Covering Metric Spaces by Few Trees

Yair Bartal
Department of Computer Science, Hebrew University of Jerusalem, Israel
yair@cs.huji.ac.il

Nova Fandina
Department of Computer Science, Hebrew University of Jerusalem, Israel
fandina@cs.huji.ac.il

Ofer Neiman
Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva, Israel
neimano@cs.bgu.ac.il

Abstract

A *tree cover* of a metric space (X, d) is a collection of trees, so that every pair $x, y \in X$ has a low distortion path in one of the trees. If it has the stronger property that every point $x \in X$ has a single tree with low distortion paths to all other points, we call this a *Ramsey* tree cover. Tree covers and Ramsey tree covers have been studied by [15, 31, 19, 30, 38], and have found several important algorithmic applications, e.g. routing and distance oracles. The union of trees in a tree cover also serves as a special type of spanner, that can be decomposed into a few trees with low distortion paths contained in a single tree; Such spanners for Euclidean pointsets were presented by [8].

In this paper we devise efficient algorithms to construct tree covers and Ramsey tree covers for general, planar and doubling metrics. We pay particular attention to the desirable case of distortion close to 1, and study what can be achieved when the number of trees is small. In particular, our work shows a large separation between what can be achieved by tree covers vs. Ramsey tree covers.

2012 ACM Subject Classification Mathematics of computing → Graph algorithms; Theory of computation → Sparsification and spanners

Keywords and phrases tree cover, Ramsey tree cover, probabilistic hierarchical family

Digital Object Identifier 10.4230/LIPIcs.ICALP.2019.20

Category Track A: Algorithms, Complexity and Games

Related Version A full version of the paper is available at `https://arxiv.org/abs/1905.07559`.

Funding *Yair Bartal*: Supported in part by a grant from the Israeli Science Foundation (1817/17).
Nova Fandina: Supported in part by a grant from the Israeli Science Foundation (1817/17).
Ofer Neiman: Supported in part by a grant from the Israeli Science Foundation (1817/17) and in part by BSF grant 2015813.

Acknowledgements We are grateful to Michael Elkin and Shay Solomon for fruitful discussions.

1 Introduction

The problem of approximating metric spaces by tree metrics has been a successful research thread in the past decades, and has found numerous algorithmic applications. This is mainly due to the fact that a tree has a very simple structure that can be exploited by the algorithm designer. While a single tree cannot provide a meaningful approximation, due to a lower bound of [44] (the metric of the n point cycle requires $\Omega(n)$ distortion for embedding into a tree), several other variants have been considered in the literature. The purpose of this paper is to study the natural question whether there exists a small collection of trees (*tree cover*) such that each pair is well preserved in at least one of them. A natural stronger demand may be that for each point all of its interpoint distances to the rest of the metric are well preserved in one of the trees (*Ramsey tree covers*).

Tree covers and Ramsey tree covers have been studied by [31, 15, 19, 30, 38], and are useful ingredients in important algorithmic applications such as routing and distance oracles.

Given a metric space (X, d_X) and an edge-weighted tree T with $X \subseteq V(T)$, for $x, y \in X$ let $d_T(x, y)$ denote the length of the path in T from x to y. We say T is *dominating* if $d_T(x, y) \geq d_X(x, y)$ for all $x, y \in X$. A dominating tree T has *distortion* α for a pair $x, y \in X$, if $d_T(x, y) \leq \alpha \cdot d_X(x, y)$. In what follows, all trees we consider are always dominating (this can be assumed w.l.o.g.).

▶ **Definition 1** (Tree cover). *Given a metric space (X, d_X), for $\alpha \geq 1$ and an integer k, a* tree cover *with distortion α and size k, (α, k)-tree cover in short, is a collection of k dominating trees $T_1, \ldots, T_k$, such that for any $u \neq v$ in X there is a tree T_i with distortion at most α for the pair u, v.*

If for each $u \in X$ there is a tree T_i with distortion at most α for each pair u, v with $v \in X$, we call this a Ramsey (α, k)-tree cover*.*

If the metric is a shortest path metric of some graph G, and the trees are subgraphs, we call this a *spanning* tree cover.

The notion of tree covers is closely related to the well studied notion of spanners. In the context of metric spaces, a *spanner* with distortion α for the metric (X, d_X), is a graph H with $X \subseteq V(H)$, so that for all $x, y \in X$, $d_X(x, y) \leq d_H(x, y) \leq \alpha \cdot d_X(x, y)$. It is often desired that the spanner would be a sparse graph. Note that taking H as the union of the trees in a (Ramsey) tree cover forms a sparse spanner with a special structure; that can be decomposed into a few trees, and every pair (or every point) has the distortion guarantee in one of these trees. In the context of graphs H is usually required to be a subgraph of the original graph, and then the same holds for spanning tree covers. Spanners are basic graph constructions, have been intensively studied [42, 7, 22, 20, 8, 24, 16, 47, 41] and have numerous applications in various settings, see e.g. [9, 43, 10, 22, 46, 25].

A related well-studied concept is probabilistic embedding of a metric space into tree metrics. This notion was introduced by Bartal [11], and a sequence of works by Bartal, and Fakcharoenphol et al. [12, 27, 13] culminated in obtaining a tight $O(\log n)$ bound. The result of [11] already implies a probabilistic construction of tree covers of size k with distortion $O(n^{2/k} \log n)$, for general metrics spaces. In Theorem 2 we improve this by constructing deterministic Ramsey tree covers with almost optimal distortion (nearly matching the lower bound in Theorem 27).

In the rest of the section we review known results on tree covers and Ramsey tree covers, and present the new results of this paper.

1.1 Tree Covers

In the context of Euclidean spanners, Arya et al. [8] used the so called dumbbell trees to build low distortion spanners. Rephrased in our context, they obtained a tree cover for Euclidean pointsets. More specifically, for any finite set of points in d-dimensional Euclidean space, and any parameter $\epsilon > 0$, they devised a $(1 + \epsilon)$-distortion tree cover with $O((d/\epsilon)^d \log(d/\epsilon))$ trees. We note that their trees are using Steiner points (i.e., points in Euclidean space that are not part of the input set), and it is not clear that such points can be removed from the spanner while maintaining $(1 + \epsilon)$ distortion.

Chan et al. [19] presented tree covers for doubling metrics. The *doubling constant* of a metric (X, d_X) is the minimal λ, so that every ball of radius $2r$ can be covered by λ balls of radius r. The *doubling dimension* of (X, d_X) is defined as $\log \lambda$, and a family of metrics is called *doubling* if every metric in it has doubling dimension $O(1)$. The result of [19] used hierarchical partitioning to construct a tree cover with distortion $O(\log^2 \lambda)$ and $O(\log \lambda \cdot \log\log \lambda)$ trees.

The notion of spanning tree covers was introduced by Gupta et al. in [31], who used these for MPLS routing. They devised spanning tree covers for planar graphs: an exact (i.e. distortion 1) tree cover with $O(\sqrt{n})$ trees (more generally $O(r(n)\log n)$ trees for graphs admitting a hierarchical $r(n)$ size separators), and a spanning tree cover with distortion 3 and only $O(\log n)$ trees. They also showed the former result for planar graphs is tight, i.e., at least $\Omega(\sqrt{n})$ trees are needed for an exact tree cover.

1.1.1 Our results

As a starting point for this study, we observe, that for general metrics, the number of trees of any tree covers with distortion α must be as large as $n^{1/\alpha}$. (This bound stems from the standard example of high girth graphs and extends a previous lower bound for spanning trees of [31]). Nearly optimal upper bounds are known even for Ramsey tree covers (see next subsection). The above lower bound also implies a lower bound of $\lambda^{1/\alpha}$ in any space with doubling constant λ.

One of our main results is a tree cover for doubling metrics. We develop a novel hierarchical clustering for such metrics, built in a bottom-up manner. We then use this new clustering to show that for any $0 < \epsilon < 1$, every metric with doubling constant λ admits a tree cover with distortion $1+\epsilon$ and only $(1/\epsilon)^{O(\log \lambda)}$ trees. Since d-dimensional Euclidean space has doubling dimension $\Theta(d)$, the number of trees in the cover is therefore $(1/\epsilon)^{O(d)}$. Hence, this can be viewed as both a generalization and improvement of the result of [8]. Moreover, we improve their result in another aspect, since unlike [8] we do not require the use of Steiner points. In particular, for any $\epsilon > 0$ our result provides a $(1+\epsilon)$-spanner with $n/\epsilon^{O(\log \lambda)}$ edges, that can be decomposed to a small number of trees, and where each pair has a $1+\epsilon$ stretch path in one of the trees. We note that the number of edges in this spanner is asymptotically optimal [45], and thus so is our result.

