sum_{Y \subseteq X} \Pr\left[\mathcal{S}_{X,i,Y,j} \wedge \overline{\mathcal{F}}_{X,i,j}\right]$).

Claim 17. *For every* $1 \le j \le n+1$, $\Pr\left[C_j \mid \mathcal{Z}_{Y,j}\right] \le (1 - \delta_j)\left(\Pr\left[\overline{\mathcal{F}_j} \mid \mathcal{Z}_{Y,j}\right] + \frac{2}{\hat{\chi}_j^2}\right)$.

Proof. For ease of notation we write simply λ ,δ, χ, a and b' instead of $\lambda_j, \delta_j, \chi_j, a_j, b'_j$, as the proof is the same for both of the cases.[10] Let ρ be the minimal number greater than a such that $W_\rho \cap B_x \ne \emptyset$. If $\rho \ge b'$, then trivially $\Pr\left[\overline{\mathcal{F}_j} \mid \mathcal{Z}_{Y,j}\right] = 1$ and hence $\mathcal{P}_r\left[C_j \mid \mathcal{Z}_{Y,j}\right] = 0$ and we are done. Otherwise, $\rho \in [a, b']$, the probability that B_x intersects X_j is

$$\Pr\left[\overline{\mathcal{F}_j} \mid \mathcal{Z}_{Y,j}\right] = \int_\rho^{b'} f(r)\,dr = \int_\rho^{b'} \frac{\lambda \cdot e^{-\lambda r}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}}\,dr = \frac{e^{-\lambda\rho} - e^{-\lambda b'}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}} \,.$$

The ball B_x is of diameter at most $2\gamma\Delta$, therefore by Claim 13, $B_x \subseteq W_{\rho + 8\gamma\Delta}$. Hence

$$\begin{aligned}
\Pr\left[C_j \mid \mathcal{Z}_{Y,j}\right] &\le \int_\rho^{\rho+8\gamma\Delta} f(r)\,dr = \int_\rho^{\rho+8\gamma\Delta} \frac{\lambda \cdot e^{-\lambda r}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}}\,dr = \frac{e^{-\lambda\rho} - e^{-\lambda(\rho+8\gamma\Delta)}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}} \\
&= \frac{e^{-\lambda\rho}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}}\left(1 - e^{-\lambda 8\gamma\Delta}\right) = (1-\delta)\,\frac{e^{-\lambda\rho} - e^{-\lambda \cdot b'} + e^{-\lambda \cdot b'}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}} \\
&= (1-\delta)\left(\frac{e^{-\lambda\rho} - e^{-\lambda \cdot b'}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}} + \frac{e^{-\lambda \cdot b'}}{e^{-\lambda \cdot a} - e^{-\lambda \cdot b'}}\right) \\
&= (1-\delta)\left(\Pr\left[\overline{\mathcal{F}_j} \middle| \mathcal{Z}_{Y,j}\right] + \frac{1}{e^{\lambda \cdot (b'-a)} - 1}\right) \\
&\le (1-\delta)\left(\Pr\left[\overline{\mathcal{F}_j} \middle| \mathcal{Z}_{Y,j}\right] + \frac{2}{\hat{\chi}^2}\right).
\end{aligned}$$

Where the last inequality follows since $\hat{\chi} \ge 2$. □

[10] Note that if $X_j = \emptyset$ is imaginary petal created only for clarity in notation, then $\Pr\left[C_j \mid \mathcal{Z}_{Y,j}\right] = 0$ and the claim is obvious.

We are now ready to bound the probability that one of the terminal petals cuts B_x.

$$\begin{aligned}\Pr\left[\bigvee_{j\le k+1} C_j\right] &= \sum_{j\le k+1}\sum_{Y\subseteq X}\Pr\left[C_j \mid \mathcal{Z}_{Y,j}\right]\cdot\Pr[\mathcal{Z}_{Y,j}]\\ &\le \sum_{j\le k+1}\sum_{Y}(1-\delta_j)\left(\Pr\left[\overline{\mathcal{F}_j}\mid \mathcal{Z}_{Y,j}\right]+\frac{2}{\hat{\chi}_j^2}\right)\cdot\Pr[\mathcal{Z}_{Y,j}]\\ &= \sum_{j\le k+1}\sum_{Y}(1-\delta_j)\Pr\left[\overline{\mathcal{F}_j}\wedge \mathcal{Z}_{Y,j}\right]+\sum_{j\le k+1}\sum_{Y}(1-\delta_j)\frac{2}{\hat{\chi}_j^2}\cdot\Pr[\mathcal{Z}_{Y,j}]\ . \quad (14)\end{aligned}$$

To see how the first equality follows, note that the probability that B_x is cut by the first $k+1$ petals ($\Pr\left[\bigvee_{j\le k+1} C_j\right]$) is equal to the sum of the probabilities that B_x cut at the first time by petal j ($\sum_{j\le k+1}\Pr\left[\mathcal{C}_j\wedge\bigwedge_{l<j}\overline{C_l}\right]$). While for each j, the probability that B_x is cut at the first time by petal j is equal to the probability that B_x is active at iteration j (i.e. $\mathcal{F}_{(<j)}$), and is indeed cut by petal j ($\sum_{Y\subseteq X}\Pr\left[\mathcal{C}_j \mid \mathcal{Z}_{Y,j}\right]\cdot\Pr[\mathcal{Z}_{Y,j}]$).

Note that for $j\le k+1$, $1-\delta_j = 1-e^{-8\lambda_j\gamma\Delta} = 1-e^{-8\cdot 160\gamma L_k\ln\hat{\chi}_j} < 2^{11}\gamma L_k\ln\hat{\chi}_j$. For a set $Y=Y_{j-1}$ such that $\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\neq 0$, necessarily a target t_j and q_j are chosen so that $B_x\cap W_{b'_j}\neq\emptyset$ (otherwise it is no possible that X_j intersects B_x). As B_x is a ball of radius $\gamma\Delta\le\Delta/(2^{10}L_k)$, increasing the petal radius by $4\gamma\Delta$ we have $x\in W_{b'_j+4\gamma\Delta}$, and moreover $B\left(x,\frac{\Delta}{2^9L_k}\right)\subseteq W_{b'_j+4\gamma\Delta+\frac{\Delta}{2^7L_k}}\subseteq W_{b'_j+\frac{\Delta}{2^8L_k}+\frac{\Delta}{2^7L_k}}\subseteq W_{b'_j+\frac{R_j}{2L_k}}=W_{b_j}$.

We will show that for such a Y, (for which $\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\neq 0$), $\log\left(\hat{\chi}_j\right)\le 6\log\left(\varphi_k\left(X,Y,x\right)\right)$.[11] We may assume that $\hat{\chi}_j=\frac{|X|_k+1}{|W_a|_k}$, as otherwise ($\hat{\chi}_j=e$) the bound is trivial. Using that $\log^{1/L_k}k\le 3$ we get:

$$\begin{aligned}\log\hat{\chi}_j &= \log\left(\frac{|X|_k+1}{\left|W_{a_j}\right|_k}\right)\\ &\le \log\left(\frac{2|X|_k}{\left|W_{a_j}\right|_k}\right)\\ &\overset{(13)}{\le} \log^{1-(q-1)/L_k}k\\ &\le 3\log^{1-q/L_k}k\\ &\overset{(13)}{\le} 3\log\left(\frac{2|X|_k}{\left|W_{b_j}\right|_k}\right)\\ &\le 3\log\left(\frac{2|X|_k}{\left|B\left(x,\Delta/2^9L_k\right)\right|_k}\right)\\ &= 3\left(\log\left(\varphi_k\left(X,Y,x\right)\right)+1\right)\\ &\le 6\log\left(\varphi_k\left(X,Y,x\right)\right)\ .\end{aligned}$$

[11] Recall that $\varphi_k\left(X,Y,x\right)=\max\left\{\frac{|X|_k}{\left|B_Y\left(x,\Delta_X/(2^8L_k)\right)\right|_k},e\right\}$.

Note that in particular it is true that $\ln \hat{\chi}_j \leq 6 \ln\left(\varphi_k\left(X,Y,x\right)\right)$. Hence we can bound the first component of the summation in (14):

$$\begin{aligned}\sum_{j\leq k+1}\sum_{Y}\left(1-\delta_j\right)\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right] &\leq 2^{11}\gamma L_k\sum_{j\leq k+1}\sum_{Y}\ln\left(\hat{\chi}_j\right)\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\\ &\leq 6\cdot 2^{11}\gamma L_k\sum_{j\leq k+1}\sum_{Y}\ln\left(\varphi_k\left(X,Y,x\right)\right)\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ .\end{aligned} \quad (15)$$

For bounding the second component $\left(\sum_{j\leq k+1}\sum_{Y}\left(1-\delta_j\right)\frac{2}{\hat{\chi}_j^2}\cdot\Pr[\mathcal{Z}_{Y,j}]\right)$, note that $\frac{1}{\hat{\chi}_j}=\min\left\{\frac{\left|W_{a_j}\right|_k}{|X|_k+1},\frac{1}{e}\right\}\leq\frac{\left|W_{a_j}\right|_k}{|X|_k}$. In addition, the probability of the event $\mathcal{Z}_{Y,j}$ is equal to the sum of probabilities of sequences $X_0,\ldots,X_s$, over all sequences for which $Y=(X_j\cup\cdots\cup X_s\cup X_0)$ (while we abuse notation, $\Pr[X_0,\ldots,X_s]$ is the probability that the petal decomposition returned the partition $X_0,\ldots,X_s$). Note also that for each such partition, $X_0,\ldots,X_s$, $\left|W_{a_j}\right|_k\leq|X_j|_k$ and $\sum_{j\leq k+1}\frac{|X_j|_k}{|X|_k}\leq 1$, because all the petals, $X_1,\ldots,X_{k+1}$, , are pairwise disjoint. We get

$$\begin{aligned}\sum_{j\leq k+1}\sum_{Y}\frac{\Pr[S_{Y,j}]}{\hat{\chi}_j} &= \sum_{j\leq k+1}\sum_{Y:\ B_x\subseteq Y}\ \sum_{(X_0,\ldots,X_s):\ Y=(X_j\cup\cdots\cup X_s\cup X_0)}\Pr[X_0,\ldots,X_s]\frac{1}{\hat{\chi}_j}\\ &\leq \sum_{j\leq k+1}\sum_{Y:\ B_x\subseteq Y}\ \sum_{(X_0,\ldots,X_s)\ :\ Y=(X_j\cup\cdots\cup X_s\cup X_0)}\Pr[X_0,\ldots,X_s]\frac{\left|W_{a_j}\right|_k}{|X|_k}\\ &\leq \sum_{j\leq k+1}\sum_{(X_0,\ldots,X_s)}\mathbf{1}_{B_x\subseteq(X_j\cup\cdots\cup X_s\cup X_0)}\cdot\Pr[X_0,\ldots,X_s]\frac{|X_j|_k}{|X|_k}\\ &\leq \sum_{j\leq k+1}\sum_{(X_0,\ldots,X_s)}\Pr[X_0,\ldots,X_s]\frac{|X_j|_k}{|X|_k}\\ &= \sum_{(X_0,\ldots,X_s)}\Pr[X_0,\ldots,X_s]\sum_{j\leq k+1}\frac{|X_j|_k}{|X|_k}\\ &\leq \sum_{(X_0,\ldots,X_s)}\Pr[X_0,\ldots,X_s]=1\ .\end{aligned}$$

Where the third inequalitie follows by the fact that for a fixed j and a particular partition $X_0,\ldots,X_s$ of X there might be only a single set Y such that $B_x\subseteq Y=X_j\cup\cdots\cup X_s\cup X_0$. As $\frac{\ln\hat{\chi}_j}{\hat{\chi}_j}\leq 1$ we can bound the second component of the summation in (14):

$$\begin{aligned}\sum_{j\leq k+1}\sum_{Y}\left(1-\delta_j\right)\frac{2}{\hat{\chi}_j^2}\cdot\Pr[\mathcal{Z}_{Y,j}] &\leq 2^{11}\gamma L_k\sum_{j\leq k+1}\sum_{Y}\ln\hat{\chi}_j\frac{2}{\hat{\chi}_j^2}\cdot\Pr[\mathcal{Z}_{Y,j}]\\ &\leq 2^{12}\gamma L_k\sum_{j\leq k+1}\sum_{Y}\frac{\Pr[\mathcal{Z}_{Y,j}]}{\hat{\chi}_j}\\ &\leq 2^{12}\gamma L_k\\ &\leq 2^{12}\gamma L_k\sum_{Y,j}\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ln\varphi_k(X,Y,x)\ .\end{aligned} \quad (16)$$

Where the last inequality follows by $\sum_{Y,j}\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ln\varphi_k(X,Y,x)\geq\sum_{Y,j}\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]=1$ (recall that all the probabilities are implicitly conditioned on $\mathcal{S}_{X,i}$, and $\overline{\mathcal{F}_j}$ has to hold for some j). We conclude:

$$\begin{aligned} Pr\left[\bigvee_{j\leq k+1}C_j\right] &\overset{(14)}{\leq} \sum_{j\leq k+1}\sum_{Y}(1-\delta_j)\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]+\sum_{j\leq k+1}\sum_{Y}(1-\delta_j)\frac{2}{\hat{\chi}_j^2}\cdot\Pr[\mathcal{Z}_{Y,j}] \\ &\overset{(15)\wedge(16)}{\leq} \left(6\cdot 2^{11}+2^{12}\right)\gamma L_k\sum_{j\leq k+1}\sum_{Y}\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ln\left(\varphi_k\left(X,Y,v\right)\right) \\ &= 2^{14}\gamma L_k\sum_{j\leq k+1}\sum_{Y}\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ln\left(\varphi_k\left(X,Y,v\right)\right)\ . \end{aligned}$$

Claim 18. *If $x\in K$, then for $j>k+1$, it holds that $\Pr\left[C_j\mid\mathcal{Z}_{Y,j}\right]=0$.*

