Bi-Lipschitz Extensions and Outlier Embeddings into Trees

Shuchi Chawla
Department of Computer Science, University of Texas at Austin, TX, USA

Arnold Filtser
Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel

Kristin Sheridan
Department of Computer Science, University of Texas at Austin, TX, USA

Yonatan Trachtenberg
Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel

Abstract

We develop low distortion embeddings with outliers from arbitrary metrics into hierarchically separated trees (HSTs). In particular, we develop an efficient algorithm that for any $\epsilon > 0$, given an input metric (X, d), and a probabilistic embedding of all but k points from X into HSTs with distortion c, samples from a probabilistic embedding of all but $O(\frac{k}{\epsilon}\log k)$ points into HSTs that achieves distortion at most $(32+\epsilon)c$.

Our results are based on two key technical components. First, we extend an algorithm of Munagala et al. [53] for minimizing the distortion of embeddings without outliers into HSTs to the setting with outliers. We combine this with new results on bi-Lipschitz extensions into trees and ℓ_1 space. In particular, we show that any probabilistic embedding into HSTs can be extended to k additional points with only a factor $O(\log k)$ of additional distortion. This bi-Lipschitz extension result utilizes a new probabilistic partitioning scheme that we call *onion partitioning*.

2012 ACM Subject Classification Theory of computation → Rounding techniques

Keywords and phrases metric embeddings, hierarchically separated trees, outliers

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.17

Category APPROX

Related Version *Full Version*: `https://arxiv.org/abs/2601.15470`

Funding *Shuchi Chawla*: This research was funded in part by NSF awards CCF-2217069 and CCF-2225259.
Arnold Filtser: This research was supported by the ISRAEL SCIENCE FOUNDATION (grant No. 1042/22).
Kristin Sheridan: This research was funded in part by NSF awards CCF-2217069 and CCF-2225259 and by JPMorgan Chase & Co. Any views or opinions expressed herein are solely those of the authors listed, and may differ from the views and opinions expressed by JPMorgan Chase & Co. or its affiliates. This material is not a product of the Research Department of J.P. Morgan Securities LLC. This material should not be construed as an individual recommendation for any particular client and is not intended as a recommendation of particular securities, financial instruments or strategies for a particular client. This material does not constitute a solicitation or offer in any jurisdiction.

1 Introduction

Low-distortion embeddings into trees are a fundamental algorithmic tool for reducing optimization problems on general graph metrics to the special case of tree metrics, where they are often easier to solve. The distortion of an embedding, defined as the maximum multiplicative change in pairwise distances, directly determines the loss in performance incurred by such a reduction. A series of works [7, 12, 13, 37] showed that *probabilistic* embeddings can embed any n-point metric into hierarchically separated trees (HSTs) with expected distortion $O(\log n)$. While uniform distortion bounds are helpful in understanding the general effectiveness of using metric embeddings to solve a problem, from a practical perspective, we may ask how well a particular given metric (X, d) can embed into trees. Indeed, for some metrics the optimal distortion is $\Omega(\log n)$ but for others it may be much smaller, e.g. $O(1)$, providing better algorithmic results.

In this paper we investigate instance-optimal distortion for probabilistic embeddings into trees by exploring the role of outliers. For a finite metric (X, d), a (k, c)-outlier embedding is a map defined on "most of" X – that is on all but k vertices, which we designate as outliers – into another metric space Y such that the distortion over the non-outliers is bounded by c. Intuitively, we can view the outliers as noisy or adversarial data that may hurt the quality of the instance-optimal embedding from X into Y. The goal of the outlier embedding is to identify and remove these outliers upfront so as to obtain a much better embedding over the remaining points. Algorithmically, given a target distortion c, we want to find the smallest k such that the removal of k outliers enables embedding the rest of the metric with distortion at most c. In this paper, we develop probabilistic outlier embeddings into trees, providing a tradeoff between outlier set size and distortion. Our main result is as follows.

▶ **Theorem 1** (HSTs admit outlier embeddings). *Let (X, d) be any metric that admits a (k, c) probabilistic outlier embedding into HSTs. Then for any constant $\epsilon > 0$, there exists an algorithm that efficiently samples from a probabilistic embedding into HSTs of some $S \subseteq X$ with $|X \setminus S| \leq O(\frac{1}{\epsilon} k \log k)$ and expected distortion at most $(32 + \epsilon)c$.*

Technical contribution: improved bi-Lipschitz extensions

A key technical component in our construction is a bi-Lipschitz extension for embeddings of arbitrary finite metrics into trees. Given a finite metric (X, d) and a low distortion embedding α from a subset $S \subset X$ into a host space Y, a bi-Lipschitz extension extends the domain of α to the entire set X, while only modestly increasing the distortion of the embedding. We show that if the distortion of the original embedding α is c_α, it is possible to obtain a bi-Lipschitz extension with distortion at most $O(c_\alpha \log k)$ where $k = |X \setminus S|$. Formally, we achieve the following bounds, where the expansion factor L_{exp} is the multiplicative increase in pairwise distances beyond the expansion of the original embedding α, and the contraction factor L_{contr} is the multiplicative decrease beyond the contraction in α.

▶ **Theorem 2** (HSTs have probabilistic bi-Lipschitz extensions). *Let $\mathcal{H}$ be the class of all HSTs. For any subset $S \subseteq X$ with $k := |X \setminus S|$ and a probabilistic embedding $\mathcal{D}_S : S \hookrightarrow \mathcal{H}$, there exists a probabilistic bi-Lipschitz extension of $\mathcal{D}_S$, $\mathcal{D} : X \hookrightarrow \mathcal{H}$, with expansion factor $L_{exp} = O(\log k)$ and contraction factor $L_{contr} = 4$.*

Further, we can efficiently sample from the distribution $\mathcal{D}$ if we can efficiently sample from $\mathcal{D}_S$.

A connection between outlier embeddings and Lipschitz extensions was first established by Chawla and Sheridan [26] in the context of deterministic embeddings into normed metrics. They showed that deterministic embeddings into ℓ_2 admit a Lipschitz extension with expansion factor $O(\log k)$ (but potentially unbounded contraction), and this suffices to obtain a bicriteria approximation for outlier embeddings into ℓ_2. We extend this approach to embeddings into HSTs by combining it with an LP-rounding approximation algorithm of Munagala et al. [53] for instance-optimal probabilistic embedding into HSTs.