We then turn to obtaining a distortion-size tradeoff for tree covers of doubling spaces with arbitrary distortion α. We improve and extend the result of [19]; for any parameter α, we use a more sophisticated construction of hierarchical partitions, to build a tree cover with distortion $O(\alpha)$ and $O(\lambda^{1/\alpha} \cdot \log \lambda \cdot \log \alpha)$ trees (note that setting $\alpha = \log \lambda$ yields distortion $O(\log \lambda)$ with $O(\log \lambda \cdot \log\log \lambda)$ trees). We note that the trees obtained here are in fact *ultrametrics*[1]. This result provides a special type of spanner with distortion $O(\alpha)$ and $O(n \cdot \lambda^{1/\alpha} \cdot \log \lambda \cdot \log \alpha)$ edges, which improves the recent spanner construction of [28], whose number of edges was larger by a factor of $O(\log_\lambda n)$ (though their spanner has additionally bounded lightness). The lower bound mentioned above for doubling spaces implies that our tree cover bounds cannot be substantially improved.

For planar graphs with n vertices, and more generally graphs excluding a fixed minor, we apply the path-separators framework of [48, 4], and show that for any $\epsilon > 0$, there exists a tree cover with distortion $1+\epsilon$ and $O((\log n)/\epsilon)^2$ trees. (Recall that for distortion 1, [31] used the planar separators of [37], but this requires $\Omega(\sqrt{n})$ trees.) We also observe that certain hierarchical partitions of [35] for planar (and fixed minor-free) graphs, can be used to obtain a tree cover with $O(1)$ distortion and only $O(1)$ trees (the obtained trees are ultrametrics).

[1] An ultrametric (U, d_U) is a metric satisfying a strong form of the triangle inequality, $\forall x, y, z$, $d_U(x,z) \le \max\{d_U(x,y), d_U(y,z)\}$. An ultrametric is both a tree metric and a Euclidean metric.

See Table 1 for a succinct comparison between our and previous results on tree covers.

Table 1 Results on tree covers for general, planar (our new results also hold for fixed minor-free graphs) and doubling metrics. (The upper bound for general metrics appears in Table 2.)

Family	Reference	Number of trees	Distortion
General metrics	**New**	$\Omega(n^{1/\alpha})$	α
Doubling metrics	[19]	$O(\log\lambda \cdot \log\log\lambda)$	$O(\log^2\lambda)$
	New	$(1/\epsilon)^{\Theta(\log\lambda)}$	$1+\epsilon$
	New	$O(\lambda^{1/\alpha}\cdot\log\lambda\cdot\log\alpha)$	$O(\alpha)$
	New	$\Omega(\lambda^{1/\alpha})$	α
Planar metrics	[31]	$\Theta(\sqrt{n})$	1
	[31]	$O(\log n)$	3
	New	$O((\log n)/\epsilon)^2$	$1+\epsilon$
	[35]+**New**	$O(1)$	O(1)

1.2 Ramsey Tree Covers

Given a metric (X, d), the metric Ramsey problem asks for a large subset $S \subseteq X$ that embeds with a given distortion into a simple metric, such as a tree metric or Euclidean space. Following [15], [38] gave a probabilistic construction that finds in any n point metric (X, d) a set $S \subseteq X$ of size at least $n^{1-1/\alpha}$ that embeds into an ultrametric with distortion $O(\alpha)$. In fact, the embedding has such distortion on all pairs in $S \times X$. Applying this iteratively, [38] obtained a collection of $O(\alpha \cdot n^{1/\alpha})$ trees, so that each point $x \in X$ has a "home tree" T_x with distortion $O(\alpha)$ for every pair containing x. We call such a collection a *Ramsey tree cover*. Some further works aim at improving the leading constant in the distortion ([14, 40, 17]) and finding a deterministic construction ([14]). Recently, in the graph setting, [3] devised a *spanning* Ramsey tree cover, where the trees are subgraphs of the input graph. They obtained the same number of trees, but with slightly larger distortion $O(\alpha \cdot \log\log n)$.

We note that the number of trees in all previous works is $\alpha \cdot n^{1/\alpha} \geq \log n$ for any value of α. It seems like a natural question to understand what can be achieved in the inverse tradeoff, where the number of trees, k, is small. We remark that the lower bound via high girth graphs is rather weak, it implies that using k trees the distortion must be only $\Omega(\log_k n)$, as that is the bound on the girth of a graph with kn edges [18].

1.2.1 Our results

We focus on the regime where the number of trees is small. We first observe that a similar method as used in [38] of iteratively extracting large Ramsey subspaces can be applied in this setting as well. Given any metric space (X, d) on n points, and a parameter $k \geq 1$, there exists a Ramsey tree cover of size k (in particular, ultrametrics) and distortion $O(n^{1/k} \cdot \log^{1-1/k} n)$. We also note that the result of [3] can be translated to this setting: given a graph $G = (V, E)$ with n vertices, we find a Ramsey *spanning* tree cover with k spanning trees and distortion $O(n^{1/k} \cdot \log^{1-1/k} n \cdot \log\log n)$. The proof of this observation is given in the full version of the paper.

Next, we investigate the tightness of this bound. We find a graph on n vertices, such that any Ramsey tree cover with k trees requires distortion $\Omega(n^{1/k})$, significantly improving the $\Omega(\log_k n)$ bound obtained from high girth graphs. This also implies that our upper bound is tight, up to lower order terms.

Moreover, the graph we construct is series-parallel (in particular a planar graph) and also has $O(1)$ doubling dimension. Thus, our lower bound indicates a large separation between what can be achieved by a tree cover vs. a Ramsey tree cover; Our upper bounds give a tree cover with $O(1)$ trees and constant distortion for planar and doubling metrics (even $1+\epsilon$ distortion for the latter), as opposed to the $n^{\Omega(1)}$ distortion required with a constant number of Ramsey trees, for both planar and doubling metrics.

We also use a result of [15] to show a lower bound for planar and doubling metrics in the low distortion regime: there are n-point planar (in fact, series-parallel) doubling metrics, such that any Ramsey tree cover with distortion α must contain at least $n^{\Omega(1/(\alpha \log \alpha))}$ trees.

Overall, for general, planar and doubling metrics, our results solve the question of covering metrics by Ramsey trees, up to logarithmic terms, in every regime of parameters. See Table 2 for a concise description of previous and our results.

Table 2 Previous and our results on Ramsey trees for general, planar and doubling metrics.

Family	Reference	Number of trees	Distortion
General metrics	[38]	$O(\alpha \cdot n^{1/\alpha})$	$O(\alpha)$
	New	k	$O(n^{1/k} \cdot \log^{1-1/k} n)$
Planar & doubling metrics	**New**	k	$\Omega(n^{1/k})$
	New	$n^{\Omega(1/(\alpha \log \alpha))}$	α

1.3 Overview of Techniques

Tree cover for doubling metrics. The standard way to construct a $(1+\epsilon)$-spanner for doubling metrics is along the following lines [29, 32]: Choose a hierarchical collection of 2^i-nets (see Section 3 for definitions), and assign every vertex to its nearest net-point at the level i when it first leaves the net hierarchy; this creates a *net tree*. Then additional edges are added to other net-points within distance $\approx 2^i/\epsilon$. This spanner cannot be decomposed into a few trees as low distortion paths for the pairs use both the net tree and the additional edges.

We use a different approach for constructing a spanner, so that it can be decomposed to trees; We first partition the hierarchical net into a small number of well-separated sub-nets (so that in level i, distances between points in the sub-net are at least $2^i/\epsilon$). Then construct a tree for each hierarchical sub-net, by iterative clustering around the sub-net points in a bottom-up manner. In order to control the radius increase caused by the clustering of lower level sub-nets, we also take sufficiently large gaps between consecutive levels used in the same tree.

Tree cover for minor-free graphs. We apply the path separators of [4], asserting that graphs excluding a fixed minor have a separator consisting of $O(1)$ shortest paths (see Section 4 for the precise definitions). Adding for each point $O(1/\epsilon)$ edges to each shortest path guarantees small distortion for all separated pairs [48, 34]. However, since we desire trees, we can allow each point to add only 1 edge (per tree) to the path separator. Using a simple randomized algorithm to choose these edge connections, we show that w.h.p. all pairs will have a low distortion tree.