Proof. Since we condition on $\mathcal{Z}_{Y,j}$, it implies that $B_x\subseteq Y_{j-1}$. As $\gamma<1/16$, the third assertion of Claim 14 implies that $X_j\cap B_x=\emptyset$. □

We are ready to prove the first statement of Lemma 15. For terminal $v\in K$, Using Claim 18,

$$\Pr[\mathcal{E}]=\Pr\left[\bigvee_{j\leq n+1}C_j\right]=\Pr\left[\bigvee_{j\leq k+1}C_j\right]\leq 2^{14}\gamma L_k\sum_{j\leq k+1}\sum_{Y}\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ln\left(\varphi_k\left(X,Y,v\right)\right)\ .$$

Using a symmetric argument, we can also show the second assertion of Lemma 15. Set $\mu=\Pr\left(\bigvee_{j\leq k+1}C_j\right)$ the probability that B_x is cut by terminal petal. Let $I_n=k+2,\ldots,n+1$ be all the indices of the non-terminal petals. Note that for $j\in I_n$, $1-\delta_j=1-e^{-8\lambda_j\gamma\Delta}=1-e^{-8\cdot128\cdot\gamma L_n\ln\hat{\chi}_j}\leq 2^{10}\gamma L_n\ln\hat{\chi}_j$. Hence for every vertex $x\in X$,

$$\begin{aligned} \Pr[\mathcal{E}] &= \Pr\left[\bigvee_{j\leq n+1}C_j\right] \\ &= \sum_{j\leq n+1}\sum_{Y\subseteq X}\Pr\left[C_j\mid\mathcal{Z}_{Y,j}\right]\cdot\Pr[\mathcal{Z}_{Y,j}] \\ &= \mu+\sum_{j\in I_n}\sum_{Y\subseteq X}\Pr\left[C_j\mid\mathcal{Z}_{Y,j}\right]\cdot\Pr[\mathcal{Z}_{Y,j}] \\ &\leq \mu+\sum_{j\in I_n}\sum_{Y}(1-\delta_j)\left(\Pr\left[\overline{\mathcal{F}_j}\mid\mathcal{Z}_{Y,j}\right]+\frac{2}{\hat{\chi}_j^2}\right)\cdot\Pr[\mathcal{Z}_{Y,j}] \\ &\leq \mu+2^{10}\gamma L_n\left(\sum_{j\in I_n}\sum_{Y}\ln\hat{\chi}_j\cdot\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]+\sum_{j\in I_n}\sum_{Y}\frac{2}{\hat{\chi}_j}\cdot\Pr[\mathcal{Z}_{Y,j}]\right) \end{aligned}$$

where the first inequality is by Claim 17. For Y such that $\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\neq 0$ it holds that $B_x\cap W_{b'_j}\neq\emptyset$. If $x\notin K$, we may assume that $\gamma\leq\frac{1}{2^{10}L_n}$ as otherwise the lemma is trivial. By Claim 13 we have that $B_x\left(x,\frac{\Delta}{2^9L_n}\right)\subseteq W_{b'_j+4\gamma\Delta+\frac{\Delta}{2^7L_n}}\subseteq W_{b'_j+\frac{\Delta}{2^8L_n}+\frac{\Delta}{2^7L_n}}\subseteq W_{b'_j+\frac{R_j}{2L_n}}=W_{b_j}$. We will show that $\ln\hat{\chi}_j\leq$

$6\ln\varphi_n(X,Y,x)$. If $\hat{\chi}_j = e$ this is trivial, hence we will assume that $\hat{\chi}_j = \frac{|X|+1}{|W_a|}$. By the maximality of q_j ($\frac{2|X|}{2^{\log^{1-(q-1)/L_n} n}} \le |W_{a_j}| \le |W_{b_j}| \le \frac{2|X|}{2^{\log^{1-q/L_n} n}}$) we get $\log\hat{\chi}_j = \log\left(\frac{|X|+1}{|W_{a_j}|}\right) \le \log\left(\frac{2|X|}{|W_{a_j}|}\right) \le \log^{1-(q-1)/L_n} n \le 3\log^{1-q/L_n} n \le 3\log\left(\frac{2|X|}{|W_b|}\right) \le 3\log\left(\frac{2|X|}{|B(v,\Delta/2^9 L_n)|}\right) \le 3\left(\log\left(\varphi_n(X,Y,x)\right)+1\right) \le 6\log\left(\varphi_n(X,Y,x)\right)$.

By similar arguments (and definitions) to (16) we get

$$
\begin{aligned}
\sum_{j\in I_n}\sum_{Y}\frac{\Pr[S_{Y,j}]}{\hat{\chi}_j} &\le \sum_{j\in I_n}\sum_{Y:\ B_x\subseteq Y}\ \sum_{(X_0,\ldots,X_s)\ :\ Y=(X_j\cup\cdots\cup X_s\cup X_0)}\Pr[X_0,\ldots,X_s]\frac{|W_{a_j}|}{|X|} \\
&\le \sum_{j\in I_n}\sum_{(X_0,\ldots,X_s)}\Pr[X_0,\ldots,X_s]\frac{|X_j|}{|X|} \\
&= \sum_{(X_0,\ldots,X_s)}\Pr[X_0,\ldots,X_s]\sum_{j\in I_n}\frac{|X_j|}{|X|} \le 1\ .
\end{aligned}
$$

As $1 \le \sum_{Y,j}\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ln\varphi_n(X,Y,x)$ we get

$$
\begin{aligned}
\Pr[\mathcal{E}] &\le \mu + 2^{10}\gamma L_n\left(6\sum_{j\in I_n}\sum_{Y}\ln\left(\varphi_k(X,Y,x)\right)\cdot\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]+2\right) \\
&\le \mu + 2^{13}\gamma L_n\sum_{j\in I_n}\sum_{Y}\ln\left(\varphi_n(X,Y,x)\right)\cdot\Pr\left[\overline{\mathcal{F}_j}\wedge\mathcal{Z}_{Y,j}\right]\ .
\end{aligned}
$$

Hence the second assertion of Lemma 15 follows as well.

Part IV

Average stretch

This part is dedicated to embedding a metric space into a single tree, such that the *average* terminal distortion is bounded by a universal constant. In Section 8 we present embedding of a metric space into an ultrametric, while in Section 9 we present embedding of a general graph into its spanning tree. Both of these results are based on the notation of scaling distortion, as presented in [ABN07].

8 Terminal Scaling Embedding into an Ultrametric

In this section we consider embedding of a metric space into a single tree, such that the *average* terminal distortion is bounded by a universal constant. The main result of this section is

Theorem 18. *Any metric space (X,d), with a set of terminals $K\subseteq X$, embeds into an ultrametric with average terminal distortion $O(1)$.*

We will follow the partition framework of [ABN07]. Roughly speaking, their algorithm finds a certain partition, such that for all values of $\epsilon\in(0,1)$ *simultaneously*, at most ϵ-fraction of the separated pairs suffer distortion which is "too large". Then we recurse this process on both of the parts. Using induction on the number of points we show that at most $O(\epsilon)$-fraction of all pairs suffer too large distortion. This can be used to bound the average distortion over all pairs, by a simple argument that appeared first in [ABN11].

In order to argue about the average over pairs containing a terminal, rather than all pairs, we will require some non-trivial modification of the [ABN07] partition algorithm. In particular, when considering which radius to pick, we have to consider both the number of points and the number of terminals on both sides of the partition, which leads to a more involved case analysis.

8.1 Definitions

Given a metric space (X, d), let $K \subseteq X$ be a subset of terminals. For a non-contractive embedding f of (X, d_X) into some metric space (Y, d_Y), let the distortion of the pair $u, v \in X$ be $\text{dist}_f(u, v) = \frac{d_Y(f(u), f(v))}{d_X(u,v)}$.

For an embedding $f : X \to Y$, the *average − terminal-distortion* of f is $Avg_f(K \times X) = \frac{1}{|K| \cdot |X|} \sum_{v \in K, u \in X \setminus \{v\}} \frac{d_Y(f(u), f(v))}{d_X(u,v)}$. Note that the pairs $K \times K$ are counted twice. (This can affect the average by a factor of 2, so we will ignore that.)

Definition 8. *[Terminal Scaling Distortion]: Given two metric spaces (X, d_X) and (Y, d_Y), an embedding $f : X \to Y$, and a function $\alpha : (0, 1) \to \mathbb{R}^+$, we say that f has terminal scaling distortion α, if for every $\epsilon \in (0, 1)$, at most $\epsilon \cdot |K| \cdot |X|$ pairs $(u, v) \in K \times X$ have distortion greater than $\alpha(\epsilon)$. Formally,*

$$\forall \epsilon \in (0, 1) \ \ |\{(u, v) \in K \times X : \ dist_f(u, v) > \alpha(\epsilon)\}| \le \epsilon \cdot |K| \cdot |X| .$$

For an embedding $f : X \to Y$, the set of pairs distorted by more than $\frac{c}{\sqrt{\epsilon}}$ is defined as $BP_\epsilon(f, Y) = \left\{(u, v) \in K \times X : \ \text{dist}_f(u, v) > \frac{c}{\sqrt{\epsilon}}\right\}$ (BP stands for bad pairs, and c is a universal constant that will be determined later). If the embedding is clear from the context we may write just $BP_\epsilon(Y)$. We will use the function $\alpha(\epsilon) = \frac{c}{\sqrt{\epsilon}}$. In this case, f has terminal scaling distortion $\frac{c}{\sqrt{\epsilon}}$ means that $\forall \epsilon \in (0, 1) \ \ |BP_\epsilon(f)| \le \epsilon \cdot |K| \cdot |X|$.

8.2 Construction

Set $c = 2^{11}$. In this section we will show that every metric (X, d), with any subset of terminals $K \subseteq X$, has an embedding f into an ultrametric with terminal scaling distortion $\frac{c}{\sqrt{\epsilon}}$. This implies that $Avg_f(K \times X)$ is a constant. The following Lemma is a simplified version of lemma of [ABN11].

Lemma 19. *For every finite metric (X, d) and a subset of terminals K, if f is an embedding from X into a metric space (Y, d_Y) with terminal scaling distortion $\frac{c}{\sqrt{\epsilon}}$, it holds that $Avg_f(K \times X) \le O(1)$.*

Proof. By definition of terminal scaling distortion for every $\epsilon > 0$, at most $\epsilon \cdot |K| \cdot |X|$ pairs have distortion greater than $\frac{c}{\sqrt{\epsilon}}$. Define for $i \ge 2$, $A_i = \left\{(v, u) \in K \times X : \ \sqrt{2^{i-1}}c < \text{dist}_f(u, v) \le \sqrt{2^i}c\right\}$ and $A_1 = \left\{(v, u) \in K \times X : \ \text{dist}_f(u, v) \le \sqrt{2}c\right\}$. By the terminal scaling distortion of f, $|A_i| \le \frac{1}{2^{i-1}} \cdot |K| \cdot |X|$ (the inequality hold trivially also in the case $i = 1$). Each pair in $K \times X$ is in some

A_i, hence $\bigcup A_i = K \times X$. We conclude:

$$\begin{aligned} Avg_f\left(K \times X\right) &\le \frac{1}{|K| \cdot |X|} \sum_{i \ge 1} \sum_{\{v,u\} \in A_i} \mathrm{dist}_f(u,v) \le \frac{1}{|K| \cdot |X|} \sum_{i \ge 1} |A_i| \cdot \sqrt{2^i} c \\ &\le \sum_{i \ge 1} \frac{1}{2^{i-1}} \cdot \sqrt{2^i} c = 2c \cdot \sum_{i \ge 1} 2^{-\frac{i}{2}} \le 5c = O\left(1\right). \end{aligned}$$

as required. □

Let (X, d) be an n point metric, fix a subset of terminals $K \subseteq X$ of size k and let $\Lambda = \mathrm{diam}\left(X\right)$. In what follows we will always assume that $\epsilon \ge \frac{1}{nk}$, because the distortion bound for $\epsilon < \frac{1}{nk}$ is required of all pairs so this case is subsumed by $\epsilon = \frac{1}{nk}$. The ultrametric will be represented by a binary tree which is induced by a laminar hierarchical partition of X; each node u corresponds to a subset $X_u \subseteq X$, such that if v, w are the children of u in the ultrametric then $X_v \cap X_w = \emptyset$, $X_v \cup X_w = X_u$. Furthermore the root r has $X_r = X$, and each leaf corresponds to a singleton.

The high level construction of T is as follows: find a partition P of X into X_1 and $X_2 = X \setminus X_1$, the root of T will be labeled Λ, and its children will be the trees T_1, T_2 formed recursively from X_1 and X_2 respectively. Since every node corrseponding to a set Y is labeled by diam(Y), it always holds that $d_T(x, y) \ge d(x, y)$, so it remains to bound $BP_\epsilon\left(T\right)$ by $\epsilon \cdot |K| \cdot |X|$. For a partition $P = (P_1; P_2)$ let

$$BPS_\epsilon\left(P_1, P_2\right) = \left\{(u,v) \in K \times X : |\{u,v\} \cap P_1| = 1,\ \frac{\Lambda}{d\left(u,v\right)} > \frac{c}{\sqrt{\epsilon}}\right\}$$

be all the pairs that will have distortion greater than $\frac{c}{\sqrt{\epsilon}}$ due to the partition $(P_1; P_2)$ (BPS stands for bad pairs due to separation). Note that $BPS_\epsilon\left(P_1, P_2\right)$ contains all the pairs in $BP_\epsilon\left(T\right)$ who are separated in the current stage of the ultrametric construction, because their distance is fixed to be Λ.