Probabilistic extension and bounded contraction

Extending the approach of [26] to HSTs poses two new technical challenges. First, while Chawla and Sheridan [26]'s construction can tolerate arbitrary contraction in the Lipschitz extension, in order to employ Munagala et al.'s LP relaxation for HST embeddings, we necessarily need to control contraction as well. The second challenge is applying the notion of Lipschitz extension to probabilistic embeddings. This is important because probabilistic tree embeddings can obtain exponentially smaller distortion than deterministic tree embeddings. Consider, in particular, a distribution $\mathcal{D}$ over embeddings from some subset S of a given metric (X, d) into a host space Y. Simply drawing an embedding α from $\mathcal{D}$ and extending it to X may result in distortion $\propto \mathbb{E}_{\alpha\sim\mathcal{D}}[\max_{u,v\in X}(d_\alpha(u, v)/d(u, v))]$. This can be much larger than the desired distortion $\max_{u,v\in X}(\mathbb{E}_\alpha[d_\alpha(u, v)]/d(u, v))$.

We resolve both challenges by constructing a bi-Lipschitz extension through a sequence of "merge" operations, each of which combines two embeddings over domains that intersect in a single point. By carefully aligning and stitching each such pair of embeddings, we ensure strong bounds on both contraction and expansion.

Improved expansion via "onion-like" partitions

Chawla and Sheridan's Lipschitz extension for ℓ_2 employs a probabilistic partitioning procedure over the outliers $X \setminus S$ similar to that of Calinescu, Karloff, and Rabani [22], and attaches an embedding of each component of the partition to its closest point in S (that we call the *anchor point*). This approach combined with the merge operation described above yields a *bi-Lipschitz* extension for HSTs with expansion factor $O(\log^2 k)$. We develop a new technique to obtain a quadratic improvement over this factor, leading to Theorem 2. Our key observation is that the largest distortion is suffered by pairs that are separated into different components by the CKR partitioning procedure – one logarithmic factor arises from the probability of separation, and the second arises from the distortion between the points and their respective anchor points.

To improve upon this distortion, we construct a probabilistic partitioning where each component has an "onion-shell" shape – all points are roughly equidistant from the anchor point. We then employ an embedding of each component that preserves distances to the anchor point within a constant factor (while potentially inflating other distances by a logarithmic factor). This allows us to bound the distortion between any pair of points by $O(\log k)$ times the distortion of the embedding over S.

Our improved partitioning approach also applies to ℓ_1, providing a bi-Lipschitz extension with expansion factor $O(\log k)$. [26] previously obtained a bi-Lipschitz extension for ℓ_1 with expansion factor $O(\frac{1}{c_S}\log^2 k + \log k)$, where c_S is the distortion of the embedding of S into ℓ_1; this can be a logarithmic factor worse than the bound we achieve when c_S is $O(1)$.

▶ **Theorem 3** (ℓ_1 has bi-Lipschitz extensions, informal). *Deterministic ℓ_1 embeddings have deterministic bi-Lipschitz extensions with expansion factor $O(\log k)$ and contraction factor $O(1)$.*

Related work

Embedding into HSTs has been studied primarily with the objective of minimizing expected distortion. See e.g. [7, 12, 13, 24, 37, 14, 46]. In our work, we will use as a subroutine the algorithm of Fakcharoenphol et al. [37], which for any n point metric samples from an $O(\log n)$ distortion probabilistic embedding into HSTs. *Instance-optimal* distortion has been studied for target spaces such as constant dimensional Euclidean space [10, 51, 30, 31, 58], the line [9, 51, 38, 57], trees [11, 27], and ultrametrics [6, 53]. The goal in such embeddings is to minimize the distortion needed for the given *specific* (X, d), which in some cases may be smaller than the best bound achievable for all metrics of that size.

The notion of outlier embeddings was initially defined by Sidiropolous et al. [59] when they studied deterministic embeddings into ultrametrics, trees, and constant-dimensional ℓ_2 space. They consider approximating *deterministic, isometric* outlier embeddings (where the distortion c is 1) and bicriteria approximations with additive distortion. The structural properties used for isometric approximations don't apply to non-isometric embeddings, and techniques for bounding additive distortion tend to be quite different from those for multiplicative distortion, the focus of this work. Chubarian and Sidiropolous [29] continued the work on outlier embeddings and studied non-isometric outlier embeddings of unweighted graph metrics into the line. Later, Chawla and Sheridan [26] studied deterministic outlier embeddings into ℓ_2 space with unbounded dimension.

Several related notions of distortion were previously studied. One notion is *embeddings with slack*, in which the goal is to minimize the number of pairs that are distorted by more than some target factor [1, 23, 47]. An even stronger notion is that of *scaling distortion*, where for every $\varepsilon > 0$, at most ε fraction of the pairs suffer from distortion $\geq f(\varepsilon)$ (for some function f), see [45, 3, 4, 16]. Another notion is *terminal distortion*, where instead of an outlier set, there is a subset of terminals $K \subseteq X$, and the goal is only to preserve distance pairs involving a terminal $K \times X$ [32, 49, 56, 28]. A stronger notion is *prioritized distortion* [33, 34, 39] where there is a priority ordering over the vertices, and the distortion guarantee is w.r.t. the priorities. Another notion is that of local embeddings [2, 8, 25] where the goal is to minimize the distortion from each point to its k nearest neighbors.

Outlier embeddings are also closely related to metric Ramsey theory [21, 17, 19, 18, 52, 55, 20, 5, 15, 41, 42]. Given inputs n and c, metric Ramsey theory asks: What is the largest m such that *all* metrics of size n have a submetric of size m that embeds into some target space with distortion at most c? In this paper we are primarily focused on *instance-optimal* solutions rather than universal bounds that apply to all metrics, and we seek to approximate the size of the outlier set rather than the size of the non-outlier set.

Many other types of embedding compositions or extensions have been considered. One particular example is the notion of a Lipschitz extension [44, 48, 54], which seeks to leave non-outliers unaltered from the original embedding and ensure the expansion (but not necessarily the contraction) of every pair of nodes is bounded. Given an embedding on $S \subseteq X$, Calinescu, Karloff, and Rabani [22] develop a partitioning algorithm for $X \setminus S$ that implicitly defines Lipschitz extensions with a contraction factor at most $O(\log |S|)$ for all metric spaces. Chawla and Sheridan [26] define essentially the same partitioning algorithm, except that in their algorithm, [22] pick a uniformly random permutation over the set S, but [26] take a random permutation over the set of $X \setminus S$'s closest neighbors in S. The resulting Lipschitz extension factor is $O(\log |X \setminus S|)$. We will permute over the same set as [26], but we follow the cleaner analysis of [22]. Another related notion is that of 0-extension [22, 36, 35, 40] where one seeks a stochastic contraction $f : X \to K$ from the metric point set to a subset K of terminals while minimizing $\max_{x,y \in X} \mathbb{E}\left[\frac{d_X(f(x),f(y))}{d_X(x,y)}\right]$.