Hierarchical partitions. We construct a collection of HST spaces (a special type of ultra-metric spaces, see Section 5) via a hierarchical probabilistic partitions similarly to [19]. Yet instead of using the basic probabilistic partitions (e.g. [11]) we use the probabilistic partitions of [1], which have two main strong properties: The padding of the partition can be set as a parameter, which may also be a constant depending on distortion α; The partitions are local, i.e. the probability of being padded is not affected by the structure of the clusters that are far enough. This allows showing that intersecting a bundle of such independent partitions achieves a good tradeoff when the number of scale levels is small. Combining these with the idea of bottom-up union of clusters similar to that of [19], we are able to construct hierarchies with diameters of clusters decreasing by a *constant factor* while letting the padding parameter depend on α, obtaining more general and improved bounds.

1.4 Related Work

We note that Charikar et al. [21] studied a related question of bounding the number of trees sufficient for probabilistic embedding. The result they obtain implies an exponentially weaker cover size than those that follow from [11] and from our construction.

In [30], Gupta et al. considered a stronger version of tree covers (stronger than Ramsey tree covers), which they used to devise an oblivious algorithm for network design problems. We note that the lower bounds given in this paper show that the bound they get using this method is almost tight, even for doubling or planar metrics.

In the context of spanning trees, the problem of computing a spanning tree with low average stretch was first studied by Alon et al. [6]. Following Elkin et al. [23], Abraham et al. [2, 5] obtained a nearly tight $O(\log n \cdot \log\log n)$ bound.

2 Ramsey Tree Covers for General Metrics with Few Trees

In this section we show a deterministic Ramsey tree cover construction for general metrics. Unlike previous works, we build a cover with a small (possible constant) number of trees.

▶ **Theorem 2.** *For any n-point metric (X, d) and any $k \geq 1$, there is a deterministic algorithm that constructs a Ramsey tree cover for X of size k with distortion $O(n^{1/k} \cdot (\log n)^{1-1/k})$.*

Our deterministic construction follows directly from the following theorem on deterministic Ramsey embedding into a tree metric that was presented by Bartal [14] and later by Abraham et al. [3] (alternatively, a randomized construction can be based on [38]).

▶ **Theorem 3** ([14, 3]). *Let (X, d) be a metric space, fix any subset $S \subseteq X$, and let $\alpha \geq 1$ be a parameter. There is a deterministic algorithm that finds a subset $Z \subseteq S$, of size $|Z| \geq |S|^{1-\frac{1}{\alpha}}$, and an embedding f of X into an ultrametric T with distortion $O(\alpha)$ for any pair $(u, v) \in Z \times X$.*

Proof of Theorem 2. Let $S_1 = X$. For $i = 1, \ldots, k-1$ iteratively apply the algorithm of Theorem 3 on the subset S_i with parameter α (to be determined later), and obtain trees $T_1, \ldots, T_{k-1}$. The last tree T_k will be constructed separately. Let $Z_i \subseteq S_i$ be the set of size at least $|S_i|^{1-1/\alpha}$ guaranteed by Theorem 2, and define $S_{i+1} = S_i \setminus Z_i$. Note that every point $x \in X \setminus S_k$ has a tree with distortion $O(\alpha)$ for all pairs in $\{x\} \times X$. For each $i \geq 1$, we have $|S_{i+1}| = |S_i| - |Z_i| \leq |S_i| \left(1 - n^{-1/\alpha}\right)$. Therefore, after $k-1$ iterations we have $|S_k| \leq n \left(1 - n^{-1/\alpha}\right)^{k-1}$. We can embed the metric S_k into an ultrametric with distortion $|S_k| - 1$ by the embedding of [15, 32]. This embedding can be extended to all of X, with distortion $O(|S_k|)$ for pairs in $S_k \times X$ by [38, Lemma 4.1]. Therefore, α should be chosen so that $n \cdot \left(1 - n^{-1/\alpha}\right)^{k-1} \leq \alpha$. Using the inequality $e^x \geq 1 + x$ for all $x \in \mathbb{R}$ we have: $n \cdot \left(1 - n^{-1/\alpha}\right)^{k-1} = n \cdot \left(1 - e^{-\frac{\ln n}{\alpha}}\right)^{k-1} \leq n \cdot \left(\frac{\ln n}{\alpha}\right)^{k-1}$. Taking $\alpha = n^{1/k} \cdot (\ln n)^{\left(1-\frac{1}{k}\right)}$ gives a Ramsey tree cover with distortion $O(\alpha)$. ◀

3 $(1+\epsilon)$-Distortion Tree Covers for Doubling Metrics

In this section we devise a tree cover for doubling metrics with distortion arbitrarily close to 1. Let (X, d) be a metric with doubling constant λ, and fix $0 < \epsilon < 1/8$.

▶ **Definition 4.** *An r-net $N \subseteq X$ is a set satisfying: 1) For every $x, y \in N$, $d(x, y) > r$, and 2) For every $u \in X$ there exists $x \in N$ with $d(x, u) \leq r$. We say that a collection $\{N_i\}$ of 2^i-nets is* hierarchical *if $N_{i+1} \subseteq N_i$.*

It is well-known that a simple greedy algorithm can construct (hierarchical) nets. Also, it is known that the size of an r-net of a ball of radius R is bounded by $\lambda^{O(\log(R/r))}$.

Let $\{N_i\}$ be a hierarchical collection of 2^i-nets of X. (It suffices to take the indices i from the range $[\log(\epsilon\delta), \log \Delta]$ where $\delta = \min_{x\neq y\in X}\{d(x, y)\}$ and $\Delta = \max_{x,y\in X}\{d(x, y)\}$.)

▷ Claim 5. There is a partition of N_i to $t = \lambda^{O(\log(1/\epsilon))}$ sets $N_{i1}, \ldots, N_{it}$, so that for every $x, y \in N_{ij}$, $d(x, y) \geq 6/\epsilon \cdot 2^i$. It is also hierarchical: if $x \in N_{ij}$ then $x \in N_{i'j}$ for every $i' < i$.

Proof. First place in N_{ij} all the points of $N_{(i+1)j}$ for each j, and denote $N'_i = N_i \setminus (\bigcup_j N_{(i+1)j})$. Next, for $j = 1, 2, ..., t$ complete N_{ij} by choosing greedily from the points remaining in $N'_i \setminus (N_{i1} \cup \cdots \cup N_{i(j-1)})$. Since for any $x \in N_i$ the ball of radius $6/\epsilon \cdot 2^i$ contains less than t net points of N_i, we will surely pick x to some N_{ij} in some iteration $j \leq t$. ◁

Construction of trees. Assume w.l.o.g that $\log(1/\epsilon)$ is an integer, we will construct $t \cdot \log(1/\epsilon)$ trees (in fact, forests). The tree $T_{j,p}$ is indexed by the pair (j, p) with $1 \leq j \leq t$ and $0 \leq p < \log(1/\epsilon)$. Fix j and p, we now describe how to build $T_{j,p}$. Let $I_p = \{i : i \equiv p \pmod{\log(1/\epsilon)}\}$. Initially all points in X are unclustered. We go over all $i \in I_p$ (from small to large in order), and for every $x \in N_{ij}$ we add an edge from x to every unclustered point $y \in X$ satisfying $d(x, y) < 3/\epsilon \cdot 2^i$, of weight $d(x, y)$. These points connected to x are now clustered. (The center x is not considered clustered.)

The following observation is proved in the full version of the paper:

▶ **Observation 6.**

a. *No point $x \in N_{ij}$ is clustered when iteration i is complete.*
b. *Every $u \in X$ can be clustered by at most 1 point.*
c. *If C_x is the connected component created by the clustering of $x \in N_{ij}$ at level $i \in I_p$, then* $\mathrm{diam}(C_x) \leq 8/\epsilon \cdot 2^i$.

▷ Claim 7. When the process completes we have a forest.

Proof. By Observation 6(b) every point u adds at most a single edge to $T_{j,p}$, at the time it becomes clustered. As u adds this edge to an unclustered point, it cannot close a cycle. (More formally, if we give each vertex a time stamp which is the time it becomes clustered, then the single edge every vertex adds is to a vertex with a higher time stamp.) ◁

▷ Claim 8. Let C_x be the connected component created when we clustered points to x at some level $i \in I_p$. Then for every point $y \in B(x, 2/\epsilon \cdot 2^i)$ we have $d_{C_x}(x, y) \leq d(x, y) + 2^{i+4}$.