Lemma 20. *For any finite metric space (X, d) of size n and a subset of terminals K of size k, there is a partition $P = (P_1; P_2)$ where $|P_i| = n_i$, $|P_i|_k = k_i$ for $i \in \{1, 2\}$, such that $n_1 \notin \{0, n\}$ and $(P_1; P_2)$ fulfills at least one of the following requirements:*

1. *For every $\epsilon \in (0, 1)$, $|BPS_\epsilon\left(P_1, P_2\right)| \le \frac{\epsilon}{2} \cdot \left(k_1 \cdot n + k \cdot n_1\right)$ and $k_1 \le \frac{k}{2}$, $n_1 \le \frac{n}{2}$.*
2. *No pairs have distortion greater than c, in other words, for every $\epsilon \in (0, 1)$, $BPS_\epsilon\left(P_1, P_2\right) = \emptyset$ (but no assumption on n_1, k_1).*

We will prove Lemma 20 in Section 8.3. Using Lemma 20, we can prove the following Theorem:

Theorem 19. *For any finite metric space (X, d) and subset of terminals K, there is an embedding f from X into an ultrametric (U, d_U) with terminal scaling distortion $\frac{c}{\sqrt{\epsilon}}$.*

Proof. Let $|X| = n$, $|K| = k$. We prove by induction on $|X| = n$ that there is an embedding f into an ultrametric, such that for every $\epsilon \in (0, 1)$, at most ϵnk pairs have distortion greater than $\frac{c}{\sqrt{\epsilon}}$. The base case, where $|X| = 2$, holds because the unique pair realizes the diameter, thus $\mathrm{dist}_f\left(u, v\right) = 1$. Assume that for any metric space with $m < n$ points and $j \le k$ terminals, we can find an ultrametric such that for every $\epsilon \in (0, 1)$ the number of pairs distorted by more than

$\frac{c}{\sqrt{\epsilon}}$ is bounded by ϵmj. Now consider the metric space (X, d) with $|X| = n$ and subset $K \subset X$ of terminals of size k. Let P be a partition $(P_1; P_2)$ guaranteed to exist by Lemma 20. For $i \in \{1, 2\}$, set $|P_i| = n_i$ and $|P_i|_k = k_i$. By induction, there exist ultrametrics T_1, T_2 for the metrics induced by P_1, P_2 such that for any $\epsilon \in (0, 1)$, $|BP_\epsilon(T_i)| \leq \epsilon n_i k_i$ for $i \in \{1, 2\}$. We obtain our ultrametric T by adding a common root to T_1 and T_2 with label Λ (diameter of X). Of course the resulting tree T is an ultrametric because the diameters of T_1,T_2 are at most Λ.

The partition P fulfills one of two requirements. Either $k_1 \leq \frac{k}{2}$, $n_1 \leq \frac{n}{2}$ and $|BPS_\epsilon(P_1, P_2)| \leq \frac{\epsilon}{2} \cdot (k_1 \cdot n + k \cdot n_1)$. Then the total number of bad pairs can be bounded as follows:

$$
\begin{aligned}
|BP_\epsilon(T)| &\leq |BPS_\epsilon(P_1, P_2)| + |BP_\epsilon(T_1)| + |BP_\epsilon(T_2)| \\
&\leq \frac{\epsilon}{2} \cdot (nk_1 + n_1 k) + \epsilon n_1 k_1 + \epsilon n_2 k_2 \\
&= \frac{\epsilon}{2} \cdot ((nk_1 + n_1 k) + 2n_1 k_1 + 2(n - n_1)(k - k_1)) \\
&= \frac{\epsilon}{2} \cdot (4n_1 k_1 - n_1 k - nk_1 + 2nk) \\
&\leq \frac{\epsilon}{2} \cdot (n_1(2k_1 - k) + k_1(2n_1 - n) + 2nk) \leq \epsilon nk.
\end{aligned}
$$

Otherwise, $BP_\epsilon(P_1, P_2) = \emptyset$, hence

$$
\begin{aligned}
|BP_\epsilon(T)| &\leq |BPS_\epsilon(P_1, P_2)| + |BP_\epsilon(T_1)| + |BP_\epsilon(T_2)| \\
&\leq \epsilon n_1 k_1 + \epsilon n_2 k_2 \leq \epsilon(n_1 + n_2)(k_1 + k_2) \leq \epsilon nk
\end{aligned}
$$

as required. □

Note that by Lemma 19 and Theorem 19, for every finite metric (X, d) and subset of terminals K, there is an embedding f from X into an ultrametric (U, d_U) such that $Avg_f(K \times X) \leq O(1)$.

8.3 Proof of Lemma 20

We start by describing the partition algorithm, then prove its correctness by considering small and large values of ϵ separately, in Claims 24,25 respectively. Set $a = 2^8$, a constant which we will use to distinguish between large and small ϵ's. Recall another constant $c = 2^{11}$ that we already set. Let $|X| = n$, $|X|_k = k$ and $\mathrm{diam}(X) = \Lambda$.

Partition Algorithm. First we find $v \in X$ such that $\left|B^o\left(v, \frac{\Lambda}{8}\right)\right| \leq \frac{n}{2}$ and $\left|B^o\left(v, \frac{\Lambda}{8}\right)\right|_k \leq \frac{k}{2}$, which will be the "center" of P_1 in our partition. Let $T(u, r_1, r_2) = \{x | \; r_1 \leq d(x, u) < r_2\}$ be all the points $x \in X$ such that $d(x, u) \in [r_1, r_2)$. Let $u, q \in V$ such that $d(u, q) = \Lambda$. W.l.o.g $\left|B^o\left(u, \frac{\Lambda}{2}\right)\right| \leq \frac{n}{2}$. If $T\left(u, \frac{1}{4}\Lambda, \frac{3}{8}\Lambda\right) = \emptyset$ then take $P_1 = B^o\left(u, \frac{1}{4}\Lambda\right)$, $P_2 = X \backslash P_1$, no pair will have distortion greater than $\frac{\Lambda}{\frac{1}{8}\Lambda} = 8 < c$ (since for every $x \in P_1$, $y \in P_2$, $d(x, y) \geq \frac{1}{8}\Lambda$). Hence for all $\epsilon > 0$, $BPS_\epsilon(P_1, P_2) = \emptyset$ and the second requirement of Lemma 20 is satisfied. Otherwise, take $w \in T\left(u, \frac{1}{4}\Lambda, \frac{3}{8}\Lambda\right)$. Then $\left|B^o\left(w, \frac{\Lambda}{8}\right)\right| < \left|B^o\left(u, \frac{\Lambda}{2}\right)\right| \leq \frac{n}{2}$. As $d(u, w) \geq \frac{1}{4}\Lambda$, $B^o\left(u, \frac{\Lambda}{8}\right) \cap B^o\left(w, \frac{\Lambda}{8}\right) = \emptyset$. Therefore $\left|B^o\left(u, \frac{\Lambda}{8}\right)\right|_k \leq \frac{k}{2}$ or $\left|B^o\left(w, \frac{\Lambda}{8}\right)\right|_k \leq \frac{k}{2}$. Let $v \in \{u, w\}$ be such that $\left|B\left(v, \frac{\Lambda}{8}\right)\right|_k \leq \frac{k}{2}$, it also holds that $\left|B\left(v, \frac{\Lambda}{8}\right)\right| \leq \frac{n}{2}$, so v is the vertex we desired. Let

$$
\hat{\epsilon} = \max\left\{\epsilon | \; \left|B\left(v, \frac{\sqrt{\epsilon}}{16}\Lambda\right)\right| \geq \epsilon \cdot n \text{ or } \left|B\left(v, \frac{\sqrt{\epsilon}}{16}\Lambda\right)\right|_k \geq \epsilon \cdot k\right\}.
$$

Note that the maximum is indeed attained because the metric space is finite. Since a ball always contains at least one point we have that $\hat{\epsilon} \geq \frac{1}{n}$. As $B(v, \frac{\sqrt{\epsilon}}{16}\Lambda) \subseteq B^o(v, \frac{\Lambda}{8})$, and by the choice of v, the ball $B(v, \frac{\sqrt{\epsilon}}{16}\Lambda)$ will never contain more than $\frac{n}{2}$ points, and similarly $B(v, \frac{\sqrt{\epsilon}}{16}\Lambda) \cap K$ will never contain more then $\frac{k}{2}$ terminals, hence $\hat{\epsilon} \leq \frac{1}{2}$. For $\epsilon > 0$, set $\delta_\epsilon = \frac{\sqrt{\epsilon}\Lambda}{c}$. Define the intervals $\hat{S} = \left[\frac{\sqrt{\hat{\epsilon}}}{16}\Lambda, \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda\right]$, $S = \left[\frac{\sqrt{\hat{\epsilon}}}{16}\Lambda + \delta_{a\hat{\epsilon}}, \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda - \delta_{a\hat{\epsilon}}\right]$, the length of S, $s = |S| = \left(\frac{1}{16} - 2\frac{\sqrt{a}}{c}\right)\sqrt{\hat{\epsilon}}\Lambda$, and the shell $Q = \left\{w \mid d(v,w) \in \hat{S}\right\}$. The partition P is defined by carefully choosing a certain $r \in S$ and letting $P_1 = B(v,r)$ and $P_2 = X \setminus P_1$. The following claim will be useful,

Claim 21. $\left|B\left(v, \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda\right)\right| \leq 4\hat{\epsilon}n$ *and* $\left|B\left(v, \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda\right)\right|_k \leq 4\hat{\epsilon}k$

Proof. If $\hat{\epsilon} \leq \frac{1}{4}$, by the maximality of $\hat{\epsilon}$, $\left|B\left(v, \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda\right)\right| = \left|B\left(v, \frac{\sqrt{4\hat{\epsilon}}}{16}\Lambda\right)\right| \leq 4\hat{\epsilon}n$, similarly for the terminals. On the other hand, if $\hat{\epsilon} > \frac{1}{4}$ then $4\hat{\epsilon} \geq 1$ and the claim follows trivially. □

We will now show that a certain choice of $r \in S$ will produce a partition that satisfies the condition of Lemma 20 for all $\epsilon \in (0, a\hat{\epsilon}]$. For any $r \in S$ and $\epsilon \leq a\hat{\epsilon}$, let $S_r(\epsilon) = (r - \delta_\epsilon, r + \delta_\epsilon) \subseteq \hat{S}$, $s(\epsilon) = 2\delta_\epsilon$ (the length of the interval $S_r(\epsilon)$) and $Q_r(\epsilon) = \{w \mid d(v,w) \in S_r(\epsilon)\}$. Notice that in the partition $P_r = (P_1^r; P_2^r) = (B(v,r), X \backslash B(v,r))$, only pairs in $P_1^r \times P_2^r$ who are both in $Q_r(\epsilon)$ can have distortion greater than $\frac{c}{\sqrt{\epsilon}}$. Indeed, assume there exist a pair $\{x,y\} \in P_1^r \times P_2^r \backslash Q_r(\epsilon) \times Q_r(\epsilon)$, by the triangle inequality $d(x,y) > \delta_\epsilon$ and hence $\text{dist}_f(u,v) < \frac{\Lambda}{\delta_\epsilon} = \frac{c}{\sqrt{\epsilon}}$. Define that property $A_r(\epsilon)$ holds if the number of pairs containing a terminal in the shell $Q_r(\epsilon)$ is sufficiently small, which will imply that cutting at radius r is "good" for ϵ. Formally,

$$A_r(\epsilon) \text{ holds } \iff |Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{2}\hat{\epsilon}nk$$

Claim 22. *For every $r \in \bar{S}$ and every $\epsilon \in (0,1)$, if $A_r(\epsilon)$ holds, then $|BPS_\epsilon(P_1,P_2)| \leq \frac{\epsilon}{2} \cdot (n \cdot k_1 + k \cdot n_1)$.*

Proof. Observe that $r \geq \frac{\sqrt{\hat{\epsilon}}}{16}\Lambda$ and thus by the definition of $\hat{\epsilon}$, $|B(v,r)| \geq \hat{\epsilon} \cdot n$ or $|B(v,r)|_k \geq \hat{\epsilon} \cdot k$. Recall that by the choice of v, $|X \backslash B(v,r)| \geq \frac{n}{2}$ and $|X \backslash B(v,r)|_k \geq \frac{k}{2}$. If it is the case that $|B(v,r)| \geq \hat{\epsilon}n$, then $A_r(\epsilon)$ implies

$$\begin{aligned} |BPS_\epsilon(P_1,P_2)| &\leq |Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{2} \cdot k \cdot \hat{\epsilon} \cdot n \leq \frac{\epsilon}{2} \cdot k \cdot |B(v,r)| \\ &\leq \frac{\epsilon}{2} \cdot (n \cdot |B(v,r)|_k + k \cdot |B(v,r)|) = \frac{\epsilon}{2} \cdot (n \cdot k_1 + k \cdot n_1). \end{aligned}$$

On the other hand, $|B(v,r)|_k \geq \hat{\epsilon} \cdot k$ implies that

$$\begin{aligned} |BPS_\epsilon(P_1,P_2)| &\leq |Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{2} \cdot \hat{\epsilon} \cdot k \cdot n \leq \frac{\epsilon}{2} \cdot (|B(v,r)|_k \cdot n) \\ &\leq \frac{\epsilon}{2} \cdot (n \cdot |B(v,r)|_k + k \cdot |B(v,r)|) \leq \frac{\epsilon}{2} \cdot (n \cdot k_1 + k \cdot n_1). \end{aligned}$$

□

We will find $r \in S$ such that for every $\epsilon \in (0, a\hat{\epsilon}]$ $A_r(\epsilon)$ holds, thus the condition of Lemma 20 is satisfied for any such ϵ. The proof is based on the `Delete-bad-intervals`(S, A, δ, σ) iterative process that greedily deletes the "worst" interval in S. The procedure input is an interval S, property A, a function $\delta : (0,1) \to \mathbb{R}^+$ and a constant $\sigma \in [0,1]$. We will refer to A as a boolean function with two inputs, note that in $A_r(\epsilon)$ the inputs are $r \in S$ and $\epsilon \in (0,1)$. The procedure returns a subset of S.