2 Definitions and results

Embeddings and distortion

A metric space is a set of points X with distances $d_X(\cdot, \cdot)$ between them that obey the triangle inequality. That is, for all $x, y, z \in X$, $d_X(x, z) \leq d_X(x, y) + d_X(y, z)$. We assume without loss of generality (via scaling) that in the metrics we consider, the minimum distance between distinct pairs of points is at least 1. For subsets $S, T \subseteq X$, denote $d(S,T) := \min_{s \in S, t \in T} \{ d(s,t) \}$ If $\alpha : X \to Y$ is a map from (X, d_X) to (Y, d_Y), we write $d_\alpha(x,y) := d_Y(\alpha(x), \alpha(y))$ to denote the distance between the images of two points $x, y \in X$ under this map.

$\mathcal{D}$ is a probabilistic embedding of (X, d_X) into $\mathcal{Y}$ if for each function α in the support of $\mathcal{D}$, $\alpha : X \to Y(\alpha)$ for some $(Y(\alpha), d_{Y(\alpha)}) \in \mathcal{Y}$. This embedding has expected *expansion* $c_{exp} \geq 1$, *contraction* $c_{contr} \geq 1$, and expected *distortion* $c_{contr} \cdot c_{exp}$ if for all $x, y \in X$ we have:

$$d_\alpha(x,y) \geq \frac{1}{c_{contr}} \cdot d(x,y) \qquad \forall \alpha \in \text{support}(\mathcal{D})$$
$$\mathop{\mathbb{E}}_{\alpha \sim \mathcal{D}}[d_\alpha(x,y)] \leq c_{exp} \cdot d(x,y).$$

Observe that *every* embedding in the support of $\mathcal{D}$ has contraction at most c_{contr}, but pairwise expansion (and thus distortion) is bounded in expectation over the randomness in $\mathcal{D}$. If the support of $\mathcal{D}$ contains a single embedding, it is a *deterministic* embedding. Further, an embedding is *expanding* or *non-contracting* if it has contraction 1.

We use the notation $\alpha : X \xhookrightarrow{c_{exp}, c_{contr}} Y$ to denote a deterministic or probabilistic embedding with expansion c_{exp} and contraction c_{contr} from (X, d_X) into (Y, d_Y). If c_{contr} is not specified, the embedding is expanding.

Bi-Lipschitz extensions

▶ **Definition 4** (Bi-Lipschitz extension). *Let (X, d) be a metric, $S \subseteq X$, and $\alpha_S : S \xhookrightarrow{c_S, c'_S} Y$ for some parameters c'_S, c_S. Then an embedding $\alpha : X \to Y$ is a (strong)* bi-Lipschitz extension *of α_S with extension factors L_{exp}, L_{contr} if* [1]

$$\text{for all } u \in S, \qquad \alpha(u) = \alpha_S(u), \tag{1}$$
$$\text{for all } u, v \in X \qquad d_\alpha(u,v) \geq \frac{1}{c'_S \cdot L_{contr}} d(u,v) \tag{2}$$
$$\text{and, for all } u, v \in X, \qquad d_\alpha(u,v) \leq L_{exp} \cdot c_S \cdot d(u,v). \tag{3}$$

If Y is a normed vector space, for Constraint 1 we require only that the restriction of α to the subspace spanned by S matches the embedding α_S.[2]

Further, α is a weak bi-Lipschitz extension *with extension factors L_{exp}, L_{contr} if Constraint 1 is replaced with the following weaker constraint:*

$$\text{for all } u, v \in S, \qquad d_\alpha(u,v) = d_{\alpha_S}(u,v).$$

In this paper, we discuss bi-Lipschitz extensions in the probabilistic setting. In particular, we model a probabilistic bi-Lipschitz extension on the weak version of a deterministic bi-Lipschitz extension, requiring that the *joint distribution* of distances between pairs of nodes in s is the same under the extended embedding.

[1] Note that L_{exp}, L_{contr} could be factors that depend on X and S.

[2] Notably, to get non-trivial extensions we must allow α to "extend" the dimension of α_S in some cases. To see this, consider an isometric (distortion 1) embedding of one or two points of a length n cycle into the real line. Any bi-Lipschitz extension that also embeds into the real line must have $L_{ext} \cdot L_{contr} = \Omega(n)$, as $\Omega(n)$ distortion is required to embed the cycle into a line.

▶ **Definition 5** (Probabilistic bi-Lipschitz extension). *Let (X, d) be a metric, $S \subseteq X$, and $\mathcal{D}_S \xhookrightarrow{c_S, c'_S} \mathcal{Y}$ for some parameters c_S, c'_S. Then a probabilistic embedding $\mathcal{D} : X \to \mathcal{Y}$ is a* probabilistic bi-Lipschitz extension *of α_S with extension factors L_{exp}, L_{contr} if*

$$(d_\alpha(x,y))_{x,y \in S \times S} \text{ has the same joint distribution for } \alpha \sim \mathcal{D} \text{ and } \alpha \sim \mathcal{D}_S \tag{4}$$

$$\text{for all } x, y \in X, \text{ and for all } \alpha \in \operatorname{support}(\mathcal{D}), \qquad d_\alpha(x,y) \geq \frac{1}{L_{contr} \cdot c'_S} \cdot d(x,y) \tag{5}$$

$$\text{and, for all } x, y \in X, \qquad \mathop{\mathbb{E}}_{\alpha \sim \mathcal{D}}[d_\alpha(x,y)] \leq L_{exp} \cdot c_S \cdot d(x,y). \tag{6}$$

Outlier embeddings

If $\mathcal{D}$ is a probabilistic embedding from (X, d_X) into $\mathcal{Y}$ and $K \subseteq X$, define $\mathcal{D}_{X \setminus K}$ to be the distribution identical to $\mathcal{D}$ but with all functions α in the support restricted to the set $X \setminus K$ (i.e. we can sample from $\mathcal{D}_{X \setminus K}$ by sampling from $\mathcal{D}$ and restricting the domain of the function drawn).

▶ **Definition 6.** *A probabilistic embedding $\mathcal{D}$ of (X, d_X) into $\mathcal{Y}$ is a* probabilistic (k, c)-outlier embedding *if there exists $K \subseteq X$ such that $|K| \leq k$ and $\mathcal{D}_{X \setminus K}$ is a probabilistic embedding of $(X \setminus K, d_X)$ into $\mathcal{Y}$ with distortion c.*

Note that in Definition 6, the subset K is chosen *deterministically* (not sampled). The distances from K to other nodes (and each other) are ignored. The randomness is only on the embedding for the remaining nodes in $X \setminus K$.