Proof. If y was unclustered when creating C_x then $d_{C_x}(x, y) = d(x, y)$ by definition. Otherwise, let z be the (unique) unclustered point in C_z, the connected component containing y before executing the clustering of level i. Let $i' < i$ be the level in which z created C_z, and by Observation 6(c) we have $\mathrm{diam}(C_z) \leq 8/\epsilon \cdot 2^{i'} \leq 8 \cdot 2^i$ (recall $i' \leq i - \log(1/\epsilon)$ since $i' \in I_p$). As $8 \leq 1/\epsilon$ it follows that $d(x, z) \leq d(x, y) + d(y, z) \leq 2/\epsilon \cdot 2^i + 8 \cdot 2^i \leq 3/\epsilon \cdot 2^i$, so x will cluster z (recall that by Observation 6(b) no other center can cluster z). Furthermore, $d_{C_x}(x, y) = d_{C_x}(x, z) + d_{C_z}(z, y) = d(x, z) + d_{C_z}(z, y) \leq d(x, y) + 2d_{C_z}(z, y) \leq d(x, y) + 2^{i+4}$. ◁

▶ **Lemma 9.** *For every $u, v \in X$, there is a tree $T = T_{j,p}$ so that $d_T(u,v) \le (1+O(\epsilon)) \cdot d(u,v)$.*

Proof. Choose i such that $2^{i-1}/\epsilon \le d(u,v) < 2^i/\epsilon$, and let $p = i \mod \log(1/\epsilon)$. Let $x \in N_i$ be the nearest net point to u, so that $d(x,u) \le 2^i$, and let $1 \le j \le t$ be such that $x \in N_{ij}$.

Observe that $d(x,v) \le d(x,u) + d(u,v) \le 2^i + 2^i/\epsilon < 2/\epsilon \cdot 2^i$, so by Claim 8

$$d_{C_x}(x,v) \le d(x,v) + 2^{i+4} \le d(x,u) + d(u,v) + 2^{i+4} = (1+O(\epsilon)) \cdot d(u,v).$$

We also have by the same claim that $d_{C_x}(x,u) \le d(x,u) + 2^{i+4} = O(\epsilon) \cdot d(u,v)$. The fact that C_x is a subtree of the forest $T = T_{j,p}$ completes the proof. ◀

Since we use edge weights that are the actual distances in (X,d), clearly $d_T \ge d$. Rescaling ϵ by a constant yields the following.

▶ **Theorem 10.** *For every metric (X,d) with doubling constant λ, and any $0 < \epsilon < 1$, there is an efficient algorithm to construct a tree cover of size $\lambda^{O(\log(1/\epsilon))}$, with distortion $1+\epsilon$.*

4 $(1+\epsilon)$-Distortion Tree Covers for Planar and Minor-Free Graphs

In this section we use *path-separators* for planar [48] and more generally minor-free graphs [4], to devise tree covers with $1+\epsilon$ distortion and $O((\log n)/\epsilon)^2$ trees. We start with some preliminary definitions.

A graph G has H as a minor if one can obtain H from G by a sequence of edge deletions, vertex deletions and edge contractions. The graph G is H-minor-free if it does not contain H as a minor.

▶ **Definition 11.** *A graph $G = (V,E)$ on n vertices is s-*path separable *if there exists an integer t and a separator $S \subseteq V$ such that:*

1. *$S = V(\mathcal{P}_0) \cup V(\mathcal{P}_1) \cup \cdots \cup V(\mathcal{P}_t)$, where for each $0 \le i \le t$, $\mathcal{P}_i$ is a collection of shortest paths in the graph $G \setminus (\bigcup_{0 \le j < i} \mathcal{P}_j)$ (and $V(\mathcal{P}_i)$ is the vertex set used by the paths in $\mathcal{P}_i$).*
2. *$\sum_{i=0}^{t} |\mathcal{P}_i| \le s$, that is, the total number of paths is at most s.*
3. *Each connected component of $G \setminus S$ is s-path separable and has at most $n/2$ vertices.*

▶ **Theorem 12** ([4])**.** *Every H-minor-free graph is s-path separable for some $s = s(H)$, and an s-path separator can be computed in polynomial time.*

The following Lemma is implicit in the works of [34, 48] (for completeness a proof is included in the full version of the paper):

▶ **Lemma 13.** *Let $G = (V,E)$ be an edge-weighted graph, fix any $0 < \epsilon < 1$, and let P be a shortest path in G. Then one can find for each $x \in V$ a set of landmarks L_x on P of size $|L_x| = O(1/\epsilon)$, such that for any $x, y \in V$ whose shortest path between them intersects P, there exists $u \in L_x$ and $v \in L_y$ satisfying $d_G(x,u) + d_P(u,v) + d_G(v,y) \le (1+\epsilon) \cdot d_G(x,y)$.*

Construction. Using these tools, we are ready to describe our tree cover for minor-free graphs. Apply the path separator of Theorem 12 on the input graph $G = (V,E)$, $|V| = n$, to obtain a collection $\mathcal{P}$ of s paths, and denote $S = V(\mathcal{P})$. For each path $P \in \mathcal{P}$, apply Lemma 13 to get a set of landmarks for each vertex, and let $\ell = \max_{x \in V}\{|L_x|\} = O(1/\epsilon)$ be the maximal size of a landmark set. Let T be a tree formed by taking P, and for each $x \in V \setminus V(P)$ add a single edge to $u \in L_x$ chosen uniformly and independently at random. Let $d_G(x,u)$ be the weight of a chosen edge. We pick $(C \log n)/\epsilon^2$ such trees independently for each path P, for sufficiently large constant C.

Next, we continue recursively on each connected component of $G \setminus S$. Since the number of vertices halves at every iteration, there will be $O(\log n)$ iterations. Furthermore, the trees of different connected components can be viewed as a forest of G (which can be arbitrarily completed to a tree), thus the total number of trees is $O((\log n)/\epsilon)^2$.

Analysis. Fix some $x, y \in V$, and let P be the first path in $\mathcal{P}$ that intersects the shortest path between x, y in G. (It may be the case that $\mathcal{P}$ is a path separator in a deep level of the recursion, when we decompose some subgraph G'. Note that $d_{G'}(x, y) = d_G(x, y)$, since no path intersected the shortest path from x to y so far. So w.l.o.g. we call the current graph G.) Let $u \in L_x$ and $v \in L_y$ be such that $d_G(x, u) + d_P(u, v) + d_G(v, y) \le (1 + \epsilon) \cdot d_G(x, y)$, which are guaranteed to exist by Lemma 13. If we choose a tree T that contains P and both edges $(x, u), (y, v)$, then T will have distortion at most $1 + \epsilon$ for the pair x, y. The probability that both x, y add these edges to T is at least $1/\ell^2 = \Omega(\epsilon^2)$. Thus, the probability that none of the trees created for the path P has distortion at most $1 + \epsilon$ for the pair x, y is at most $(1 - \Omega(\epsilon^2))^{(C \log n)/\epsilon^2} \le e^{-3 \ln n} = 1/n^3$, whenever C is sufficiently large. By the union bound over the $\binom{n}{2}$ pairs, with high probability all pairs have a tree with distortion $1 + \epsilon$ in that tree. We have proven the following.

▶ **Theorem 14.** *Let G be a graph on n vertices that is H-minor-free. For any $0 < \epsilon < 1$, there is a randomized efficient algorithm that w.h.p. constructs a tree cover for G containing $O((\log n)/\epsilon)^2$ trees with distortion $1 + \epsilon$. (The constant in the O-notation depends on $|H|$.)*

5 Tree Covers for Doubling Metrics with Distortion-Size Tradeoff

In this section we prove that any metric space with doubling constant λ has a tree cover with distortion $O(\alpha)$ of size $O(\lambda^{1/\alpha} \log \lambda \log \alpha)$.

Recall that ultrametric is a metric space obeying a strong form of the triangle inequality. It is well known that any finite ultrametric (U, ρ) can be represented by a finite labeled tree T, with the points of U being the leaves of T. Each node $u \in T$ has a label $\Delta(u) \ge 0$ and the label of each leaf is 0. For any two nodes u and v, such that v is a child of u, $\Delta(u) \ge \Delta(v)$. For $u, v \in U$, the distance $\rho(u, v)$ is defined to be the label of their least common ancestor. We refer to ultrametrics by their tree representation. If the labels in an ultrametric tree T are decreasing by a factor at most $\mu > 1$, then T is called a μ-Hierarchically Separated Tree metric (μ-HST) [11]. We note that an ultrametric space can also be represented as a shortest path metric on a Steiner tree.