Algorithm 7 $I_t =$ `Delete-bad-intervals`(S, A, δ, σ)

1: Let $I_0 = S$, and $j = 1$;
2: If for all $r \in I_{j-1}$ and for all $\epsilon \in (0,\sigma)$ property $A_r(\epsilon)$ holds, then set $t = j-1$, stop the iterative process and output I_t;
3: Let $\mathcal{S}_j = \{(\epsilon, r) \ : \ r \in I_{j-1}, \epsilon \in (0,\sigma), \neg A_r(\epsilon)\}$. We greedily remove the interval $(r-\delta_\epsilon, r+\delta_\epsilon)$ that has maximal ϵ. Formally, let r_j, ϵ_j be parameters such that $(\epsilon_j, r_j) \in \mathcal{S}_j$ and $\epsilon_j = \max\{\epsilon \ : \ \exists (\epsilon_j, r_j) \in \mathcal{S}_j\}$.
4: Set $I_j = I_{j-1} \setminus (r-\delta_\epsilon, r+\delta_\epsilon)$, set $j = j+1$, and go to line 2;

The next claim bounds the measure of all the intervals deleted by the `Delete-bad-intervals` procedure.

Claim 23. *Fix an interval S, constants $\hat{\epsilon}$,α,β, a function $\delta(\epsilon) = \alpha\sqrt{\epsilon}$, function $Q : S \times (0,1) \to P(X)$ ($P(X)$ is the power set of X) such that $\tilde{Q} = \cup_{(r,\epsilon)} Q_r(\epsilon)$, $\left|\tilde{Q}\right| \le 4\hat{\epsilon}n$, $\left|\tilde{Q}\right|_k \le 4\hat{\epsilon}k$. Also fix a set of pairs $(\epsilon_1, r_1), \ldots, (\epsilon_t, r_t)$ such that each point $x \in X$ may be included in at most two sets $Q_{r_i}(\epsilon_i), Q_{r_j}(\epsilon_j)$, and property A which is a boolean function from $S \times (0,1)$, such that A holds iff $|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \le \frac{\epsilon}{\beta}\hat{\epsilon}nk$, and A does not holds for any pair (ϵ_i, r_i). Then it holds that $\sum_{i \le t} 2\delta_{\epsilon_i} \le 32\alpha\sqrt{\beta}\sqrt{\hat{\epsilon}}$.*

Proof. From the assumption that each point $x \in X$ may be included in at most two sets $Q_{r_i}(\epsilon_i), Q_{r_j}(\epsilon_j)$ it follows that

$$\sum_i |Q_{r_i}(\epsilon_i)| \le 2\left|\tilde{Q}\right| \le 8\hat{\epsilon}n \tag{17}$$

and similarly

$$\sum_i |Q_{r_i}(\epsilon_i)|_k \le 8\hat{\epsilon}k. \tag{18}$$

For every i, $|Q_{r_i}(\epsilon_i)| > \sqrt{\epsilon_i \frac{\hat{\epsilon}}{\beta}}n$ or $|Q_{r_i}(\epsilon_i)|_k > \sqrt{\epsilon_i \frac{\hat{\epsilon}}{\beta}}k$, as otherwise $A_r(\epsilon)$ holds. Define $\boldsymbol{N} = \left\{i \ : \ |Q_{r_i}(\epsilon_i)| > \sqrt{\epsilon_i \frac{\hat{\epsilon}}{\beta}}n\right\}$ and similarly $\boldsymbol{K} = \left\{i \ : \ |Q_{r_i}(\epsilon_i)|_k > \sqrt{\epsilon_i \frac{\hat{\epsilon}}{\beta}}k\right\}$, $\boldsymbol{N} \cup \boldsymbol{K} = \{1, \ldots t\}$. We show that

$$\sum_{i=1}^{t} \sqrt{\epsilon_i} \quad \le \quad 16\sqrt{\beta\hat{\epsilon}} \tag{19}$$

by considering two cases:

- $\sum_{i \in \boldsymbol{N}} \sqrt{\epsilon_i} \ge \sum_{i \in \boldsymbol{K}} \sqrt{\epsilon_i}$
  This implies that $\sum_{i \in \boldsymbol{N}} \sqrt{\epsilon_i} \ge \frac{1}{2}\sum_i \sqrt{\epsilon_i}$. We conclude:

$$\sum_i |Q_{r_i}(\epsilon_i)| \ge \sum_{i \in \boldsymbol{N}} |Q_{r_i}(\epsilon_i)| > \sum_{i \in \boldsymbol{N}} \sqrt{\epsilon_i \frac{\hat{\epsilon}}{\beta}}n \ge \frac{1}{2}\sqrt{\frac{\hat{\epsilon}}{\beta}}n \sum_i \sqrt{\epsilon_i}$$

combining this with (17) proves (19) in this case.

- $\sum_{i\in \boldsymbol{K}} \sqrt{\epsilon_i} \geq \sum_{i\in \boldsymbol{N}} \sqrt{\epsilon_i}$
  This implies that $\sum_{i\in \boldsymbol{K}} \sqrt{\epsilon_i} \geq \frac{1}{2}\sum_i \sqrt{\epsilon_i}$. We conclude:

$$\sum_{i\leq t} |Q_{r_i}(\epsilon_i)|_k \geq \sum_{i\in \boldsymbol{K}} |Q_{r_i}(\epsilon_i)|_k \geq \sum_{i\in \boldsymbol{K}} \sqrt{\epsilon_i \frac{\hat{\epsilon}}{\beta} k} \geq \frac{1}{2}\sqrt{\frac{\hat{\epsilon}}{\beta} k} \sum_i \sqrt{\epsilon_i}$$

combining this with (18) proves (19) in this case.

Finally

$$\sum_{i\leq t} 2\delta_{\epsilon_i} = \sum_{i\leq t} 2\alpha\sqrt{\epsilon_i} \leq 2\alpha 16\sqrt{\beta\hat{\epsilon}} = 32\alpha\sqrt{\beta}\sqrt{\hat{\epsilon}}.$$

□

Claim 24. *There exists some $r \in S$ such that property $A_r(\epsilon)$ holds for all $\epsilon \leq a\hat{\epsilon}$ simultaneously.*

Proof. Let $I_t = \texttt{Delete-bad-intervals}(S, A, \delta, a\hat{\epsilon})$. We will conclude from Claim 23 that $I_t \neq \emptyset$ and hence an appropriate value $r \in S$ can be found. Let $(\epsilon_1, r_1), \ldots, (\epsilon_t, r_t)$ be all the selected variables. Any number $y \in \hat{S}$ can be covered by at most two intervals $S_{r_j}(\epsilon_j)$, $S_{r_i}(\epsilon_i)$ for some $1 \leq j < i \leq t$. This holds because once y is covered from the left by $S_{r_j}(\epsilon_j)$ (that is, $r_j \leq y$) then this interval is removed in step 3. from I_j. By maximality of ϵ_j, for any $j < i \leq t$ we have that $s(\epsilon_i) \leq s(\epsilon_j)$, so as $\epsilon_i \in I_j$ it must be that the interval $S_{r_i}(\epsilon_i)$ covers y from the right (that is, $r_i \geq y$), and no other interval will cover y in the remainder of the process. Observe that $S_{r_j}(\epsilon_j) \subseteq \hat{S}$ for any $1 \leq j \leq t$. For $x \in Q$, let $r = d(v, x)$. As $x \in Q_{r_j}(\epsilon_j)$ iff $r \in S_{r_j}(\epsilon_j)$, it follows that any $x \in Q$ appears in at most two sets $Q_{r_j}(\epsilon_j)$, $Q_{r_i}(\epsilon_i)$. Note that $Q = \bigcup_{(r,\epsilon)} Q_r(\epsilon) \subseteq B\left(v, \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda\right)$, and from Claim 21, $|Q| \leq 4\hat{\epsilon}n$, $|Q|_k \leq 4\hat{\epsilon}k$. Hence by Claim 23, with parameters $\alpha = \frac{\Lambda}{c}$, $\beta = 2$, it holds that the measure of all the intervals deleted by the `Delete-bad-intervals` procedure is bounded by (recall that $c = 2^{11}, a = 2^8$)

$$\sum_{i\leq t} s(\epsilon_i) = \sum_{i\leq t} 2\delta_{\epsilon_i} \leq 32\alpha\sqrt{\beta}\sqrt{\hat{\epsilon}} = 32\frac{\Lambda}{c}\sqrt{2}\sqrt{\hat{\epsilon}} = \frac{\Lambda}{2^{5.5}}\sqrt{\hat{\epsilon}}$$

But since the length of $I_0 = S$ is

$$s \quad = \quad \left(\frac{1}{16} - 2\frac{\sqrt{a}}{c}\right)\sqrt{\hat{\epsilon}}\Lambda = \left(\frac{1}{2^4} - \frac{1}{2^6}\right)\sqrt{\hat{\epsilon}}\Lambda > \frac{1}{2^{5.5}}\sqrt{\hat{\epsilon}}\Lambda,$$

it is impossible that the entire interval I_0 was removed, therefore $I_t \neq \emptyset$, and in fact any $r \in I_t$ satisfies the first condition of the Lemma. □

Finally, we show that any choice of $r \in S$ will produce a partition that satisfies the first condition of Lemma 20 for all $\epsilon \in (a\hat{\epsilon}, 1)$.

Claim 25. *If $\epsilon \in (a\hat{\epsilon}, 1)$, $r \in S$ and $P_r = (P_1^r; P_2^r) = (B(v, r), X \backslash B(v, r))$ then $|BPS_\epsilon(P_1^r, P_2^r)| < \frac{\epsilon}{2} \cdot (k_1 \cdot n + k \cdot n_1)$.*

Proof. By the triangle inequality only pairs $\{x,y\}$ such that $x \in B(v,r)$ and $y \in B(v, r+\delta_\epsilon)$ can be included in $BPS_\epsilon(P_1^r, P_2^r)$. As $r \le \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda$, we get that

$$\begin{aligned}
|B(v, r+\delta_\epsilon)| &= \left|B\left(v, r+\frac{\sqrt{\epsilon}}{c}\Lambda\right)\right| \le \left|B\left(v, \frac{\sqrt{\hat{\epsilon}}}{8}\Lambda + \frac{\sqrt{\epsilon}}{c}\Lambda\right)\right| \\
&\le \left|B\left(v, \frac{\sqrt{\epsilon}}{\sqrt{a}8}\Lambda + \frac{\sqrt{\epsilon}}{c}\Lambda\right)\right| = \left|B\left(v, \frac{\left(\frac{2}{\sqrt{a}} + \frac{16}{c}\right)\sqrt{\epsilon}}{16}\Lambda\right)\right| \\
&\overset{(1)}{<} \left(\frac{2}{\sqrt{a}} + \frac{16}{c}\right)^2 \epsilon \cdot n = \left(\frac{1}{2^3} + \frac{1}{2^7}\right)^2 \epsilon \cdot n < \frac{1}{2}\epsilon \cdot n
\end{aligned}$$

where (1) holds by the fact that $\left(\frac{2}{\sqrt{a}} + \frac{16}{c}\right)^2 \epsilon \ge \frac{4}{a} \cdot a \cdot \hat{\epsilon} > \hat{\epsilon}$ and the maximality of $\hat{\epsilon}$. Similarly $|B(v, r+\delta_\epsilon)|_k < \frac{1}{2}\epsilon \cdot k$. We get that

$$\begin{aligned}
BPS_\epsilon(P_1, P_2) &\le |B(v,r)| \cdot |B(v, r+\delta_\epsilon)|_k + |B(v,r)|_k \cdot |B(v, r+\delta_\epsilon)| \\
&< \frac{1}{2}\epsilon \cdot k \cdot n_1 + \frac{1}{2}\epsilon \cdot n \cdot k_1 = \frac{\epsilon}{2} \cdot (k \cdot n_1 + n \cdot k_1)
\end{aligned}$$

□

9 Terminal Scaling Embedding into a Spanning Tree

In this section we consider embedding of graphs into their spanning tree, such that the *average* terminal distortion is bounded by a universal constant. The main result of this section is

Theorem 20. *Any graph $G = (V, E)$, with a set of terminals $K \subseteq V$, has a spanning tree with average terminal distortion $O(1)$.*

We shall take inspiration from the techniques of the ultrametrics case in Section 8. In particular, we use a similar partitioning scheme, in which for every $\epsilon \in (0,1)$, at most an $\epsilon-$fraction of the pairs that contain a terminal and are separated by the partition are "bad". However, obtaining a subgraph tree is considerably more involved, and requires control over the radius of the final tree (which is trivial in the ultrametric case). While [ABN07] used a complicated construction and argument in order to bound the final radius of the tree, we can use the more recently developed petal decomposition method of [AN12], which arguably simplifies the proof of [ABN07]. We assume that the reader is familiar with petal decomposition presented in Section 7, which constructs a spanning tree with radius proportional to that of the original graph, while allowing large freedom in the choice of radii for the clusters (called petals).

Set $c = 2^{18}$. We recall some former definitions (in addition to some new ones): given a weighted graph $G = (V, E, w)$, a subset of terminals $K \subseteq V$ and a spanning tree T, the average terminal distortion of T is $Avg_T(K \times V) = \frac{1}{|K|\cdot|V|}\sum_{(v,u)\in K\times V} \mathrm{dist}_T(u,v)$. For $\epsilon > 0$, let $BP_\epsilon(T) = \left\{(u,v) \in K \times V :\ \mathrm{dist}_T(u,v) \ge \frac{c}{\sqrt{\epsilon}}\right\}$ be the set of all pairs with distortion greater than $\frac{c}{\sqrt{\epsilon}}$. We say that an embedding f has terminal scaling distortion $\frac{c}{\sqrt{\epsilon}}$ if for all $\epsilon \in (0,1)$, $|BP_\epsilon(f)| \le \epsilon \cdot |K| \cdot |V|$. Note that as our embedding f is the identity into a spanning tree T, we consider T as both spanning tree and embedding, and denote $|BP_\epsilon(T)|$ instead of $|BP_\epsilon(f)|$.