Hierarchically Separated Trees (HSTs), ultrametrics, and ℓ_p

▶ **Definition 7.** *A metric (X, d) is a β-*hierarchically separated tree (β-HST) *if there is a map from X to the leaves of a tree such that the following hold:*

- *Each node u in the tree is associated with ("labeled" with) some value η_u, and $\eta_v = \beta \cdot \eta_u$ whenever u is a child of v and u is not a leaf.*[3] *Further, for all leaves v, $\eta_v = 0$, and for nodes v that are parents of leaves, $\eta(v) = 1$.*
- *For two leaves $x, y \in X$ with least common ancestor z, $d(x, y) = \eta_z$.*

For simplicity, we will primarily focus on 2-HSTs, and in referring to HSTs, we mean 2-HSTs unless otherwise noted. We use $\mathcal{H}$ to denote the set of all 2-HSTs. Note that focusing on β-HSTs for a different constant β at worst changes results by a constant factor.

▶ **Definition 8.** *An* ultrametric *is a metric (X, d) that fulfills a strong triangle inequality guarantee: for all $x, y, z \in X$, $d(x, y) \leq \max\{ d(x, z), d(z, y) \}$.*

HSTs form a subset of the set of *ultrametrics*. Further, any ultrametric embeds into an HST with distortion at most 2 [17]. We will use $\mathcal{U}$ to denote the set of all ultrametrics.

▶ **Definition 9** (ℓ_p space). *ℓ_p in dimension δ, denoted ℓ_p^δ, is $(\mathbb{R}^\delta, d_{\ell_p})$ where for $x, y \in \mathbb{R}^\delta$, d_{ℓ_p} is defined as*

$$d_{\ell_p}(x,y) := \left(\sum_{i=1}^{\delta} |x_i - y_i|^p \right)^{1/p}.$$

We let ℓ_p space refer to the set of all spaces ℓ_p^δ for $\delta \in [1, \infty)$.

---

[3] In this paper, we will use the term β-HSTs to refer to what other sources sometimes call exact β-HSTs. These sources typically define non-exact β-HSTs as only requiring $\eta_v \geq \beta \cdot \eta_u$.

Main results

Our primary results are for probabilistic bi-Lipschitz extensions of embeddings into HSTs and deterministic bi-Lipschitz extensions of embeddings into ℓ_1 space.

▶ **Theorem 2** (HSTs have probabilistic bi-Lipschitz extensions). *Let $\mathcal{H}$ be the class of all HSTs. For any subset $S \subseteq X$ with $k := |X \setminus S|$ and a probabilistic embedding $\mathcal{D}_S : S \hookrightarrow \mathcal{H}$, there exists a probabilistic bi-Lipschitz extension of $\mathcal{D}_S$, $\mathcal{D} : X \hookrightarrow \mathcal{H}$, with expansion factor $L_{exp} = O(\log k)$ and contraction factor $L_{contr} = 4$.*

Further, we can efficiently sample from the distribution $\mathcal{D}$ if we can efficiently sample from $\mathcal{D}_S$.

We can also argue for the existence of deterministic extensions for ℓ_1 embeddings, though we do not know how to find such deterministic embeddings efficiently.

▶ **Theorem 3** (ℓ_1 has bi-Lipschitz extensions). *For any subset $S \subseteq X$ with an embedding $\alpha_S : S \xhookrightarrow{c_S, c'_S} \ell_1$, there exists*

1. *a deterministic* weak *bi-Lipschitz extension of α_S, $\alpha : X \hookrightarrow \ell_1$ with extension factors $L_{exp} = O(\log k)$ and $L_{contr} = 1$, where $k := |X \setminus S|$ and*
2. *a deterministic strong bi-Lipschitz extension with $L_{exp} = O(\log k)$ and $L_{contr} = 7$.*

Our bi-Lipschitz extensions into ℓ_1 are asymptotically optimal. Some metrics on n points (namely expander graph metrics) require $\Omega(\log n)$ distortion for embedding into ℓ_1 [50]. We can always "extend" a distortion 1 embedding on one or two points to an embedding on an expander graph metric, obtaining an embedding of distortion at most $L_{exp} \cdot L_{contr}$. Thus, we must have $L_{exp} \cdot L_{contr} = \Omega(\log k)$. Likewise, some metrics require $\Omega(\log n)$ distortion to embed probabilistically into HSTs [12], so our HST results are also asymptotically optimal.

Our results for probabilistic bi-Lipschitz extensions into HSTs allow us to obtain a bicriteria approximation for probabilistic outlier embeddings into HSTs, which we discussed in the introduction.

▶ **Theorem 1** (HSTs admit outlier embeddings). *Let (X, d) be any metric that admits a (k, c) probabilistic outlier embedding into HSTs. Then for any constant $\epsilon > 0$, there exists an algorithm that efficiently samples from a probabilistic embedding into HSTs of some $S \subseteq X$ with $|X \setminus S| \leq O(\frac{1}{\epsilon} k \log k)$ and expected distortion at most $(32 + \epsilon)c$.*

In Section 6, we expand our results to the weighted outlier setting, and we use a modification of the LP from Munagala et al. [53] to prove Theorem 1. Our LP-based techniques also allow us to obtain the following result for the weighted variant of the problem.[4]

▶ **Corollary 10.** *Given a metric (X, d), node weights w, and target distortion c, let w^* be the optimal outlier set cost for probabilistically embedding (X, d) into HSTs with c distortion and let k be the size of such an optimal outlier set. Then for any constant $\epsilon > 0$, there exists an algorithm that efficiently samples from a probabilistic embedding into HSTs of some $S \subseteq X$ with $w(X \setminus S) \leq O(\frac{1}{\epsilon} w^* \log k)$ and expected distortion at most $(32 + \epsilon)c$.*

[4] In the weighted outlier set problem we are given weights w_i on all nodes i in the metric space. The cost of a subset K of the metric space is $w(K) := \sum_{i \in K} w_i$, and the goal is to find an outlier embedding with some target distortion such that the cost of the outlier set is minimized.

3 Components of a bi-Lipschitz extension

The bi-Lipschitz extensions we discuss in this paper have three primary components. In particular, given a probabilistic embedding $\mathcal{D}_S : S \to \mathcal{Y}$ for $S \subseteq X$, we develop bi-Lipschitz extensions for $\mathcal{D}_S$ when the relevant spaces have certain properties. Namely, the space X should admit what we refer to as "onion partitions" and the space $\mathcal{Y}$ should admit "bounded-diameter embeddings" and "perfect merge functions." We show in Section 4 that onion partitions always exist, so our bi-Lipschitz extensions depend on the existence of bounded-diameter embeddings and perfect merge functions for the space into which we are embedding.

At a high level, we will use the onion partitions to find clusters of the set $K = X \setminus S$ and assign an "anchor point" in S to each such cluster. Then we find a bounded-diameter embedding on each cluster with its anchor point, and we "merge" each of these embeddings with the original embedding on S using a perfect merge function.