For a finite metric (X, d), let $d_{\max} = \max_{x \neq y \in X}\{d(x, y)\}$, and $d_{\min} = \min_{x \neq y \in X}\{d(x, y)\}$. Let $\Phi(X) := d_{\max}/d_{\min}$ denote the aspect ratio of X.

5.1 Probabilistic Hierarchical Partition Family

With start with the neceassary definitions. For any $\Delta > 0$, a Δ-bounded partition P of a finite metric space (X, d) is a collection of pairwise disjoint clusters $P_i \subseteq X$, such that $\cup P_i = X$, and for each cluster $P_i \in P$, $\mathrm{diam}(P_i) \le \Delta$. We assume that each cluster has some point designated as its center. For a point $x \in X$, let $P(x) \in P$ denote the cluster that contains x. A Δ-bounded probabilistic partition of X is a distribution $\mathcal{P}$ over a set of Δ-bounded partitions of X.

The notion of a padding parameter of a random partition is studied in various papers [36, 33, 11, 27]. We use a stronger definition given by Abraham et al. in [1], where the padding parameter depends on the desired probability of success. The following is a rephrased version of their original definition ([1], Definition 17):

▶ **Definition 15** (Padded Probabilistic Partition). *Let $\eta(\delta)\colon (0,1] \to (0,1]$ be some function, and $(a,b] \subseteq (0,1]$ be some range. A Δ-bounded probabilistic partition $\mathcal{P}$ is $\eta(\delta)$-padded on the range $(a,b]$, if for all $x \in X$ and for all $\delta \in (a,b]$, $\Pr_{P\sim\mathcal{P}}[B(x,\eta(\delta)\cdot\Delta) \subseteq P(x)] \geq \delta$.*

In addition, the authors defined a notion of a *locally* padded probabilistic partition (on the range $(a,b]$): $\mathcal{P}$ is $\eta(\delta)$-locally padded if for all $a < \delta \leq b$ the event $B(x,\eta(\delta)\cdot\Delta) \subseteq P(x)$ occurs with probability at least δ regardless of the structure of the partition outside the ball $B(x,2\Delta)$. Formally stated, for all $x \in X$, for all subsets $C \subseteq X \setminus B(x,2\Delta)$ and all partitions P' of C, $\Pr_{P\sim\mathcal{P}}[B(x,\eta(\delta)\cdot\Delta) \subseteq P(x) \mid P[C] = P'] \geq \delta$, where $P[C]$ denotes the restriction of the partition P to C. Our construction uses their random partitions as a building block:

▶ **Lemma 16** ([1], Lemma 8). *Given a finite metric space X with doubling constant λ, and given any $0 < \Delta < \mathrm{diam}(X)$, there is a Δ-bounded, $\left(\frac{\log(1/\delta)}{2^6\log\lambda}\right)$-locally padded probabilistic partition $\mathcal{P}$ of X, for $\delta \in \left[\lambda^{-2^{12}}, 1\right]$.*

A set of nested partitions of X forms a hierarchy:

▶ **Definition 17** (Hierarchical Partition). *For all $\mu > 1$, $\Delta \leq d_{\max}(X)$ and integer $1 \leq B \leq \log_\mu \Phi(X)$, let $\Delta_i = \Delta/\mu^i$, for all $0 \leq i \leq B$. A μ-Hierarchical Partition of X for range $[\Delta, \Delta_B]$, is a collection $H = \{P_0,\ldots,P_B\}$ of partitions of X such that: For all $0 \leq i \leq B$, P_i is a Δ_i-bounded partition of X; Each P_{i+1} is a refinement of P_i, i.e. each cluster in P_i is a union of some clusters in P_{i+1}. Let μ-$HP_B(\Delta)$ denote such a collection.*

A full range Hierarchical Partition, denoted by μ-HP, is the μ-$HP_B(\Delta)$, for $\Delta = d_{\max}(X)$ and $B = \log_\mu \Phi(X)$ (we assume this is an integer).

There is a natural way to associate a dominating μ-HST tree to a μ-HP. For each cluster of the partition P_i there is a node in the tree. The nodes associated with clusters of the partition P_{i+1} are the children of nodes associated with clusters of P_i. The label of all level i nodes in the tree is Δ/μ^i. The points of X are at the leaves.

▶ **Definition 18** (η-Padded μ-Hierarchical Partition Family). *Let $\eta < 1$ and $\mu > 1$. For a finite metric space (X,d), an η-padded μ-Hierarchical Partition Family of X, (η,μ)-HPF, is a set $\mathcal{H}$ of μ-Hierarchical Partitions $\{H^j\}_{j\geq 1}$ of X such that: For all $x \in X$ and for all scales $0 \leq i \leq \log_\mu \Phi(X)$, there is an $H^j \in \mathcal{H}$ such that $B(x,\eta\Delta_i) \subseteq P_i^{(j)}(x)$, where $P_i^{(j)}$ is a Δ_i-bounded partition of H^j. The size of $\mathcal{H}$ is the number of hierarchical partitions it has.*

The following lemma shows the connection between hierarchical family and a tree cover:

▶ **Lemma 19.** *If there is an (η,μ)-HPF of size k of X, then there is an $(\mu/\eta,k)$-tree cover of X.*

Proof. Let $H^1,\ldots,H^k$ be an (η,μ)-HPF of X. Consider an associated collection of dominating μ-HST trees $T_1,\ldots,T_k$. Given any $x \neq y \in X$, let i be the minimal index such that $d(x,y) \geq \eta\Delta_i$. If $i = 0$, then by the construction, for any tree T_j, $d_{T_j}(x,y) \leq \Delta_0$, implying $d_{T_j}(x,y)/d(x,y) \leq 1/\eta$. If $i \geq 1$, then $\eta\Delta_i \leq d(x,y) \leq \eta\Delta_{i-1}$. The padding property implies that there is H^j such that $B(x,\eta\Delta_{i-1}) \subseteq P_{i-1}^{(j)}(x)$. As $y \in B(x,\eta\Delta_{i-1})$, it holds that $d_{T^j}(x,y) \leq \Delta_{i-1}$. Therefore, $d_{T^j}(x,y)/d(x,y) \leq \Delta_{i-1}/\eta\Delta_i = \mu/\eta$. ◀

In what follows, we construct $(\Omega(1/\alpha),2)$-HPF of X, of size $O(\lambda^{1/\alpha}\log\lambda\log\alpha)$. We note that the notion of hierarchical family also appeared in [35], where the authors constructed an $(\Omega(s^{-2})), O(s^2))$-HPF of size 3^s for any metric of a $K_{s,s}$-minor free graph. As a corollary, we conclude

▶ **Corollary 20.** *For any metric induced on a $K_{s,s}$-minor free graph, there is a tree cover with distortion $O(s^4)$, of size 3^s.*

In our proofs we will use the following version of the Lovasz Local Lemma:

▶ **Lemma 21** ([26]). *Let $\mathcal{E}_1, \ldots, \mathcal{E}_n$ be a family of events. Let $G(V, E)$ be a directed graph on n vertices with out-degree at most d, where each vertex corresponds to an event. Assume that for all $1 \le i \le n$, for all $Q \subseteq \{j \mid (\mathcal{E}_i, \mathcal{E}_j) \notin E\}$, $\Pr[\mathcal{E}_i \mid \bigcap_{j \in Q} \neg\mathcal{E}_j] \le p$. If $ep(d+1) \le 1$, then $\Pr\left[\bigcap_{i\in[1,n]} \neg\mathcal{E}_i\right] > 0$.*

5.2 Constructing Hierarchical Padded Family of Bounded Size

Our main hierarchical partitions result is:

▶ **Theorem 22.** *For any finite metric space X with doubling constant λ and for any $\alpha \ge 2$, there is an $\Omega(1/\alpha)$-padded 2-Hierarchical Partition Family of X, of size $O\left(\lambda^{1/\alpha} \log \alpha \log \lambda\right)$.*

Note that taking $\alpha = O(\log \lambda)$, we obtain a hierarchical family with padding $\Omega(1/\log \lambda)$, of size $O(\log \lambda \log \log \lambda)$, which is an improvement over the result of [19]: $O(\log \lambda)$-hierarchical partitions with padding $\Omega(1/\log \lambda)$, of the same size. They construct a family of hierarchies, where each hierarchy is constructed in a bottom-up manner: the clusters of larger diameters are the union of the clusters of lower diameters. Preserving the padding parameter requires the diameters of the clusters to increase by a factor of $O(\log \lambda)$, thus covering only $\log \log \lambda$ of all the distance scales in the metric space. This results in $O(\log^2 \lambda)$ distortion. Using the Lovasz Local Lemma they were able to bound the size of this family.