Algorithm 8 $T = \texttt{hierarchical-petal-decomposition}(G[X], x_0, t)$

```
If |X| = 1, then return G[X];
Call (X_0, ..., X_s, (y_1, x_1), ..., (y_s, x_s), t_0, ..., t_s) = petal-decomposition(G[X], x_0, t);
for j ∈ [0, ..., s] do
    Let T_j = hierarchical-petal-decomposition(G[X_j], x_j, t_j);
end for
return The tree T, formed by connecting T_0, ..., T_s using the edges {y_1, x_1}, ..., {y_s, x_s};
```

Recall that the radius of G with respect to center vertex x_0 is $\Delta_{G,x_0} = \max_{v \in V} d_G(x_0, v)$, we often will omit the subscript and write just Δ if x_0 and G are clear from the context. For a graph G and two vertices v, u, $P_{v,u}$ is the shortest path from v to u in G.

The only (major) change in the petal decomposition algorithm presented here, compared to the algorithm from [AN12], is the method of choice for the radius in the `create-petal` procedure. Hence, Facts 1,2,3,4 and 5 remain valid.

Here are two of the main technical difficulties that we will face, which do not appear in the ultrametric case:

1. The petal decomposition algorithm picks the vertex v around which we will grow the petal, and we will not be able to assume that in some area of v there are at most $\frac{n}{2}$ vertices and $\frac{k}{2}$ terminals.

2. For a partition (P_1, P_2) of the graph, not only the separated pairs may suffer large distortion due to the partition, but also pairs in the same part (P_1 for example), whose shortest path uses vertices from the second part.

We will deal with the first difficulty by case analysis, according to the density of terminals and vertices around v, and with the second difficulty by amortized analysis, by bounding the total number of such pairs.

Our tree, is returned by the call of `hierarchical-petal-decomposition` algorithm with parameters G, x_0, x_0 for arbitrary x_0.

9.1 Analysis

In this section we present the procedure for choosing the radius r in line 3 of the `create-petal` algorithm and its analysis. Define $a = 2^6$ and recall $c = 2^{18}$. Let $G = (V, E, w)$ be a weighted graph, $K \subseteq V$ subset of terminals, and a root vertex $x_0 \in V$. Suppose G have n vertices and k terminals. For a partition (P_1, P_2) of V, we set $n_1 = |P_1|$, $k_1 = |P_1 \cap K|$ and similarly n_2, k_2. Let $\Delta = \Delta_{G,x_0}$. For $\epsilon > 0$, set $\delta_\epsilon = \frac{32}{c}\sqrt{\epsilon}\Delta$. We define the sets of pairs of vertices with large distortion with respect to ϵ. For a partition (P_1, P_2) of V, let $BPS_\epsilon(P_1, P_2) = \left\{(u, v) \in K \times V : |\{u, v\} \cap P_1| = 1,\ d_G(u, v) \le 8\frac{\sqrt{\epsilon}}{c}\Delta\right\}$ be all the pairs which will have too large distortion due to the partition of V to P_1 and P_2. Recall that in the previous section we had similar definition. BPS stands for Bad Pairs due to Separation. For a subset P_i of V, restricting G to P_i may increase the distance between vertices in P_i. Thus any such pair is in danger

Algorithm 9 $(X_0, \ldots, X_s, \{y_1, x_1\}, \ldots, \{y_s, x_s\}, t_0, \ldots, t_s) =$ `petal-decomposition`$(G[X], x_0, t)$

Let $\Delta = \Delta_{x_0,X}$; Let $Y_0 = X$; Set $j = 1$;
if $d_X(x_0, t) \geq 5\Delta/8$ **then**
  Create a special first petal:
  Let $(X_1, x_1) =$ `create-petal`$(X, X, t, x_0, [d_X(x_0, t) - 5\Delta/8, d_X(x_0, t) - \Delta/2])$;
  Let $Y_1 = Y_0 \setminus X_1$;
  Let y_1 be the neighbor of x_1 on $P_{x_0 t}$ (the one closer to x_0); Set $t_0 = y_1$, $t_1 = t$;
  Set $j = 2$;
else
  Set $t_0 = t$;
end if
Creating the petals:
while $Y_{j-1} \setminus B_X(x_0, 3\Delta/4) \neq \emptyset$ **do**
  Let $t_j \in Y_{j-1}$ be an arbitrary point satisfying $d_X(x_0, t_j) > 3\Delta/4$;
  Let $(X_j, x_j) =$ `create-petal`$(X, Y_{j-1}, t_j, x_0, [0, \Delta/8])$;
  Let $Y_j = Y_{j-1} \setminus X_j$;
  For each edge $e \in P_{x_j t_j}$, set its weight to be $w(e)/2$;
  Let y_j be the neighbor of x_j on $P_{x_0 t_j}$ (the one closer to x_0);
  Let $j = j + 1$;
end while
Let $s = j - 1$;
Creating the stigma X_0:
Let $X_0 = Y_s$;

Algorithm 10 $(W, x) =$ `create-petal`$(X, Y, t, x_0, [lo, hi])$

Let $R = hi - lo$;
Let $W_r = \bigcup_{p \in P_{x_0 t}:\ d_Y(p,t) \leq r} B_{(Y, \rho(Y, x_0, p))}(p, (r - d_Y(p, t))/2)$;
Pick $r \in [lo, hi]$ according to Lemma 29;
Let $r' \leq r$ be the maximal such that there exists a point $p_{r'}$ of distance r' from t on $P_{x_0 t}$;
return $(W_r, p_{r'})$;

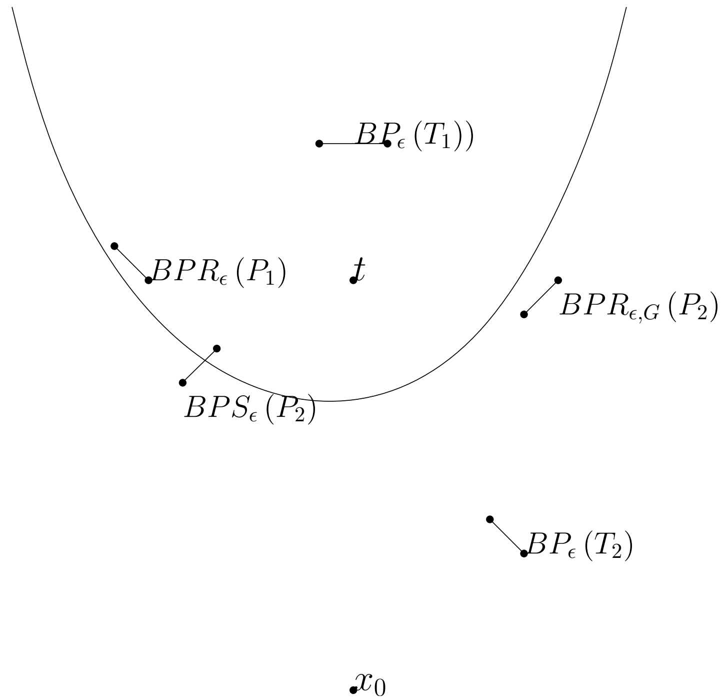


Figure 10: Example of possible members in each of the sets. Note that for each edge in $BPS_\epsilon(P_1, P_2)$ the vertices are in different sets of the partition. The edges in $BPR_\epsilon(P_1)$ are close to the boundary, and edges in $BP_\epsilon(T_1)$ can be found in any place inside P_1.

of having distortion which is too large, and we will count some of these pairs as bad pairs. Let $BPR_\epsilon(P_i) = \left\{(u,v) \in K \times V \cap P_i \times P_i \mid d_{G[P_i]}(u,v) > d_G(u,v) \text{ and } d_G(u,v) \leq 8\frac{\sqrt{\epsilon}}{c}\Delta\right\}$ be all the pairs in P_i which may have distortion which is too large due to the restriction to P_i. BPR stands for Bad Pairs due to Restriction. Note that if $d_G(u,v) \geq 8\frac{\sqrt{\epsilon}}{c}\Delta$ then u, v will have distortion at most $\frac{c}{\sqrt{\epsilon}}$ (as the diameter of the tree is bounded by 8Δ), which is small enough for our goals (showing that it is a scaling distortion). Now we present some properties of these sets.

Claim 26. *For a graph H, with vertex x_0, radius at most Δ, and a partition (P_1, P_2) of H where P_1 is a petal chosen by the petal decomposition algorithm, T_1 is the tree created from the hierarchical petal decomposition of P_1 and similarly T_2 of P_2 and T of H. It holds that for all $\epsilon > 0$ that: $BP_\epsilon(T) \subseteq BP_\epsilon(T_1) \cup BP_\epsilon(T_2) \cup BPS_\epsilon(P_1, P_2) \cup BPR_\epsilon(P_1) \cup BPR_\epsilon(P_2)$.*

Proof. Let $\{u, v\} \in BP_\epsilon(T)$. We consider the following cases:

- $|\{u,v\} \cap P_1| = 1$: By Fact 4 the diameter of T is at most 8Δ. Therefore $\frac{c}{\sqrt{\epsilon}} \leq \frac{d_T(u,v)}{d_H(u,v)} \leq \frac{8\Delta}{d_H(u,v)}$ which implies $d_H(u,v) \leq 8\frac{\sqrt{\epsilon}}{c}\Delta$ and $(u,v) \in BPS_\epsilon(P_1, P_2)$.
- $|\{u,v\} \cap P_1| = 2$: Since T has diameter at most 8Δ, and (u,v) has distortion greater than $\frac{c}{\sqrt{\epsilon}}$, necessarily $d_H(u,v) \leq \frac{d_T(u,v)}{\frac{c}{\sqrt{\epsilon}}} \leq 8\frac{\sqrt{\epsilon}}{c}\Delta$. Therefore, if $d_{H[P_1]}(u,v) > d_H(u,v)$ then

$\{u, v\} \in BPR_\epsilon(P_1)$. Otherwise $d_{H[P_1]}(u, v) = d_H(u, v)$ and

$$\frac{c}{\sqrt{\epsilon}} \le \frac{d_T(u,v)}{d_H(u,v)} = \frac{d_T(u,v)}{d_{H[P_1]}(u,v)} = \frac{d_{T_1}(u,v)}{d_{H[P_1]}(u,v)}$$

so $(u, v) \in BP_\epsilon(T_1)$.

- $|\{u, v\} \cap P_1| = 0$: This case is similar to the previous one. We get $(u, v) \in BPR_\epsilon(P_2) \cup BP_\epsilon(T_2)$.

□

We now look on some specific stage in the algorithm. We have some cluster Y, root vertex x_0 and target vertex t. We will refer to $W_r(Y, x_0, t)$ as W_r. Define $Q_r(\epsilon) = W_{r+\delta_\epsilon} \backslash W_{r-\delta_\epsilon}$ (with respect to Y, x_0 and t). The next claim shows that for each bad pair that is created by the partition $(W_r, Y \backslash W_r)$, both of its members are in $Q_r(\epsilon)$. This claim will be used for bounding the number of bad pairs created in particular stages of the recursion.

Claim 27. *For a graph $H(V_H, E_H, w_H)$, with center vertex x_0, target vertex t, radius at most Δ, and any partition (P_1, P_2) where $P_1 = W_r$, $P_2 = V_H \backslash W_r$, it holds that $BPS_\epsilon(P_1, P_2) \cup BPR_\epsilon(P_1) \cup BPR_\epsilon(P_2) \subseteq Q_r(\epsilon) \times Q_r(\epsilon)$.*

Proof. The proof is by case analysis. For a pair $\{u, v\}$, in each case it holds that $d_H(u, v) \le 8\frac{\sqrt{\epsilon}}{c}\Delta = \frac{\delta_\epsilon}{4}$.

- $\{u, v\} \in BPS_\epsilon(P_1, P_2)$: W.l.o.g assume that $u \in W_r$. Seeking contradiction, assume $\{u, v\} \notin Q_r(\epsilon) \times Q_r(\epsilon)$. Then $u \in W_{r-\delta_\epsilon}$ or $v \notin W_{r+\delta_\epsilon}$. By Claim 13, if $u \in W_{r-\delta_\epsilon}$ then $v \in W_{r-\delta_\epsilon+4d_H(u,v)} \subseteq W_r$, contradiction. If $v \notin W_{r+\delta_\epsilon}$ then by similar argument $u \notin W_r$, contradiction. Hence $(u, v) \in Q_r(\epsilon) \times Q_r(\epsilon)$.

- $\{u, v\} \in BPR_\epsilon(P_1)$: Then both $u, v \in W_r$. Since $d_{H[P_1]}(u, v) > d_H(u, v)$, it implies that the shortest path from u to v is going through some vertex $z \in P_2$. Seeking contradiction assume $u \in W_{r-\delta_\epsilon}$, then by Claim 13, $d_H(u, z) \ge \frac{\delta_\epsilon}{4}$ (otherwise we will have $z \in W_r$), so $d_H(u, v) = d_H(u, z) + d_H(z, v) > \frac{\delta_\epsilon}{4}$ contradiction. The same argument shows that $v \notin W_{r-\delta_\epsilon}$.

- $\{u, v\} \in BPR_\epsilon(P_2)$: This case is symmetric to the previous one.

□

The next corollary shows that for $\epsilon > 0$, all the bad pairs created by the partition $(W_r, Y \backslash W_r)$, are included in the union of balls with radius $\frac{\delta_\epsilon}{4}$ around the terminals of $Q_r(\epsilon)$. In addition, this union is included in $Q_r(4\epsilon)$. The corollary will be used in the amortized analysis.