3.1 Onion partitions

We begin by defining what we call onion partitions. Given a metric (X, d) and a subset $S \subseteq X$ with $K = X \setminus S$, an onion partition of K is a partition $\mathcal{K} = \{ K_i \}$ of K such that each K_i has an assigned "anchor point" from S. For any point $x \in K_i$, we refer to its cluster's assigned point in S as the anchor point of x. The partition should satisfy three properties: (1) for any $x, y \in K$, the probability that x and y are separated in the partition is proportional to $d(x, y)/d(\{ x, y \}, S)$; (2) for each $x \in K$, the distance between x and its anchor point is (deterministically) similar to $d(x, S)$; and (3) for any cluster $K_i \in \mathcal{K}$, the distance between $x \in K_i$ and its assigned anchor point is (deterministically) not too much smaller than the diameter of the cluster K_i itself.

▶ **Definition 11** (Onion partitions). *Let (X, d) be a metric and let $S \subseteq X$ with $K := X \setminus S$. Then an* onion partition *of K is a joint distribution over $(\mathcal{K} = \{ K_i \}_{i \in \mathbb{N}}, r : \mathcal{K} \to S)$ such that $\mathcal{K}$ is a partition of K.*

For any $(\mathcal{K}, r)$ in the support of the distribution and $K_i \in \mathcal{K}$, we call $r(K_i)$ the anchor point *of K_i. We extend the domain of r to points in K: for $x \in K$, $r(x)$ denotes the anchor point assigned to the component containing x in the random partition $(\mathcal{K}, r)$. Likewise, let Δ_{K_i} denote the diameter of a component K_i, and Δ_x denote the (random) diameter of the component containing x in the random partition $(\mathcal{K}, r)$. Additionally, for any fixed $(\mathcal{K}, r)$ we let K_i^+ denote the set $K_i \cup \{ r(K_i) \}$ for each $K_i \in \mathcal{K}$.*

An onion partition has separation parameter $\rho > 0$, closeness $b > 1$, *and* band thickness $\gamma > 1$ *if it meets the following three criteria:*

1. ***Low separation probability.*** *For any $x, y \in K$, the probability that x and y are in different components of the partition is upper bounded by $\rho \cdot \frac{d(x,y)}{d(\{ x,y \},S)}$*
2. ***An anchor point is a "close" point in S.*** *For any $x \in K$ and any $(\mathcal{K}, r)$ in the support of the distribution, $d(x, r(x)) \leq b \cdot d(x, S)$ with probability 1.*
3. ***Cluster points are arranged in a thin band around their anchor point.*** *For any $x \in K$ and any $(\mathcal{K}, r)$ in the support of the distribution, $d(x, r(x)) \geq \Delta_x / \gamma$ with probability 1.*

Figure 1 gives a visual depiction of the third property of onion partitions, and in Section 4, we show the following theorem about their existence.

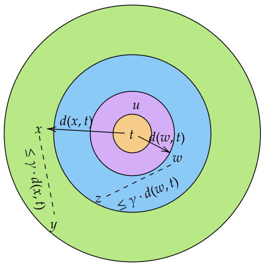

Figure 1 This figure depicts the third property of onion partitions. In particular, for any point $t \in S$, we think of the clusters with anchor point t as rings around t (like the rings of an onion). For any $x, y \in K$, if x and y are in the same cluster with anchor point $r(x) = r(y) = t$, then the third criterion of onion partitions states that they must be a distance at most $\gamma \cdot d(x,t)$ apart.

▶ **Theorem 12** (Onion partitions). *For any metric (X, d) and any $S \subseteq X$, there exists an onion partition on $K = X \setminus S$ with separation parameter $O(\log k)$, closeness 2, and band thickness 6, where $k = |K|$.*

3.2 Bounded-diameter embeddings

The second component of our extension method involves what we call bounded-diameter embeddings. In particular, we consider expanding embeddings that are parameterized not only by their worst expected expansion for any pair of nodes, but also by their expected expansion of the *diameter* of the original metric. That is, a bounded-diameter embedding can place a smaller upper bound on the expected expansion for pairs of nodes that are relatively far apart in the original metric than it does for other pairs of nodes. We define this formally in Definition 13.

▶ **Definition 13** (Bounded diameter embedding). *A probabilistic embedding $\mathcal{D} : X \to \mathcal{Y}$ is an expanding* bounded-diameter embedding *with expansion c and* diameter expansion *c_Δ if the following conditions hold:*

$$d_\alpha(x, y) \geq d(x, y) \qquad \forall x, y \in X, \forall \alpha \in \text{support}(\mathcal{D})$$
$$\mathop{\mathbb{E}}_{\alpha \sim \mathcal{D}}[d_\alpha(x, y)] \leq c \cdot d(x, y) \qquad \forall x, y \in X$$
$$\mathop{\mathbb{E}}_{\alpha \sim \mathcal{D}}[d_\alpha(x, y)] \leq c_\Delta \cdot \max_{x', y' \in X} \{ d(x', y') \} \qquad \forall x, y \in X$$

Further, we say $\mathcal{Y}$ admits bounded-diameter embeddings with expansion c_k and diameter expansion c_Δ *if for all metrics (X, d) of size k, there is a bounded-diameter embedding of X into $\mathcal{Y}$ with the given parameters.*

Notably, if all metrics of size k can embed into $\mathcal{Y}$ with expected expansion at most c_k, then $\mathcal{Y}$ admits bounded-diameter embeddings with expansion and diameter expansion equal to c_k. However, in the settings we consider here (namely HSTs and ℓ_1 space), we will consider bounded-diameter embeddings where c_Δ is tighter than c_k. In this paper, we will use FRT embeddings [37] as our bounded diameter embeddings.

▶ **Theorem 14** ([37]). *For any metric (X, d) with $|X| = n$, there exists a probabilistic expanding bounded diameter embedding (an FRT embedding) of X into HSTs with expansion $O(\log k)$ and diameter expansion 2.*

In fact, for HST embeddings, we can assume without loss of generality an even stronger property than that required by Definition 13: For any embedding α in the support of an HST embedding on X, the diameter of the image of α is at most twice the diameter of X. To see this, let Δ be the diameter of X and consider an HST embedding with diameter larger than $2^{\lceil \log \Delta \rceil}$. The nodes of the HST at level $2^{\lceil \log \Delta \rceil}$ or higher can all be merged into a single node without incurring any new contraction between any pair of points, as their new distance under this contraction is the minimum of their old distance and $2^{\lceil \log \Delta \rceil}$.

Further, trees embed isometrically into ℓ_1, so we immediately get the following corollary.