In our construction, we combine the bottom-up union of clusters technique with an *intersection of clusters* procedure. Essentially, there are two steps. First, we use the locally padded partitions of Lemma 16 to create a padded hierarchy with diameters decreasing by a constant factor, by intersecting the clusters of levels of the hierarchy of larger diameter. Using the locality property and the fact that the padding parameter depends on the success probability, we show that using $\log \log \lambda$ such levels of partitions with diameters increasing by factor 2, results in a 2-hierarchy with $\Omega(1/\alpha)$ padding, thus covering the $\log \log \lambda$ scales uncovered by the construction of [19]. We apply the Lovasz Local Lemma to bound the size of the family of such hierarchies by $O\left(\lambda^{1/\alpha} \log \alpha \log \lambda\right)$. Second, we combine the hierarchies obtained by cutting clusters, in a bottom-up manner, by defining higher scales clusters as the union of lower level clusters, thus obtaining a hierarchy with diameters decreasing by a factor of 2 in all its levels, while padding is $\Omega(1/\alpha)$.

To prove Theorem 22, we consider hierarchical partitions that cover a range of scales: for any Δ and an integer B we build a family $\{H^j\}_{j\ge1}$, where each H^j is a μ-$\mathrm{HP}_B(\Delta)$. The padding property is then required to hold for all points $x \in X$ and for all scales $\Delta_i \in [\Delta, \Delta_B]$. We call such family as (η, μ)-HPF for range $[\Delta, \Delta_B]$.

The following lemma is used as a subroutine in the construction of the hierarchical family:

▶ **Lemma 23.** *Let X be a finite metric space with doubling constant λ . For a given $\alpha \ge 2$, $\Delta \le \mathrm{diam}(X)$ and an integer $1 \le B \le \log_\mu \Phi(x)$, there exists an $(\Omega(1/\alpha), 2)$-HPF for range $[\Delta, \Delta_B]$, of size $O\left(\lambda^{1/\alpha} \log \lambda (\log \alpha + B)\right)$.*

Proof. For a given distortion $\alpha \ge 2$, let $\delta = \lambda^{-1/(2\alpha)}$. Therefore, for such δ we have $\eta(\delta) := \frac{\log(1/\delta)}{2^6 \log \lambda} = 2^{-7}/\alpha$. Also note that for any $\alpha \ge 1$, it holds that $\delta \in [\lambda^{-2^{12}}, 1]$. Thus, we will show that there exists a hierarchical family with padding $\Omega(\eta(\delta))$, of size $k := O(\left(\lambda^{1/\alpha} \log \lambda (\log \alpha + B)\right))$.

Let $N \subseteq X$ be an $(\eta(\delta)\Delta_B/4)$-net of X. We show the claim is true for N and the extension of it to X is immediate, with a constant factor loss in distortion. In the sequel, all the balls are balls of metric space N.

Let $\Delta_i = \Delta/2^i$, for all $0 \leq i \leq B$. Consider the following random process: For each scale Δ_i in the range $[\Delta, \Delta_B]$, independently generate Δ_i-bounded partitions $P_0, \ldots P_B$ of N by invoking the locally padded probabilistic decomposition of Lemma 16. To obtain a 2-Hierarchical Partition H for the scales $[\Delta_0, \Delta_B]$ we cut all the clusters of all the partitions, to get Δ_i-bounded nested partitions $\hat{P}_0, \ldots, \hat{P}_B$. Let $\hat{P}_0 = P_0$, for all $i \geq 1$, define $\hat{P}_i = \cup_{\hat{C} \in \hat{P}_{i-1}} \cup_{C \in P_i} C \cap \hat{C}$.

Now, independently repeat the above random process k times to obtain a randomly generated family $H^{(1)}, \ldots, H^{(k)}$ of 2-Hierarchical Partitions of the net N, for range $[\Delta, \Delta_B]$. Each hierarchical partition $H^{(t)}$ consists of Δ_i-bounded partitions, denoted by $\hat{P}_i^{(t)}$.

For each $x \in N$ and for each scale $\Delta_i \in [\Delta, \Delta_B]$, let $\mathcal{E}_{x,i}$ be an event that the ball $B(x, \eta(\delta)\Delta_i)$ is not padded at the i-th level partition $\hat{P}_i^{(t)}$ in any of the hierarchical partitions $H^{(1)}, \ldots, H^{(k)}$. We use the Lovasz Local Lemma (Lemma 21) to prove that for the chosen value of k, $\Pr\left[\bigcap_{\substack{x \in X,\\ 0 \leq i \leq B}} \neg \mathcal{E}_{x,i}\right] > 0$. Let $G = (V, E)$ be a directed graph with $V = \{\mathcal{E}_{x,i}\}$, for all $x \in N$ and $0 \leq j \leq B$. The vertex $\mathcal{E}_{x,i}$ is connected with an out-edge with all the verticies $\mathcal{E}_{y,j}$, such that $y \in B(x, 2\Delta)$ and $0 \leq j \leq B$. In the full version we prove the following lemma:

▶ **Lemma 24.** *For all $Q \subseteq N \backslash B(x, 2\Delta)$, for all $J \subseteq [0, B]$, $Pr[\mathcal{E}_{x,i} \mid \bigcap_{\substack{y \in Q,\\ j \in J}} \neg \mathcal{E}_{y,j}] \leq (1 - \delta^2)^k$.*

Then, for $\delta = \lambda^{-1/(2\alpha)}$, and $k = O\left(\lambda^{1/\alpha} \log \lambda (\log \alpha + B)\right)$, $(1 - \delta^2)^k \leq e^{-\delta^2 k} \leq \lambda^{-\Theta(\log \alpha + B)}$.

In addition, the out degree d of G is bounded by

$$d = B \cdot |N \cap B(x, 2\Delta)| = B \cdot O\left(\left(\frac{\Delta}{\eta(\delta)\Delta_B}\right)^{\log \lambda}\right) = B \cdot \lambda^{O(\log(1/\eta(\delta)) + B)} = \lambda^{O(\log(1/\eta(\delta)) + B)}.$$

Thus, the LLL can be applied to conclude the proof. ◀

Proof of Theorem 22. Let $\Phi = \Phi(X)$, $\Delta_0 = d_{\max}(X)$, and for all $1 \leq i \leq \log \Phi$, $\Delta_i = \Delta_0/2^i$. Let $\mathcal{I} = \{\Delta_i \mid 0 \leq \ i \leq \log \Phi\}$. We build a small family of 2-HP's, such that the padding property is satisfied for all points in X and for all scales $\Delta_i \in \mathcal{I}$, with padding parameter $\Omega(1/\alpha)$. Let $B = \lceil \log(2\alpha/c') \rceil$, where $c' < 1$ will be defined later, and let $k = O\left(\lambda^{1/\alpha} \log \alpha \log \lambda\right)$. We will build a family $\mathcal{H} \cup \mathcal{R}$ such that:

1. $\mathcal{H}$ is a collection of size k of 2-HP's. For[2] $\mathcal{I}_{\mathcal{H}} := \left\{\Delta_j \mid \Delta_j \in \bigcup_{0 \leq l \leq L} \left[\Delta_{2lB}, \Delta_{(2l+1)B}\right]\right\} \subseteq \mathcal{I}$, where $L = \log(\Phi^{1/(2B)}) - 1/2$, the following padding property holds: For all $x \in X$ and for all scale $\Delta_j \in \mathcal{I}_{\mathcal{H}}$, there is a 2-HP $H \in \mathcal{H}$ such that $B(x, \Omega(1/\alpha)\Delta_j) \subseteq P_j(x)$, for the j-th level partition $P_j \in H$.
2. $\mathcal{R}$ is a collection of size k of 2-HP's. The padding property as for $\mathcal{H}$ holds for all $x \in X$ and for all scales $\Delta_j \in \mathcal{I}_{\mathcal{R}} := \mathcal{I} \setminus \mathcal{I}_{\mathcal{H}}$.

Namely, the hierarchical partitions of $\mathcal{R}$ are padded for the scales that are not padded in the partitions of $\mathcal{H}$. Thus, together these two collections constitute an $\Omega(1/\alpha)$-padded 2-HP Family for X, of size $2k$. We describe the construction of $\mathcal{H}$, while $\mathcal{R}$ is constructed similarly.