Corollary 28. *For a graph $H(V_H, E_H, w_H)$, with root vertex x_0, target vertex t, radius at most Δ, and partition (P_1, P_2) where $P_1 = W_r$, $P_2 = V_H \backslash W_r$, it holds that $BPS_\epsilon(P_1, P_2) \cup BPR_\epsilon(P_1) \cup BPR_\epsilon(P_2) \subseteq \bigcup_{v \in Q_r(\epsilon) \cap K} \{v\} \times B_H\left(v, \frac{\delta_\epsilon}{4}\right) \subseteq (Q_r(\epsilon) \cap K) \times Q_r(4\epsilon)$.*

Proof. By Claim 27, each pair $\{v, u\} \in BPS_\epsilon(P_1, P_2) \cup BPR_\epsilon(P_1) \cup BPR_\epsilon(P_2)$ is included in $Q_r(\epsilon) \times Q_r(\epsilon)$ (note that $v \in K$). By definition, $d_H(v, u) \le \frac{\delta_\epsilon}{4}$, hence the first containment is trivial. For the second containment, it is enough to show that for every $v \in Q_r(\epsilon) \cap K$, $B_H\left(v, \frac{\delta_\epsilon}{4}\right) \subseteq Q_r(4\epsilon)$. Let $u \in B_H\left(v, \frac{\delta_\epsilon}{4}\right)$, by Fact 4, $u \notin W_{r-2\delta_\epsilon}$, as otherwise $v \in W_{r-\delta_\epsilon}$. In addition $v \in W_{r+\delta_\epsilon}$ implies $u \in W_{r+2\delta_\epsilon}$ hence $u \in W_{r+2\delta_\epsilon} \backslash W_{r-2\delta_\epsilon} = W_{r+\delta_{4\epsilon}} \backslash W_{r-\delta_{4\epsilon}} = Q_r(4\epsilon)$. □

The next lemma is the main lemma in this section, it asserts that there exists a "good" choice of radius for all ϵ simultaneously. It is similar in spirit to Lemma 20, but it is more involved because of the difficulty that arises due to bad pairs of types $BPR_\epsilon(P_1), BPR_\epsilon(P_2)$, and because the center point cannot be chosen, thus we have no control over the density of vertices and terminals around the center.

Lemma 29. *Given a graph $H(V_H, E_H, w_H)$, two vertices $x_0, t \in V$, some Δ such that $\Delta_{H,x_0} \leq \Delta$, and an interval $[lo, hi]$ such that $hi - lo = \frac{1}{8}\Delta$ and $lo \in \left[0, \frac{3}{8}\Delta\right]$. There exists $r \in [lo, hi]$ such that for any $\epsilon \in (0,1)$, the partition $P_1 = W_r,\ P_2 = V_H \backslash W_r$ will satisfies at least one of the following:*

1. $|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{2} \cdot (|W_r| \cdot |V \backslash W_r|_k + |W_r|_k \cdot |V \backslash W_r|)$.
2. $|Q_r(4\epsilon)| \leq \frac{\epsilon}{2} \cdot |V|$.

where $Q_r(\epsilon)$ taken with respect to Δ.

We postpone the proof of Lemma 29 to Section 9.2. The next theorem proves that the tree created from the hierarchical petal decomposition has scaling distortion $\frac{c}{\sqrt{\epsilon}}$, given Lemma 29.

Theorem 21. *For a weighted graph $G = (V, E, w)$ and a subset of terminals $K \subseteq V$, the hierarchical petal decomposition algorithm returns a spanning tree T of G with terminal scaling distortion $\frac{c}{\sqrt{\epsilon}}$.*

Proof. Fix $\epsilon > 0$, each terminal will have tokens, which it may use to compensate for bad pairs. We will denote by $BP_\epsilon^M(T)$ (M for miser) all the pairs in $BP_\epsilon(T)$ that no token was paid for, and by $BP_\epsilon^F(T)$ (F for financed) all the pairs in $BP_\epsilon(T)$ which a token was paid for. The algorithm constructs a hierarchical tree, in each step it constructs a petal $P_1 = W_r$, and recursively constructs a tree T_1 for it. For $P_2 = V \backslash W_r$ the algorithm proceeds using the radius Δ of G and not the radius of $X \backslash W_r$ which can only be smaller. The choice of the petal radius in Lemma 29 requires just an upper bound on the radius, hence we can relate to the construction of the tree T_2 on $V \backslash W_r$ as an independent part. Our final tree T is obtain by taking a union of T_1, T_2 with an additional edge. For $i \in \{1, 2\}$ set $n_i = |P_i|$, $k_i = |P_i|_k$.

For each partition (P_1, P_2) of some cluster $Y \subseteq V$ performed during the algorithm, the charging scheme will be as follows: by Lemma 29 there are two possibilities: if $|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{2} \cdot (n_1 \cdot k_2 + k_1 \cdot n_2)$ then no terminal pays tokens. On the other hand if $|Q_r(4\epsilon)| \leq \frac{\epsilon}{2} \cdot |Y|$, then each terminal $v \in Q_r(\epsilon) \cap K$, will pay a token for each vertex in $B_Y\left(v, \frac{\delta_\epsilon}{4}\right)$.

We prove by induction on the cardinality of V, that $\left|BP_\epsilon^M(T)\right| \leq \frac{\epsilon}{2} \cdot |K| \cdot |V|$. In the base case ($|V| \in \{1, 2\}$) there are no bad pairs, hence the claim is trivial. Assume the induction for values of $|V|$ smaller than n and we will prove it for n. Let T_1, T_2 be two trees constructed by the algorithm for P_1, P_2 respectively, such that the spanning tree of V constructed by combining T_1 with T_2. By the induction hypothesis:

$$\left|BP_\epsilon^M(T_1)\right| \leq \frac{\epsilon}{2} \cdot k_1 n_1 \text{ and } \left|BP_\epsilon^M(T_2)\right| \leq \frac{\epsilon}{2} \cdot k_2 n_2.$$

If $|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{2} \cdot (n_1 \cdot k_2 + k_1 \cdot n_2)$, then by Claim 27

$$|BPS_\epsilon(P_1, P_2) \cup BPR_\epsilon(P_1) \cup BPR_\epsilon(P_2)| \leq |Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{2} \cdot (n_1 \cdot k_2 + k_1 \cdot n_2).$$

From Claim 26, and the induction hypothesis we get:

$$\begin{aligned}\left|BP^M_\epsilon(T)\right| &\le \left|BP^M_\epsilon(T_1)\right| + \left|BP^M_\epsilon(T_2)\right| + \left|BPS_\epsilon(P_1,P_2)\cup BPR_\epsilon(P_1)\cup BPR_\epsilon(P_2)\right| \\ &\le \frac{\epsilon}{2}n_1k_1 + \frac{\epsilon}{2}n_2k_2 + \frac{\epsilon}{2}\cdot(n_2k_1+n_1k_2) = \frac{\epsilon}{2}\cdot(k_1+k_2)\cdot(n_1+n_2) = \frac{\epsilon}{2}\cdot n\cdot k\end{aligned}$$

as required. Otherwise, if $|Q_r(4\epsilon)| \le \frac{\epsilon}{2}\cdot|V|$, then by Corollary 28, a token is paid for each bad pair created due to the partition of V to P_1, P_2. Hence no new miser pairs are created and we get $\left|BP^M_\epsilon(T)\right| \le \left|BP^M_\epsilon(T_1)\right| + \left|BP^M_\epsilon(T_2)\right| \le \frac{\epsilon}{2}n_1k_1 + \frac{\epsilon}{2}n_2k_2 \le \frac{\epsilon}{2}\cdot(k_1+k_2)\cdot(n_1+n_2) = \frac{\epsilon}{2}\cdot n\cdot k$, as required.

Next, we count how many tokens were paid throughout the construction. During the partition of some cluster Y, tokens are paid when Lemma 29 guarantees that $|Q_r(4\epsilon)| \le \frac{\epsilon}{2}\cdot|Y|$. Then each terminal $v \in Q_r(\epsilon)\cap K$ pays a token for each vertex in $B_Y\left(v, \frac{\delta_\epsilon}{4}\right) \subseteq Q_r(4\epsilon)$. (The containment is by Corollary 28.) Observe that for any vertex $v \in Y$, after any number of steps of the algorithm, v will be contained in some subset $P \subset Y$ with a smaller radius then the radius of Y (using Fact 1). Hence $B_P\left(v, \frac{8}{c}\sqrt{\epsilon}\Delta_P\right) \subseteq B_Y\left(v, \frac{8}{c}\sqrt{\epsilon}\Delta_Y\right)$. So the first time a terminal decides to use its tokens is also the last time, because any time in the "future", it will try to pay tokens for vertices which it already paid for. We conclude that the sum of all tokens which were paid by all the terminals combined is at most $\frac{\epsilon}{2}\cdot n\cdot k$.

We conclude that

$$|BP_\epsilon(T)| = \left|BP^M_\epsilon(T)\cup BP^R_\epsilon(T)\right| \le \frac{\epsilon}{2}\cdot n\cdot k + \frac{\epsilon}{2}\cdot n\cdot k \le \epsilon\cdot n\cdot k = \epsilon\cdot|K|\cdot|V|.$$

Thus the tree T created by the hierarchical petal decomposition has scaling distortion $\frac{c}{\sqrt{\epsilon}}$, as required. □

Note that by Lemma 19 and Theorem 21, for every graph $G = (V,E,w)$ and a subset of terminals K, there is a spanning tree T of G such that $Avg_f(K\times V) \le O(1)$.

9.2 Proof of Lemma 29.

In this proof, we are given a particular cluster Y root x_0 and target t. For simplicity in notation we will call the vertices of this cluster V, and the graph induced by them G. We will also assume for simplicity that $lo = 0$, while if $lo > 0$ the proof will work in a similar manner. We will consider the following cases, according to the cardinality of vertices and terminals around the target vertex t. Note that $W_{\frac{\Delta}{32}} \cap \left(V\backslash W_{\frac{\Delta}{16}}\right) = \emptyset$ hence $\left|W_{\frac{\Delta}{32}}\right| \le \frac{n}{2}$ or $\left|V\backslash W_{\frac{\Delta}{16}}\right| \le \frac{n}{2}$, and $\left|W_{\frac{\Delta}{32}}\right|_k \le \frac{k}{2}$ or $\left|V\backslash W_{\frac{\Delta}{16}}\right|_k \le \frac{k}{2}$. Lemmas 30,33,36,38 will handle each case:

- If $\left|W_{\frac{\Delta}{32}}\right|_k \le \frac{k}{2}$ and $\left|W_{\frac{\Delta}{32}}\right| \le \frac{n}{2}$ then we will use Lemma 30.
- If $\left|W_{\frac{\Delta}{32}}\right|_k \le \frac{k}{2}$ and $\left|V\backslash W_{\frac{\Delta}{16}}\right| \le \frac{n}{2}$ then we will use Lemma 38.
- If $\left|V\backslash W_{\frac{\Delta}{16}}\right|_k \le \frac{k}{2}$ and $\left|W_{\frac{\Delta}{32}}\right| \le \frac{n}{2}$ then we will use Lemma 36.
- If $\left|V\backslash W_{\frac{\Delta}{16}}\right|_k \le \frac{k}{2}$ and $\left|V\backslash W_{\frac{\Delta}{16}}\right| \le \frac{n}{2}$ then we will use Lemma 33.

The proof of all the four Lemmas will be very similar to the proof of Lemma 20. Recall that $c = 2^{18}, a = 2^6$.

Lemma 30. *Let $G(V,E,w)$ be a graph with two points $x_0, t \in V$, such that the radius from x_0 is at most Δ. If $\left|W_{\frac{\Delta}{32}}\right| \le \frac{n}{2}$ and $\left|W_{\frac{\Delta}{32}}\right|_k \le \frac{k}{2}$ then there is $r \in \left[0, \frac{1}{8}\Delta\right]$, such that the partition $P_1 = W_r,\ P_2 = V \backslash W_r$ satisfies the requirements of Lemma 29.*

Proof. We assume $\left|W_{\frac{\Delta}{32}}\right| \le \frac{n}{2}, \left|W_{\frac{\Delta}{32}}\right|_k \le \frac{k}{2}$ which implies for every $r \in \left[0, \frac{\Delta}{32}\right]$ that $|V\backslash W_r| \ge \frac{n}{2}, |V\backslash W_r|_k \ge \frac{k}{2}$. Set

$$\hat{\epsilon} = \max\left\{\epsilon \middle| \ \left|W_{\frac{\sqrt{\epsilon}}{64}\Delta}\right| \ge \epsilon \cdot n \text{ or } \left|W_{\frac{\sqrt{\epsilon}}{64}\Delta}\right|_k \ge \epsilon \cdot k\right\}.$$

Note that $\hat{\epsilon}$ is well defined: $\frac{1}{n} \le \hat{\epsilon} \le \frac{1}{2}$ (since $t \in W_0$ and $\left|W_{\frac{\Delta}{32}}\right| \le \frac{n}{2}, \left|W_{\frac{\Delta}{32}}\right|_k \le \frac{k}{2}$). Define $\hat{S} = \left[\frac{\sqrt{\hat{\epsilon}}}{64}\Delta, \frac{\sqrt{\hat{\epsilon}}}{32}\Delta\right]$, $S = \left[\frac{\sqrt{\hat{\epsilon}}}{64}\Delta + \delta_{a\hat{\epsilon}}, \frac{\sqrt{\hat{\epsilon}}}{32}\Delta - \delta_{a\hat{\epsilon}}\right]$, $s = \frac{\sqrt{\hat{\epsilon}}}{64}\Delta - 2\delta_{a\hat{\epsilon}}$ and the shell $Q = W_{\frac{\sqrt{\hat{\epsilon}}}{32}\Delta} \backslash W_{\frac{\sqrt{\hat{\epsilon}}}{64}\Delta}$.