▶ **Corollary 15.** *For any metric (X, d) with $|X| = n$, there exists a probabilistic expanding bounded diameter embedding (an FRT embedding) of X into ℓ_1 with expansion $O(\log k)$ and diameter expansion 2.*

3.3 Perfect merge functions

In the final step of our extension method, we "merge" the embedding α_S on $S \subseteq X$ with bounded-diameter embeddings on the clusters of our onion partition. In particular, a *merge function* is a function whose input is two embeddings whose domains overlap at a single point and whose output is a new embedding on the union of their domains. For our purposes, we want to ensure some of the necessary distance properties are maintained. In particular, for any pair of nodes that appeared in the domain of the same input embedding, the distance under the new embedding should be the same as under the original one. Additionally, we want to ensure that no pair of nodes is allowed to be "too close" in the new embedding. This extra requirement will allow us to bound contraction in our final bi-Lipschitz extension.

▶ **Definition 16** (Perfect merge). *A procedure* `Merge` *is a* perfect merge function *for embedding X into $\mathcal{Y}$ with* contraction factor *η if for any $Z_1, Z_2 \subseteq X$ with $Z_1 \cap Z_2 = \{v\}$ and any deterministic embeddings $\alpha_1 : Z_1 \to Y_1$, $\alpha_2 : Z_2 \to Y_2$ for $Y_1, Y_2 \in \mathcal{Y}$, it produces an embedding $\alpha : Z_1 \cup Z_2 \to Y$ for some $Y \in \mathcal{Y}$ such that:*

1. *For all $x, y \in Z_1$, $d_\alpha(x, y) = d_{\alpha_1}(x, y)$.*
2. *For all $x, y \in Z_2$, $d_\alpha(x, y) = d_{\alpha_2}(x, y)$.*
3. *For $x \in Z_1$ and $y \in Z_2$, we have $d_\alpha(x, y) \geq \frac{d_{\alpha_1}(x,v)+d_{\alpha_2}(v,y)}{\eta}$.*

We say $\mathcal{Y}$ admits perfect merge functions if there is a perfect merge function for any metric X.

In Section 5, we prove the following existence results about perfect merge functions for HSTs and ℓ_1.

▶ **Lemma 17.** *HSTs admit a perfect merge function with contraction factor 2.*

▶ **Lemma 18.** *ℓ_1 admits a perfect merge function with contraction factor 1.*

Putting the components together

With our three components defined, we now put them together to construct bi-Lipschitz extensions for amenable target spaces $\mathcal{Y}$. In particular, given a metric (X, d) with subset $S \subseteq X$ and $\mathcal{D}_S : S \to \mathcal{Y}$, our extension proceeds in the following manner:

1. Sample an embedding on α_S from $\mathcal{D}_S$.
2. Sample an onion partition $(\mathcal{K}, r)$ of $K = X \setminus S$.
3. For each $K_i \in \mathcal{K}$, sample an embedding α_i from a bounded diameter embedding of $K_i^+ = K_i \cup \{ r(K_i) \}$.
4. Use a perfect merge function to sequentially merge each α_i with α_S.

Algorithm 1 formalizes this sampling process, and Theorem 19 states that the embedding sampled by Algorithm 1 samples from a bi-Lipschitz extension of the input embedding $\mathcal{D}_S$. We highlight in red the places where Algorithm 1 uses each of the three components we have discussed so far.

Algorithm 1 Bi-Lipschitz extension: Algorithm for finding a probabilistic bi-Lipschitz extension when the requirements of Theorem 19 are met.

Input: Metric space (X, d), subset $S \subseteq X$, embedding $\mathcal{D}_S : S \to \mathcal{Y}$; the spaces X and $\mathcal{Y}$ should meet the conditions of Theorem 19, and `PerfectMerge` will be the designated perfect merge function

Output: A randomly sampled embedding $\alpha : X \to \mathcal{Y}$.

1: Sample a partition $(\mathcal{K} = \{ K_i \}_{i \in \mathbb{N}}, r)$ from an onion partition of $K = X \setminus S$
2: Let K_i^+ be $K_i \cup \{ r(K_i) \}$ for all $K_i \in \mathcal{K}$
3: Sample $\alpha \sim \mathcal{D}_S$
4: **for** each non-empty K_i **do**
5: Let $\mathcal{D}_i$ be an expanding bounded-diameter embedding of K_i^+ into $\mathcal{Y}$
6: Sample $\alpha_i \sim \mathcal{D}_i$
7: $\alpha \leftarrow$ `PerfectMerge`(α, α_i)
8: Return α

▶ **Theorem 19.** *Let (X, d) be a metric with $S \subseteq X$ and let $\mathcal{D}_S : S \to \mathcal{Y}$ be a probabilistic embedding with contraction at most c'_S and expansion at most c_S.*
Then, if:

1. *X admits an onion partition with separation parameter ρ, closeness 2,*[5] *and band thickness γ,*
2. *$\mathcal{Y}$ admits expanding bounded-diameter embeddings on sets of size k with expansion c_k and diameter expansion c_Δ, and*
3. *$\mathcal{Y}$ admits a perfect merge function with contraction factor η,*

$\mathcal{D}_S$ has a bi-Lipschitz extension to all of X with extension factors $L_{exp} = O(\gamma \cdot \rho \cdot (c_\Delta/c_S + 1) + c_k/c_S)$ and $L_{contr} = \eta^2$. Further, Algorithm 1 samples from such an extension.

Before proving this theorem, we observe that due to the first two properties of a perfect merge function, it is sufficient to bound the first extension Algorithm 1 produces that includes both x and y in the domain.

▶ **Observation 20.** *Let $x, y \in X$. Consider the first iteration of Algorithm 1 in which both x and y are in the domain of α, and let $\hat{d}$ be $d_\alpha(x, y)$ at that time step. Then at every subsequent iteration of the algorithm, $d_\alpha(x, y) = \hat{d}$.*

We are now ready to bound the contraction and expansion of the embedding procedure, which we do in Lemmas 21 and 22. We defer the full proofs of these lemmas to the full version of this paper and just give a brief overview of the ideas here.

▶ **Lemma 21** (Contraction bound). *Let $X, S, \mathcal{D}_S$ be as in Theorem 19 and let $\alpha : X \to \mathcal{Y}$ be the embedding returned by Algorithm 1. Then for all $x, y \in X$, $d_\alpha(x, y) \geq d(x, y)/\eta^2$.*

[5] This could be defined for more general closeness parameters, but it greatly complicates the algebra, so we set $b = 2$ in our analysis.