---

[2] For the simplicity of representation, we assume that B is integer and that $\log \Phi$ is a multiple of B.

Let $\mathcal{H} = H^{(1)}, \ldots, H^{(k)}$ denote the set of 2-HP's. We construct it iteratively in a bottom up fashion. Assume by induction that we have already constructed a family $\hat{\mathcal{H}} = \hat{H}^{(1)}, \ldots, \hat{H}^{(k)}$ such that: Each $\hat{H}^{(t)}$ is a 2-HP for range $[\Delta_{2B}, \Delta_{(2L+1)B}]$; The padding property holds with parameter $\Omega(1/\alpha)$ for all scales $\Delta_j \in \mathcal{I}_{\mathcal{H}} \setminus \{\Delta_i \in [\Delta_0, \Delta_B]\}$.

Let $c = 1 + \frac{1}{2^{B-1}-1}$, by Lemma 23 there is $(\Omega(1/\alpha), 2)$-HPF $F^{(1)}, \ldots, F^{(k)}$ for range $[\tilde{\Delta}_0, \tilde{\Delta}_B]$, where $\tilde{\Delta}_j = \Delta_j/c$, for $0 \le j \le B$. For each $1 \le t \le k$, $H^{(t)}$ is obtained by adding the partitions of $F^{(t)}$ to $\hat{H}^{(t)}$ in the following way. Let $\hat{P}^{(t)}_{2B}$ denote the Δ_{2B}-bounded partition of $\hat{H}^{(t)}$. First, for all scale $\Delta_j \in [\Delta_{B+1}, \Delta_{2B}]$ add to $H^{(t)}$, Δ_j-bounded partition $P^{(t)}_j := \hat{P}^{(t)}_{2B}$ (these artificial partitions are added to have a well defined 2-HP family). Next, let $\{\tilde{P}^{(t)}_j\}_{0 \le j \le B}$ denote the set of $\tilde{\Delta}_j$-bounded partitions of $F^{(t)}$. For all j starting from $j = B$ down to $j = 0$, the partition $P^{(t)}_j$ is constructed a s follows: for each $\tilde{C} \in \tilde{P}^{(t)}_j$, add a cluster C to $P^{(t)}_j$, defined by $C = \cup\{C' \in P^{(t)}_{2B} \mid \text{center of } C' \in \tilde{C}\}$. Finally, the partitions of $\hat{H}^{(t)}$ are unchanged.

For all $B \le j \le 2B$, $P_j(t)$ is Δ_j-bounded, since $\Delta_{2B} \le \Delta_j$. For $0 \le j \le B$ the diameter of each cluster in partition $P^{(t)}_j$ is bounded by $\tilde{\Delta}_j + 2\Delta_{2B} = \Delta_j/c + \Delta_B/2^{B-1} \le \Delta_j \left(1/c + 1/2^{B-1}\right) = \Delta_j$, for a chosen value of c. In addition, by the construction, the partitions $P^{(t)}_j$ form a hierarchy. It is left to show that the padding property holds in $\mathcal{H}$ for the scales $[\Delta_0, \Delta_B]$. By Lemma 23, for any $x \in X$, for any $\tilde{\Delta}_j \in [\tilde{\Delta}_0, \tilde{\Delta}_B]$, there is $F^{(t)}$ such that $B(x, (c'/\alpha)\tilde{\Delta}_j) \subseteq \tilde{P}^{(t)}_j(x)$, for $\tilde{P}^{(t)}_j \in F^{(t)}$, for some constant c'. Consider some cluster $P^{(t)}_j(x)$, for some $x \in X$. In the process of constructing $P^{(t)}_j(x)$ some points from $\tilde{P}^{(t)}_j(x)$ may be removed, due to removal of some cluster $C' \in P^{(t)}_{2B}$ whose center falls outside the cluster $\tilde{P}^{(t)}_j(x)$. For B as defined above, for $r = \frac{c'}{\alpha}\tilde{\Delta}_j - \Delta_{2B} \ge \left(\frac{c'}{2\alpha}\right)\Delta_j$, we have that $B(x, r) \subseteq P^{(t)}_j(x)$. This completes the proof. ◀

▶ **Theorem 25.** *For any finite metric space (X, d) with doubling constant λ, for any $\alpha \ge 2$, there is a tree cover of X with distortion $O(\alpha)$ and of size $O\left(\lambda^{1/\alpha} \log \alpha \log \lambda\right)$.*

Proof. Apply Lemma19 on the Hierarchical Family of Theorem 22. ◀

Note that the tree cover of Theorem 25 can be deterministically constructed in polynomial time via the constructive local lemma due to [39].

6 Lower Bounds

In full version of the paper we show that there is an n-point metric (X, d) (the metric of a high girth graph) with doubling constant λ, such that any tree cover for X with distortion α requires at least $\Omega(\lambda^{1/\alpha})$ trees. Additionally, we show below a nearly tight lower bound for Ramsey tree covers of a doubling planar metric space.

6.1 Lower Bound on Ramsey Tree Cover

In this section we show an asymptotically tight lower bound on Ramsey tree covers, in the regime where the number of trees is small. The lower bound on high girth graphs mentioned above can only give distortion $\Omega(\log_k n)$ for tree covers with k trees. Here we use a different example (which is a planar metric with $O(1)$ doubling dimension) that strengthen the lower bound on the distortion to $\Omega(n^{1/k})$.

We will need the following notion of a composition of metric spaces that was introduced in [15] (we present here a simplification of the original definition).

▶ **Definition 26.** *Let (S, d_S), (T, d_T) be finite metric spaces. For $\beta \geq 1/2$, the β-composition of S with T, denoted by $Z = S_\beta[T]$, is a metric space of size $|Z| = |S| \cdot |T|$ constructed by replacing each point $u \in S$ with a copy of T, denoted by $T^{(u)}$. Let $\gamma = \frac{\max_{t \neq t' \in T}\{d_T(t,t')\}}{\min_{s \neq s' \in S}\{d_S(s,s')\}}$. For $z_i \neq z_j \in Z$ such that $z_i \in T^{(u)}$ and $z_j \in T^{(v)}$ the distance is defined as follows: if $u = v$, then $d_Z(z_i, z_j) = \frac{1}{\beta\gamma} \cdot d_T(z_i, z_j)$, otherwise (if $u \neq v$), $d_Z(z_i, z_j) = d_S(u, v)$.*

It is easily checked that the choice of the factor $1/(\beta\gamma)$ guarantees that d_Z is indeed a metric. For a finite metric space S and an integer $t \geq 1$, let $[S]^t_\beta$ denote the metric space obtained by β-composition of S with itself t times. The following theorem asserts that when the number of trees is small our upper bound on the distortion of Ramsey tree covers is tight up to a logarithmic factor. Although the example is not described as a planar and doubling metric space, in the full version of the paper we show that it can be approximated with constant distortion by a shortest path metric on a series-parallel graph with constant doubling dimension.

▶ **Theorem 27.** *For any $k \geq 1$ and large enough n, there is an n-point doubling metric space X, such that any Ramsey tree cover of X of size k, has distortion $\Omega\left(n^{\frac{1}{k}}\right)$.*

Proof. Let C_N denote the shortest path metric on the unweighted N-point cycle graph. For any integers $k, N \geq 1$ and for any $\beta \geq 1/2$, consider the metric space $Z_k(N) = [C_N]^k_\beta$. We prove by induction on k, that any Ramsey tree cover of $Z_k(N)$ with k trees has distortion at least $\frac{1}{3}|Z_k(N)|^{1/k} - 1$. The details are differed to the full version of the paper. ◀

References

1. Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026–3126, 2011. doi:10.1016/j.aim.2011.08.003.
2. Ittai Abraham, Yair Bartal, and Ofer Neiman. Embedding Metrics into Ultrametrics and Graphs into Spanning Trees with Constant Average Distortion. *SIAM J. Comput.*, 44(1):160–192, 2015. doi:10.1137/120884390.
3. Ittai Abraham, Shiri Chechik, Michael Elkin, Arnold Filtser, and Ofer Neiman. Ramsey Spanning Trees and their Applications. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 1650–1664, 2018. doi:10.1137/1.9781611975031.108.
4. Ittai Abraham and Cyril Gavoille. Object location using path separators. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing, PODC 2006, Denver, CO, USA, July 23-26, 2006*, pages 188–197, 2006. doi:10.1145/1146381.1146411.
5. Ittai Abraham and Ofer Neiman. Using petal-decompositions to build a low stretch spanning tree. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 395–406, 2012. doi:10.1145/2213977.2214015.
6. 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.
7. Ingo Althöfer, Gautam Das, David Dobkin, and Deborah Joseph. Generating sparse spanners for weighted graphs. In John R. Gilbert and Rolf Karlsson, editors, *SWAT 90*, pages 26–37, Berlin, Heidelberg, 1990. Springer Berlin Heidelberg.
8. S. Arya, G. Das, D. M. Mount, J. S. Salowe, and M. H. M. Smid. Euclidean spanners: short, thin, and lanky. In *Proc. of 27th STOC*, pages 489–498, 1995.
9. Baruch Awerbuch. Complexity of network synchronization. *J. ACM*, 32(4):804–823, October 1985. doi:10.1145/4221.4227.
10. Baruch Awerbuch, Bonnie Berger, Lenore Cowen, and David Peleg. Near-Linear Cost Sequential and Distribured Constructions of Sparse Neighborhood Covers. In *34th Annual Symposium on Foundations of Computer Science, Palo Alto, California, USA, 3-5 November 1993*, pages 638–647, 1993. doi:10.1109/SFCS.1993.366823.