Claim 31. $\left|W_{\frac{\sqrt{\hat{\epsilon}}}{32}\Delta}\right| \le 4\hat{\epsilon}n$ *and* $\left|W_{\frac{\sqrt{\hat{\epsilon}}}{32}\Delta}\right|_k \le 4\hat{\epsilon}k$

Proof. If $\hat{\epsilon} \le \frac{1}{4}$, from the maximality of $\hat{\epsilon}$, $\left|W_{\frac{\sqrt{\hat{\epsilon}}}{32}\Delta}\right| = \left|W_{\frac{\sqrt{4\hat{\epsilon}}}{64}\Delta}\right| \le 4\hat{\epsilon}n$, similarly for the terminals. On the other hand if $\hat{\epsilon} > \frac{1}{4}$ then $4\hat{\epsilon} \ge 1$ and the Claim follows trivially. □

We will find r that will be good for all the values of ϵ simultaneously. First we will find r such that any $\epsilon \le a \cdot \hat{\epsilon}$ will satisfy $|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \le \frac{\epsilon}{2} \cdot \left(|W_r| \cdot |V\backslash W_r|_k + |W_r|_k \cdot |V\backslash W_r|\right)$. Then we will show that the same r also satisfies $|Q_r(4\epsilon)| \le \frac{\epsilon}{2} \cdot |V|$ for every $\epsilon > a \cdot \hat{\epsilon}$. For $r \in S$, define $S_r(\epsilon) = (r - \delta_\epsilon, r + \delta_\epsilon) \subseteq \hat{S}$, $s(\epsilon) = 2\delta_\epsilon$. For r and ϵ we say that $A_r(\epsilon)$ holds if the shell $Q_r(\epsilon)$ is small enough which implies that cutting at radius r is "good" for ϵ. Formally

$$A_r(\epsilon) \text{ holds } \iff |Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \le \frac{\epsilon}{4}\hat{\epsilon}nk$$

Notice that since $\frac{\sqrt{\hat{\epsilon}}}{64}\Delta \le r \le \frac{\sqrt{\hat{\epsilon}}}{32}\Delta$, then it must be that $|W_r| \ge \hat{\epsilon}n$ or $|W_r|_k \ge \hat{\epsilon}k$. If $|W_r| \ge \hat{\epsilon}n$ then $A_r(\epsilon)$ implies

$$|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \le \frac{\epsilon}{4} \cdot k \cdot \hat{\epsilon} \cdot n \le \frac{\epsilon}{4} \cdot 2 \cdot |V\backslash W_r|_k \cdot |W_r| \le \frac{\epsilon}{2} \cdot \left(|V\backslash W_r| \cdot |W_r|_k + |V\backslash W_r|_k \cdot |W_r|\right)$$

On the other hand, $|W_r|_k \ge \hat{\epsilon} \cdot k$ implies that

$$|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \le \frac{\epsilon}{4} \cdot \hat{\epsilon} \cdot k \cdot n \le \frac{\epsilon}{4} \cdot |W_r|_k \cdot 2 \cdot |V\backslash W_r| \le \frac{\epsilon}{2} \cdot \left(|V\backslash W_r| \cdot |W_r|_k + |V\backslash W_r|_k \cdot |W_r|\right)$$

That is, the condition of Lemma 30 is satisfied for ϵ. Hence for $\epsilon \in (0, a\hat{\epsilon}]$ the following is sufficient:

Claim 32. *There exists $r \in S$ such that $A_r(\epsilon)$ holds for all $\epsilon \le a \cdot \hat{\epsilon}$.*

Proof. We start by running Algorithm 7. Let $I_t = \texttt{Delete-bad-intervals}(S, A, \delta, a\hat{\epsilon})$. In what follows we will conclude from Claim 23 that $I_t \neq \emptyset$ and hence an appropriate value $r \in S$ can be found. Let $(\epsilon_1, r_1), \ldots, (\epsilon_t, r_t)$ be all the selected variables. Similarly to Claim 24, any point $y \in \hat{S}$ can be covered by at most two intervals $S_{r_j}(\epsilon_j)$, $S_{r_i}(\epsilon_i)$ for some $1 \leq j < i \leq t$. For $x \in Q$, let $q \in S$ be minimal such that $x \in W_q$. As $x \in Q_{r_j}(\epsilon_j)$ iff $q \in S_{r_j}(\epsilon_j)$, this suggests that any $x \in Q$ appears in at most two sets $Q_{r_j}(\epsilon_j)$, $Q_{r_i}(\epsilon_i)$. Note that $Q = \bigcup_{(r,\epsilon)\in S\times(0,a\hat{\epsilon}]} Q_r(\epsilon) \subseteq W_{\frac{\sqrt{\hat{\epsilon}}}{32}\Delta}$, and from Claim 31 $|Q| \leq 4\hat{\epsilon}n$, $|Q|_k \leq 4\hat{\epsilon}k$. Hence by Claim 23, with parameters $\alpha = 32\frac{\Delta}{c}$, $\beta = 4$, it holds that the sum of lenghts of all the intervals deleted by the `Delete-bad-intervals` procedure is bounded by

$$\sum_{i\leq t} s(\epsilon_i) = \sum_{i\leq t} 2\delta_{\epsilon_i} \leq 32\alpha\sqrt{\beta}\sqrt{\hat{\epsilon}} = 32\cdot 32\frac{\Delta}{c}2\sqrt{\hat{\epsilon}} = \frac{\Delta}{2^7}\sqrt{\hat{\epsilon}}$$

But since the length of $I_0 = S$ is

$$s = \frac{\sqrt{\hat{\epsilon}}}{64}\Delta - 2\delta_{a\hat{\epsilon}} = \frac{\sqrt{\hat{\epsilon}}}{64}\Delta - 2\frac{32}{c}\sqrt{a\hat{\epsilon}}\Delta = \left(\frac{1}{2^6} - \frac{2^6\sqrt{a}}{c}\right)\sqrt{\hat{\epsilon}}\Delta = \left(\frac{1}{2^6} - \frac{1}{2^9}\right)\sqrt{\hat{\epsilon}}\Delta > \frac{\Delta}{2^7}\sqrt{\hat{\epsilon}}$$

it is impossible that the entire interval I_0 was removed, therefore $I_t \neq \emptyset$, and in fact any $r \in I_t$ satisfies the condition of the claim. □

Next, we show that any choice of $r \in S$ will produce a partition that for every $\epsilon > a\cdot\hat{\epsilon}$ will satisfy $|Q_r(4\epsilon)| \leq \frac{\epsilon}{2}\cdot|V|$. We get:

$$\begin{aligned}
|Q_r(4\epsilon)| &\leq \left|W_{(r+\delta_{4\epsilon})}\right| \leq \left|W_{\left(\frac{\sqrt{\hat{\epsilon}}}{32}\Delta + \frac{32\sqrt{4\epsilon}}{c}\Delta\right)}\right| \leq \left|W_{\left(\frac{\sqrt{\epsilon}}{\sqrt{a}32}\Delta + \frac{2^6\sqrt{\epsilon}}{c}\Delta\right)}\right| \\
&\leq \left|W_{\left(\left(\frac{2}{\sqrt{a}} + \frac{2^{12}}{c}\right)\frac{\sqrt{\epsilon}}{64}\Delta\right)}\right| \overset{(1)}{\leq} \left(\frac{2}{\sqrt{a}} + \frac{2^{12}}{c}\right)^2\cdot\epsilon\cdot n = \left(\frac{1}{2^2} + \frac{1}{2^6}\right)^2\cdot\epsilon\cdot n \leq \frac{\epsilon}{2}\cdot n
\end{aligned}$$

where (1) holds by the fact $\left(\frac{2}{\sqrt{a}} + \frac{2^{11}}{c}\right)^2\epsilon \geq \frac{4}{a}\cdot a\cdot\hat{\epsilon} > \hat{\epsilon}$ and the maximality of $\hat{\epsilon}$. We found $r \in S$ such that for every $\epsilon \in (0,1)$ at least one property of Lemma 29 is satisfied. □

Lemma 33. *Let $G(V,E,w)$ be a graph with two points $x_0, t \in V$, such that the radius from x_0 is at most Δ. If $\left|V\backslash W_{\frac{\Delta}{16}}\right| \leq \frac{n}{2}$ and $\left|V\backslash W_{\frac{\Delta}{16}}\right|_k \leq \frac{k}{2}$ then there is $r \in \left[0, \frac{1}{8}\Delta\right]$, such that the partition $P_1 = W_r$, $P_2 = V\backslash W_r$ satisfies the requirements of Lemma 29.*

Proof. The proof of this Lemma is fully symmetric to Lemma 30. Instead of growing the petal W_r we think about it as we grow $V\backslash W_r$. We give most of the details for completeness.

We assume $\left|V\backslash W_{\frac{\Delta}{16}}\right| \leq \frac{n}{2}$ and $\left|V\backslash W_{\frac{\Delta}{16}}\right|_k \leq \frac{k}{2}$, hence for every $r \in \left[\frac{\Delta}{16}, \frac{\Delta}{8}\right]$, $|W_r| \geq \frac{n}{2}$ and $|W_r| \geq \frac{k}{2}$. Set $\gamma = \frac{1}{8}\Delta$ and

$$\hat{\epsilon} = \max\left\{\epsilon \middle| \left|V\backslash W_{\gamma - \frac{\sqrt{\epsilon}}{32}\Delta}\right| \geq \epsilon\cdot n \text{ or } \left|V\backslash W_{\gamma - \frac{\sqrt{\epsilon}}{32}\Delta}\right|_k \geq \epsilon\cdot k\right\}.$$

Note that $\hat{\epsilon}$ is well defined: $\frac{1}{n} \leq \hat{\epsilon} \leq \frac{1}{2}$ (since $x_0 \in V\backslash W_\gamma$ and $\left|V\backslash W_{\frac{\Delta}{16}}\right| \leq \frac{n}{2}$, $\left|V\backslash W_{\frac{\Delta}{16}}\right|_k \leq \frac{k}{2}$). Let $\hat{S} = \left[\gamma - \frac{\sqrt{\hat{\epsilon}}}{16}\Delta, \gamma - \frac{\sqrt{\hat{\epsilon}}}{32}\Delta\right]$, $S = \left[\gamma - \frac{\sqrt{\hat{\epsilon}}}{16}\Delta + \delta_{a\hat{\epsilon}}, \gamma - \frac{\sqrt{\hat{\epsilon}}}{32}\Delta - \delta_{a\hat{\epsilon}}\right]$, $s = \frac{\sqrt{\hat{\epsilon}}}{32}\Delta - 2\delta_{a\hat{\epsilon}}$ and the shell $Q = W_{\gamma - \frac{\sqrt{\hat{\epsilon}}}{32}\Delta}\backslash W_{\gamma - \frac{\sqrt{\hat{\epsilon}}}{16}\Delta}$. The following claim is similar to Claim 31 and we omit its proof.

Claim 34. $\left|V\backslash W_{\gamma-\frac{\sqrt{\hat{\epsilon}}}{16}\Delta}\right|\leq 4\hat{\epsilon}n$ *and* $\left|V\backslash W_{\gamma-\frac{\sqrt{\hat{\epsilon}}}{16}\Delta}\right|_{k}\leq 4\hat{\epsilon}k$

First we will find $r\in S$ that will be good for any $\epsilon\leq a\cdot\hat{\epsilon}$. Recall the definitions: for $r\in S$, $S_r(\epsilon)=(r-\delta_\epsilon,r+\delta_\epsilon)\subseteq\hat{S}$, $s(\epsilon)=2\delta_\epsilon$. For r and ϵ we say that $A_r(\epsilon)$ holds if $|Q_r(\epsilon)|\cdot|Q_r(\epsilon)|_k\leq\frac{\epsilon}{4}\cdot k\cdot\hat{\epsilon}\cdot n$. Since $r<\gamma-\frac{\sqrt{\hat{\epsilon}}}{32}\Delta$, it holds that $|V\backslash W_r|\geq\hat{\epsilon}n$ or $|V\backslash W_r|_k\geq\hat{\epsilon}k$. If $|V\backslash W_r|\geq\hat{\epsilon}n$ then $A_r(\epsilon)$ implies

$$|Q_r(\epsilon)|\cdot|Q_r(\epsilon)|_k\leq\frac{\epsilon}{4}\cdot k\cdot\hat{\epsilon}\cdot n\leq\frac{\epsilon}{4}\cdot 2\cdot|W_r|_k\cdot|V\backslash W_r|\leq\frac{\epsilon}{2}\cdot\left(|V\backslash W_r|\cdot|W_r|_k+|V\backslash W_r|_k\cdot|W_r|\right)$$

Symmetrically, $|V\backslash W_r|_k\geq\hat{\epsilon}k$ implies that $|Q_r(\epsilon)|\cdot|Q_r(\epsilon)|_k\leq\frac{\epsilon}{2}\cdot(|V\backslash W_r|\cdot|W_r|_k+|V\backslash W_r|_k\cdot|W_r|)$. That is, the condition of Lemma 30 is satisfied for ϵ. Hence for $\epsilon\in(0,a\hat{\epsilon}]$ the following is sufficient:

Claim 35. *There exists $r\in S$ such that $A_r(\epsilon)$ holds for all $\epsilon\leq a\cdot\hat{\epsilon}$.*

Proof. This proof is fully symmetric to the proof of Claim 32. Let I_t = `Delete-bad-intervals`$(S,A,\delta,a\hat{\epsilon})$ and $(\epsilon_1,r_1),\ldots,(\epsilon_t,r_t)$ be all the variables selected by the algorithm. As in Claim 32, any point $x\in Q$ appears in at most two sets $Q_{r_j}(\epsilon_j)$, $Q_{r_i}(\epsilon_i)$. By Claim 34 $|Q|\leq 4\hat{\epsilon}n$, $|Q|_k\leq 4\hat{\epsilon}k$. Hence by Claim 23, with parameters $\alpha=32\frac{\Delta}{c}$, $\beta=4$ it holds that that the sum of lengths of all the intervals deleted by the `Delete-bad-intervals` procedure is bound by $\sum_{i\leq t}s(\epsilon_i)\leq\frac{\Delta}{2^7}\sqrt{\hat{\epsilon}}$, but since the length of $I_0=S$ is $s=\frac{\sqrt{\hat{\epsilon}}}{32}\Delta-2\delta_{a\hat{\epsilon}}>\frac{\Delta}{2^7}\sqrt{\hat{\epsilon}}$ it is impossible that the entire interval I_0 was removed, therefore $I_t\neq\emptyset$, and any $r\in I_t$ satisfies the condition of the Claim. □