To intuitively understand the contraction bound, note that when some node $v \in K_i$ is added to the domain of α, there is no contraction between v and other nodes in K_i^+, as $\mathcal{D}_i$ is non-contracting and we use a perfect merge function to combine this with α. Further, for any node u already in the domain of α when v was added, property 3 of perfect merge functions ensures $d_\alpha(u, v) \geq (d_\alpha(u, r(v)) + d_\alpha(v, r(v)))/\eta$. Applying the triangle inequality and appropriate contraction bounds on α prior to this point gives us the result.

We are now ready to bound the expected expansion for each pair of nodes. This argument follows a similar line of reasoning to that for bi-Lipschitz extensions for ℓ_1 in [26], but because we are using onion partitions rather than CKR partitions, we are able to better bound the distortion in the "bad case" of the embedding.

More specifically, the bad case for the embedding is when a pair of nodes $x, y \in K$ is separated by the partition and they are much closer to each other than they are to S (so routing their path through S potentially causes a large distortion). In this situation, we bound the distance between x and y in terms of $d(x, S)$ and $d(y, S)$, and we use the fact that the probability of separation is bounded by $\rho \cdot \frac{d(x,y)}{d(\{x,y\},S)}$ to obtain our final bound. Crucially, the bound we obtain for the distance between $x \in K_i$ and $y \in K_j$ depends on walking from x to $r(x)$ (the anchor point of x) through the embedding on K_i^+, then from $r(x)$ to $r(y)$ (the anchor point of y) through the embedding on S, and then from $r(y)$ to y through the embedding on K_j^+ (see Figure 2). In the simplest sense, we could apply a bound of c_S for the expansion of the distance between $r(x)$ and $r(y)$ and a bound of c_k (the distortion bound for embedding subsets of size k) for the distances between x and s and between y and t. Because we must pay ρ as part of the separation probability in the partition, our overall expected distortion goes up to $O(\rho \cdot (c_S + c_k))$.[6]

However, because of the onion partitions we use here, $d(x, s)$ is forced to be within a factor of γ of the diameter Δ_x of the cluster x is in. Further, because we use diameter-bounded embeddings on the set K_i^+, we know that the maximum distance between any pair of nodes in that embedding is at most $c_\Delta \cdot \Delta_x$. Thus, the distance between x and $r(x)$ in this embedding is within a constant factor of their true distance, allowing us to reduce our expansion factor in this bad case to $O(\rho \cdot \gamma \cdot (c_S + c_\Delta))$. We potentially still incur a cost of c_k if x and y are in the same component, so our overall cost is $O((c_S + c_\Delta) \cdot \rho \cdot \gamma + c_k)$, which gives us a potential savings if c_S and c_Δ are much smaller than c_k and γ is a constant. Our analysis follows a similar reasoning to that of Calinescu, Karloff, and Rabani [22] for the 0-extension problem, though we deal with an additional merge step and resulting effect on the embedding. See the full version of the paper for a full proof.

▶ **Lemma 22** (Expansion bound). *Let $X, S, \mathcal{D}_S$ be as in Theorem 19, and let $\mathcal{D}$ be the distribution over embeddings from which Algorithm 1 draws when given inputs $X, S, \mathcal{D}_S$. Then for all $x, y \in X$*

$$\mathop{\mathbb{E}}_{\alpha \sim \mathcal{D}}[d_\alpha(x, y)] \leq O(c_k + \gamma \cdot \rho \cdot (c_S + c_\Delta)) \cdot d(x, y),$$

where c_k is the bound on the expansion of the bounded-diameter embeddings for each part of the partition used by the algorithm, c_S is the expansion of $\mathcal{D}_S$, c_Δ is the diameter expansion of the bounded-diameter embeddings used by the algorithm, and $k = |X \setminus S|$.

---

[6] In fact, this bound applies even if we do not have a bounded diameter embedding (i.e. we just have an embedding with expected expansion c_k on sets of size k) and even if the partitions we use do not meet the third criterion of onion partitions (requiring that the distance between a cluster node and S be similar to the diameter of the cluster).

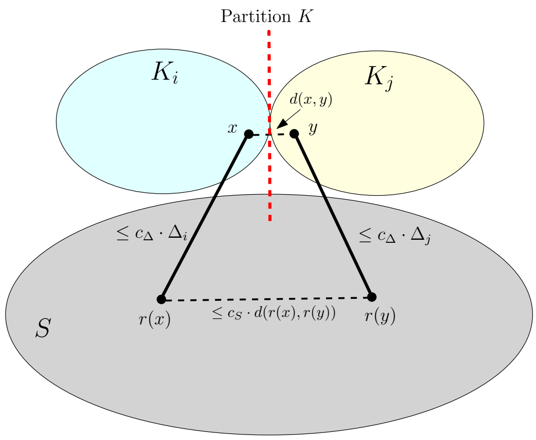

Figure 2 The "bad case" routing scenario for expansion bounds. Nodes $x \in K_i$ and $y \in K_j$ are physically close but separated by the onion terminal partition $\mathcal{K}$. The expected distance between x and y in the merged embedding is bounded by routing the path through their respective anchor points $r(x)$ and $r(y)$ within the base embedding S.

Noting that Algorithm 1 always outputs an embedding in which the distances on pairs of points in S are the same as under the sampled embedding α_S, Theorem 2 (finding bi-Lipschitz extensions for HSTs) is immediate from Theorems 12, 14, and 19 and Lemma 17. Theorem 3 (finding bi-Lipschitz extensions for ℓ_1) takes a little more work, and we prove it in Section 5.

4 Onion partitions

In this section we prove Theorem 12.

▶ **Theorem 12** (Onion partitions). *For any metric (X, d) and any $S \subseteq X$, there exists an onion partition on $K = X \setminus S$ with separation parameter $O(\log k)$, closeness 2, and band thickness 6, where $k = |K|$.*

We begin with a discussion of the partition algorithm given by Calinescu, Karloff, and Rabani in their work on the 0-extension problem [22], which we refer to as CKR partitions. Our onion partitions will consist of CKR partitions intersected with a randomized bucketing scheme for $K = X \setminus S$. The CKR partitions ensure that nodes are separated with a probability proportional to their distance, and the randomized bucketing allows us to control the diameter of each cluster relative to the designated anchor point without increasing the probability of separation between x and y by too much.

Given a metric (X, d) and a subset $S \subseteq X$, [22] finds a partition $\{K_1, K_2, \ldots, K_{|S|}\}$ of $K := X \setminus S$ such that the "anchor point" of K_s is $s \in S$, the probability of $x, y \in K$ being separated is $O(\log k \cdot \frac{d(x,y)}{d(\{x,y\},S)})$, and $x \in K_s$ for $s \in S$ implies $d(x, s) \leq 2d(x, S)$. Algorithm 2 presents a variation of the CKR algorithm in which we use a subset $T \subseteq S$ as the set of anchor points, rather than the entire set S.