11 Yair Bartal. Probabilistic Approximations of Metric Spaces and Its Algorithmic Applications. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 184–193, 1996.

12 Yair Bartal. On Approximating Arbitrary Metrices by Tree Metrics. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 161–168, 1998. `doi:10.1145/276698.276725`.

13 Yair Bartal. Graph Decomposition Lemmas and their Role in Metric Embedding Methods. In *12th Annual European Symposium on Algorithms*, pages 89–97, 2004. URL: `http://springerlink.metapress.com/openurl.asp?genre=article&issn=0302-9743&volume=3221&spage=89`.

14 Yair Bartal. Lecture notes. In *Metric embedding theory and its algorithmic applications course*, 2011. URL: `http://moodle.cs.huji.ac.il/cs10/file.php/67720/GM_Lecture6.pdf`.

15 Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric ramsey-type phenomena. *Annals Math*, 162(2):643–709, 2005.

16 S. Baswana and S. Sen. A Simple Linear Time algorithm for Computing a $(2k-1)$-Spanner of $O(n^{1+1/k})$ Size in Weighted Graphs. In *Proceedings of the 30th International Colloquium on Automata, Languages and Programming*, volume 2719 of *LNCS*, pages 384–396. Springer, 2003.

17 Guy E. Blelloch, Yan Gu, and Yihan Sun. A New Efficient Construction on Probabilistic Tree Embeddings. *CoRR*, abs/1605.04651, 2016. `arXiv:1605.04651`.

18 Béla Bollobas. *Extremal Graph Theory*. Dover Publications, Inc., New York, NY, USA, 2004.

19 T-H. Hubert Chan, Anupam Gupta, Bruce M. Maggs, and Shuheng Zhou. On Hierarchical routing in Doubling Metrics. In *Proc. 16th Ann. ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2005.

20 B. Chandra, G. Das, G. Narasimhan, and J. Soares. New sparseness results on graph spanners. *Int. J. Comput. Geometry Appl.*, 5:125–144, 1995.

21 Moses Charikar, Chandra Chekuri, Ashish Goel, Sudipto Guha, and Serge Plotkin. Approximating a Finite Metric by a Small Number of Tree Metrics. In *FOCS '98: Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 379, Washington, DC, USA, 1998. IEEE Computer Society.

22 Edith Cohen. Fast algorithms for constructing t-spanners and paths with stretch t. In *34th Annual Symposium on Foundations of Computer Science, Palo Alto, California, USA, 3-5 November 1993*, pages 648–658, 1993. `doi:10.1109/SFCS.1993.366822`.

23 Michael Elkin, Yuval Emek, Daniel A. Spielman, and Shang-Hua Teng. Lower-stretch spanning trees. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 494–503, New York, NY, USA, 2005. ACM Press. `doi:10.1145/1060590.1060665`.

24 Michael Elkin and David Peleg. (1+epsilon, beta)-Spanner Constructions for General Graphs. *SIAM J. Comput.*, 33(3):608–631, 2004. `doi:10.1137/S0097539701393384`.

25 Michael Elkin and Seth Pettie. A Linear-Size Logarithmic Stretch Path-Reporting Distance Oracle for General Graphs. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 805–821, 2015. `doi:10.1137/1.9781611973730.55`.

26 P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and finite sets*, 10(2):609–â€“627, 1975.

27 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, November 2004. `doi:10.1016/j.jcss.2004.04.011`.

28 Arnold Filtser and Ofer Neiman. Light Spanners for High Dimensional Norms via Stochastic Decompositions. In *26th Annual European Symposium on Algorithms*, 2018.

29 Jie Gao, Leonidas J. Guibas, and An Nguyen. Deformable Spanners and Applications. In *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, SCG '04, pages 190–199, New York, NY, USA, 2004. ACM. doi:10.1145/997817.997848.

30 Anupam Gupta, Mohammad T. Hajiaghayi, and Harald Räcke. Oblivious Network Design. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*, SODA '06, pages 970–979, Philadelphia, PA, USA, 2006. Society for Industrial and Applied Mathematics. URL: http://dl.acm.org/citation.cfm?id=1109557.1109665.

31 Anupam Gupta, Amit Kumar, and Rajeev Rastogi. Traveling with a Pez Dispenser (or, Routing Issues in MPLS). *SIAM J. Comput.*, 34(2):453–474, 2004. doi:10.1137/S0097539702409927.

32 Sariel Har-Peled and Manor Mendel. Fast Construction of Nets in Low-Dimensional Metrics and Their Applications. *SIAM J. Comput*, 35(5):1148–1184, 2006.

33 Philip Klein, Serge A. Plotkin, and Satish Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, STOC '93, pages 682–690, New York, NY, USA, 1993. ACM. doi:10.1145/167088.167261.

34 Philip N. Klein. Preprocessing an undirected planar network to enable fast approximate distance queries. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA.*, pages 820–827, 2002. URL: http://dl.acm.org/citation.cfm?id=545381.545488.

35 Robert Krauthgamer, James R. Lee, Manor Mendel, and Assaf Naor. Measured descent: A new embedding method for finite metrics. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 434–443. IEEE, October 2004.

36 Nathan Linial and Michael Saks. Decomposing graphs into regions of small diameter. In *SODA '91: Proceedings of the second annual ACM-SIAM symposium on Discrete algorithms*, pages 320–330, Philadelphia, PA, USA, 1991. Society for Industrial and Applied Mathematics.

37 Richard J. Lipton and Robert Endre Tarjan. Applications of a Planar Separator Theorem. *SIAM J. Comput.*, 9(3):615–627, 1980. doi:10.1137/0209046.

38 Manor Mendel and Assaf Naor. Ramsey partitions and proximity data structures. *Journal of the European Mathematical Society*, 9(2):253–275, 2007.

39 Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovasz Local Lemma. *CoRR*, abs/0903.0544, 2009. arXiv:0903.0544.

40 Assaf Naor and Terence Tao. Scale-oblivious metric fragmentation and the nonlinear Dvoretzky theorem. *Israel Journal of Mathematics*, 192(1):489–504, 2012. doi:10.1007/s11856-012-0039-7.

41 G. Narasimhan and M. Smid. *Geometric Spanner Networks*. Cambridge University Press, 2007.

42 D. Peleg and A. Schäffer. Graph Spanners. *J. Graph Theory*, 13:99–116, 1989.

43 D. Peleg and E. Upfal. A tradeoff between size and efficiency for routing tables. *J. of the ACM*, 36:510–530, 1989.

44 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. URL: http://link.springer.de/link/service/journals/00454/bibs/19n1p79.html.

45 Satish Rao and Warren D. Smith. Approximating Geometrical Graphs via "Spanners" and "Banyans". In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 540–550, 1998. doi:10.1145/276698.276868.

46 M. Thorup and U. Zwick. Approximate Distance Oracles. In 33^{rd} *Annual ACM Symposium on Theory of Computing (STOC)*, pages 183–192, Hersonissos, Crete, Greece, July 2001.

47 M. Thorup and U. Zwick. Spanners and emulators with sublinear distance errors. In *Proc. of Symp. on Discr. Algorithms*, pages 802–809, 2006.

48 Mikkel Thorup. Compact oracles for reachability and approximate distances in planar digraphs. *J. ACM*, 51(6):993–1024, 2004. doi:10.1145/1039488.1039493.