Next we show that any choice of $r\in S$ will produce a partition that satisfies for every $\epsilon>a\cdot\hat{\epsilon}$, $|Q_r(4\epsilon)|\leq\frac{\epsilon}{4}\cdot|V|$. We get:

$$\begin{aligned}|Q_r(4\epsilon)| &\leq \left|V\backslash W_{(r-\delta_{4\epsilon})}\right|\leq\left|V\backslash W_{\left(\gamma-\frac{\sqrt{\hat{\epsilon}}}{16}\Delta-\frac{32\sqrt{4\epsilon}}{c}\Delta\right)}\right|\leq\left|V\backslash W_{\left(\gamma-\frac{\sqrt{\epsilon}}{\sqrt{a}16}\Delta-\frac{2^6\sqrt{\epsilon}}{c}\Delta\right)}\right|\\ &\leq \left|V\backslash W_{\left(\gamma-\left(\frac{2}{\sqrt{a}}+\frac{2^{11}}{c}\right)\frac{\sqrt{\epsilon}}{32}\Delta\right)}\right|\overset{(1)}{\leq}\left(\frac{2}{\sqrt{a}}+\frac{2^{11}}{c}\right)^2\cdot\epsilon\cdot n=\left(\frac{1}{2^2}+\frac{1}{2^7}\right)^2\cdot\epsilon\cdot n\leq\frac{\epsilon}{2}\cdot n\end{aligned}$$

where (1) holds by the fact $\left(\frac{2}{\sqrt{a}}+\frac{2^{11}}{c}\right)^2\epsilon\geq\frac{4}{a}\cdot a\cdot\hat{\epsilon}>\hat{\epsilon}$ and the maximality of $\hat{\epsilon}$. We found $r\in S$ such that for every $\epsilon\in(0,1)$ at least one property of Lemma 29 is satisfied. □

Lemma 36. *Let $G(V,E,w)$ be a graph with two points $x_0,t\in V$, such that the radius from x_0 is at most Δ. If $\left|W_{\frac{\Delta}{32}}\right|\leq\frac{n}{2}$ and $\left|V\backslash W_{\frac{\Delta}{16}}\right|_k\leq\frac{k}{2}$ then there is $r\in\left[0,\frac{1}{8}\Delta\right]$, such that the partition $P_1=W_r,\ P_2=V\backslash W_r$ satisfies the requirements of Lemma 29.*

Proof. This case is actually a bit easier than the former ones. If $\left|W_{\frac{\Delta}{32}}\right|_k\leq\frac{k}{2}$ then by Lemma 30 we are done, hence we can assume $\left|W_{\frac{\Delta}{32}}\right|_k>\frac{k}{2}$. Similarly, if $\left|V\backslash W_{\frac{\Delta}{16}}\right|\leq\frac{n}{2}$ then we done by Lemma 33, hence we assume $\left|V\backslash W_{\frac{\Delta}{16}}\right|>\frac{n}{2}$. We will use the same basic technique but without defining $\hat{\epsilon}$. One can think that we choose $\hat{\epsilon}=1$. Define $\hat{S}=\left[\frac{\Delta}{32},\frac{\Delta}{16}\right]$ and $S=\left[\frac{\Delta}{32}+\delta_1,\frac{\Delta}{16}-\delta_1\right]\subseteq\hat{S}$ where $\delta_1=\frac{32}{c}\Delta$, $s=\frac{1}{32}\Delta-2\delta_1$, and the shell $Q=W_{\frac{\Delta}{16}-\delta_1}\backslash W_{\frac{\Delta}{32}+\delta_1}$.

For $r \in S$ and $\epsilon > 0$ define $S_r(\epsilon) = (r - \delta_\epsilon, r + \delta_\epsilon) \subseteq S$, $s(\epsilon) = 2\delta_\epsilon$. We say that for $r \in S$ and $\epsilon \in (0,1)$, $A_r(\epsilon)$ holds if $|Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k \leq \frac{\epsilon}{8} \cdot n \cdot k$. $A_r(\epsilon)$ implies that the partition $(W_r, V \backslash W_r)$ is good for ϵ, because than

$$\begin{aligned} |Q_r(\epsilon)| \cdot |Q_r(\epsilon)|_k &\leq \frac{\epsilon}{8} \cdot n \cdot k < \frac{\epsilon}{8} \cdot 2 \cdot \left|V \backslash W_{\frac{\Delta}{16}}\right| \cdot 2 \cdot \left|W_{\frac{\Delta}{32}}\right|_k \\ &\leq \frac{\epsilon}{2} \cdot |V \backslash W_r| \cdot |W_r|_k \leq \frac{\epsilon}{2} \cdot \left(|W_r| \cdot |V \backslash W_r|_k + |W_r|_k \cdot |V \backslash W_r|\right). \end{aligned}$$

That is, the first condition of Lemma 29 holds for ϵ. Hence the following is sufficient:

Claim 37. *There exists $r \in S$ such that $A_r(\epsilon)$ holds for all $\epsilon \in (0,1)$.*

Proof. This proof is similar to the proof of Claim 32. Let $I_t = \texttt{Delete-bad-intervals}(S, A, \delta, a)$ and $(\epsilon_1, r_1), \ldots, (\epsilon_t, r_t)$ be all the variables selected by the algorithm. As in Claim 32, any point $x \in Q$ appears in at most two sets $Q_{r_j}(\epsilon_j)$, $Q_{r_i}(\epsilon_i)$. Trivially $|Q| \leq 4n$ and $|Q|_k \leq 4k$. Hence by Claim 23, with parameters $\alpha = 32\frac{\Delta}{c}$, $\beta = 8$, it holds that the measure of all the intervals deleted by the `Delete-bad-intervals` procedure is bounded by

$$\sum_{i \leq t} s(\epsilon_i) = \sum_{i \leq t} 2\delta_{\epsilon_i} \leq 32\alpha\sqrt{\beta} = 32 \cdot 32\frac{\Delta}{c}\sqrt{8} = \frac{\Delta}{2^{6.5}}$$

But since the length of $I_0 = S$ is

$$s = \frac{1}{32}\Delta - 2\delta_1 = \frac{1}{32}\Delta - 2\frac{32}{c}\Delta = \left(\frac{1}{2^5} - \frac{2^6}{c}\right)\Delta = \left(\frac{1}{2^5} - \frac{1}{2^{12}}\right)\Delta > \frac{\Delta}{2^{6.5}}$$

it is impossible that the entire interval I_0 was removed, therefore $I_t \neq \emptyset$, and in fact any $r \in I_t$ satisfies the condition of the Claim. □

We conclude that the first requirement of Lemma 29 is satisfied for all $0 < \epsilon < 1$. □

Lemma 38. *Let $G(V, E, w)$ be a graph with two points $x_0, t \in V$, such that the radius from x_0 is at most Δ. If $\left|V \backslash W_{\frac{\Delta}{16}}\right| \leq \frac{n}{2}$ and $\left|W_{\frac{\Delta}{32}}\right|_k \leq \frac{k}{2}$ then there is $r \in \left[0, \frac{1}{8}\Delta\right]$, such that the partition $P_1 = W_r$, $P_2 = V \backslash W_r$ satisfies the requirements of Lemma 29.*

The proof of Lemma 38 is perfectly symmetric to the proof of Lemma 36 and we will omit it.

Part V

Conclusion, Future Work and Open Problems

We considered two notions of terminal embeddings, one that has a guarantee for all pairs containing a terminal, and the strong setting, which has a (different) guarantee also for all other pairs. We showed results that match the state-of-the-art in various settings, for both notions, while replacing the dependence of the distortion on the size of the metric, with the cardinality of the set of terminals. Recently we have been working on a strengthening of the notion of the terminal distortion, which we call for now *priority distortion*. Given a metric space (X, d), a monotone function γ and a priority list $\pi = \{v_1, v_2, \ldots, v_n\}$ over the points of X (some permutation), we say that an embedding f has priority distortion γ with respect to π if the distortion of the pair $\{v_i, v_j\}$ is bounded by $\gamma(\min\{i, j\})$. Some preliminary results are: embedding any metric space into a distribution over ultrametrics with $\gamma(i) = O(\log i)$, and into ℓ_p with $\gamma(i) = \tilde{O}(\log i)$.

We leave a number of interesting open problems for future research.

- Can the dimension reduction result of [JL84] be attained in a terminal setting, without losing a constant factor in the distortion? That is, if (X, d) is an ℓ_2 metric, the terminals $K \subseteq X$ of size $|K| = k$, and $\epsilon > 0$, we would like an embedding into $\ell_2^{g(\epsilon)\cdot\log k}$ with terminal distortion $1 + \epsilon$, for some function $g : \mathbb{R} \to \mathbb{R}$ (ideally $g(\epsilon) = O(1/\epsilon^2)$). Recall that by Corollary 1 we obtain terminal distortion $O(1)$.

- For a weighted graph, and some constants α, β is it NP hard to determine if there exists an $(\alpha, \beta) - k - LTT$? In this thesis we prove it only for $\beta = 1$.

References

[ABN07] Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding metrics into ultrametrics and graphs into spanning trees with constant average distortion. In *SODA*, pages 502–511, 2007.

[ABN08] Ittai Abraham, Yair Bartal, and Ofer Neiman. Nearly tight low stretch spanning trees. In *FOCS '08: Proceedings of the 2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 781–790, Washington, DC, USA, 2008. IEEE Computer Society.

[ABN11] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026 – 3126, 2011.

[AF09] Reid Andersen and Uriel Feige. Interchanging distance and capacity in probabilistic mappings. Technical report, 2009.

[AKPW95] 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.

[ALN08] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. *Journal of the American Mathematical Society 21*, 1:1–21, 2008.

[AN12] Ittai Abraham and Ofer Neiman. Using petal-decompositions to build a low stretch spanning tree. In *STOC*, pages 395–406, 2012.

[BAP92] A. Baratz B. Awerbuch and D. Peleg. Efficient broadcast and light-weight spanners. Technical Report CS92-22, The Weizmann Institute of Science, Rehovot, Israel., 1992.

[Bar96] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *FOCS*, pages 184–193, 1996.

[Bar04] Yair Bartal. Graph decomposition lemmas and their role in metric embedding methods. In *ESA*, pages 89–97, 2004.

[BLMN05] Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. Some low distortion metric ramsey problems. *Discrete & Computational Geometry*, 33(1):27–41, 2005.

[Bou85] J. Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics*, 52(1-2):46–52, 1985.

[CKR04] Gruia Călinescu, Howard J. Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. *SIAM Journal on Computing*, 34(2):358–372, 2004.

[CLRS09] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, Third Edition.* The MIT Press, 3rd edition, 2009.

[CXKR06] T.-H. Chan, Donglin Xia, Goran Konjevod, and Andrea Richa. A tight lower bound for the steiner point removal problem on trees. In *Proceedings of the 9th International Conference on Approximation Algorithms for Combinatorial Optimization Problems, and 10th International Conference on Randomization and Computation*, APPROX'06/RANDOM'06, pages 70–81, Berlin, Heidelberg, 2006. Springer-Verlag.

[EEST08] Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. Lower-stretch spanning trees. *SIAM Journal on Computing*, 38(2):608–628, 2008.

[EFN] Michael Elkin, Arnold Filtser, and Ofer Neiman. Terminal embeddings. *To appear.*

[EGK+10] Matthias Englert, Anupam Gupta, Robert Krauthgamer, Harald Räcke, Inbal Talgam-Cohen, and Kunal Talwar. Vertex sparsifiers: New results from old techniques. In *Proceedings of the 13th International Conference on Approximation, and 14 the International Conference on Randomization, and Combinatorial Optimization: Algorithms and Techniques*, APPROX/RANDOM'10, pages 152–165, Berlin, Heidelberg, 2010. Springer-Verlag.

[FRT04] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. Syst. Sci.*, 69(3):485–497, 2004.

[GNR10] Anupam Gupta, Viswanath Nagarajan, and R. Ravi. An improved approximation algorithm for requirement cut. *Oper. Res. Lett.*, 38(4):322–325, 2010.

[GP68] E. N. Gilbert and H. O. Pollak. Steiner minimal trees. *SIAM Journal on Applied Mathematics*, 16(1):1–29, Jan 1968.

[Gup01] Anupam Gupta. Steiner points in tree metrics don't (really) help. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '01, pages 220–227, Philadelphia, PA, USA, 2001. Society for Industrial and Applied Mathematics.

[JL84] William Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary Mathematics*, 26:189206, 1984.

[KKN14] Lior Kamma, Robert Krauthgamer, and Huy L. Nguyen. Cutting corners cheaply, or how to remove steiner points. In *SODA*, pages 1029–1040, 2014.

[KLMN05] Robert Krauthgamer, James R. Lee, Manor Mendel, and Assaf Naor. Measured descent: a new embedding method for finite metrics. *Geometric and Functional Analysis*, 15(4):839–858, 2005.

[KRY93] Samir Khuller, Balaji Raghavachari, and Neal E. Young. Balancing minimum spanning and shortest path trees. In *SODA*, pages 243–250, 1993.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[MN06] Manor Mendel and Assaf Naor. Ramsey partitions and proximity data structures. In *FOCS*, pages 109–118, 2006.

[Moi09] Ankur Moitra. Approximation algorithms for multicommodity-type problems with guarantees independent of the graph size. In *FOCS*, pages 3–12, 2009.

[Räc08] Harald Räcke. Optimal hierarchical decompositions for congestion minimization in networks. In *STOC*, pages 255–264, 2008.

[RR98] Yuri Rabinovich and Ran Raz. Lower bounds on the distortion of embedding finite metric spaces in graphs. *Discrete & Computational Geometry*, 19(1):79–94, 1998.