The CKR partitioning algorithm works by selecting a random permutation on some set of terminals and then "assigning" points in $X \setminus S$ to terminals in $T \subseteq S$ such that (1) a node in $X \setminus S$ is always assigned to a terminal that is fairly close to it, and (2) the probability that a pair of nodes $x, y \in X$ are placed in separate part of the partition is proportional to $d(x, y)/d(\{x, y\}, T)$. We state these two properties as Observation 23 and Lemma 24. Note that if we let the anchor point of a cluster K_t for $t \in T$ be t itself, then these two facts imply that CKR partitions meet criteria (1) and (2) of onion partitions, so we're left with fulfilling criterion (3).

Algorithm 2 CKR partitioning.

Input: Metric space (X, d), subset $S \subseteq X$, a set of terminals $T \subseteq S$
Output: A partition $\{ K_t \}_{t \in T}$ of $X \setminus S$

1: Sample a permutation π uniformly at random over T
2: Sample $\mu \in [1, 2)$ uniformly at random.
3: **for** Each node $x \in X \setminus S$ **do**
4: Let $t \in T$ be the first terminal under the ordering π such that $d(x, t) \le \mu \cdot d(x, T)$
5: Add x to K_t
6: Return $\{ K_t \}_{t \in T}$, where the anchor point of K_t is set to t

▶ **Observation 23** (Deterministic Proximity of CKR Rounding, implicit in [22]). *For any $x \in X \setminus S$, if Algorithm 2 assigns x to K_t for some $t \in T$, then*

$$d(x, t) \le \mu \cdot d(x, T) \le 2 \cdot d(x, T),$$

where the last inequality holds because $\mu \in [1, 2)$.

▶ **Lemma 24** (CKR Separation, Lemma 2.2 in [22]). *Fix $x, y \in X$ and let $\delta := d(x, y)$. If $0 < \delta \le \frac{1}{4} \min\{d(x, T), d(y, T)\}$, then:*

$$\Pr[x, y \textit{ are separated}] \le 4\mathcal{H}_{|T|} \left(\frac{\delta}{d(x, T)} + \frac{\delta}{d(y, T)} \right),$$

where $\mathcal{H}_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$ denotes the m-th harmonic number.

For our partitions, we will pick T to be the set of "nearest neighbors" of nodes in K. That is, $T := \{ t \in \arg\min_{s \in S} \{ d(s, x) \} \mid x \in K \}$. We assume that for each $x \in K$, we add only one of the nearest neighbors of x to T so that $|T| \le k$. Moreover, by the choice of T, we also have $d(x, T) = d(x, S)$ for all $x \in K$.

Randomized bucketing

Now we show how to refine the CKR partition technique so that the diameter of each partition is bounded by a constant factor times the distance from that partition to the terminal set T. In particular, we will use a technique called *randomized scale partitioning*. For any point $x \in K$, let $A_x = d(x, T)$ denote its distance to the nearest terminal. We choose a shift parameter $u \in [0, 1)$ uniformly at random and define $\beta = 2^u$. We then (randomly) bucket K into disjoint scales B_i by distance from T:

$$B_i := \{x \in K \mid \beta \cdot 2^i \le A_x < \beta \cdot 2^{i+1}\}.$$

Given a metric (X, d) with diameter Δ, subset $S \subseteq X$, and terminal set $T \subseteq S$, we will randomly partition $K = X \setminus S$ into sets $\{ K_t \}_{t \in T}$ and buckets $\{ B_i \}_{i \in [\lceil \log \Delta \rceil]}$,[7] where the K_t are chosen by finding a CKR partition on K with terminal set T and the B_i are chosen by randomized scale partition. Then we define a new partition $\{ K_{t,i} \}_{t \in T, i \in [\lceil \log \Delta \rceil]}$, where $K_{t,i} := (K_t \cap B_i)$ and let the anchor point of this set be t – that is, $r(K_{t,i}) = t$.

[7] Here, $[n]$ denotes $\{ 0, 1, 2, \ldots, n \}$.

The primary properties of this partitioning that we will use are (1) for any node $x \in K_{t,i}$, the distance from x to another node in $K_{t,i}$ is not too much larger than the distance from x to T, and (2) the probability that a pair of nodes x and y are separated in this partition is not too much larger than the probability that x and y are separated under the CKR embedding. We express these as Observation 25 and Lemma 26, respectively.

▶ **Observation 25** (Partition diameter bound). *For any part $K_{t,i}$ of our scale partition and any node $x \in K_{t,i}$, if $\Delta_{t,i}$ is the diameter of $K_{t,i}$, then $d(x,t) \geq \Delta_{t,i}/6$*

Proof. For any pair $x, y \in K_{t,i}$, we have $d(x,y) \leq d(x,t) + d(t,y) \leq 2d(x,T) + 2d(y,T) \leq 6d(x,T)$. The first inequality is by the deterministic properties of CKR partitions, and the second is because x and y being in the same bucket implies $d(y,T)$ is no more than twice $d(x,T)$. Thus, $\Delta_{t,i} \leq 6 \cdot d(x,T) \leq 6 \cdot d(x,t)$. ◀

▶ **Lemma 26** (Scale Partition Separation). *For any pair of points $x, y \in K = X \setminus S$, let $\delta := d(x,y)$, and assume without loss of generality that $A_x \leq A_y$. The probability that x and y assigned to distinct buckets B_i and B_j is bounded by $O\left(\frac{\delta}{A_x}\right)$.*

Thus, by the union bound and Lemma 24, the probability that x and y are in different parts of the partition $\{ K_{t,i} \}$ is at most $O(\log |T| \cdot \frac{\delta}{A_x})$.

Proof. Let x be in bucket i and y be in bucket j. The scale boundaries are determined by $\beta 2^i = 2^{i+u}$, which is equivalent to placing a grid of integers shifted by a uniformly random $u \in [0,1)$ on a logarithmic scale. The probability that an integer boundary falls strictly between $\log_2 A_x$ and $\log_2 A_y$ is bounded by the length of the interval between their logarithmic distances. By the metric triangle inequality, $A_y \leq A_x + \delta$. Therefore:

$$\begin{aligned}
\Pr[i \neq j] &\leq \log_2(A_y) - \log_2(A_x) \\
&= \log_2\left(\frac{A_y}{A_x}\right) \\
&\leq \log_2\left(\frac{A_x+\delta}{A_x}\right) = \log_2\left(1 + \frac{\delta}{A_x}\right) \\
&\leq \frac{\delta}{A_x \ln 2} \;=\; O\left(\frac{\delta}{A_x}\right)
\end{aligned}$$

where the final inequality follows from the standard logarithmic bound $\ln(1+z) \leq z$ for all $z \geq 0$. ◀

Thus, the partitions we have constructed inherit property (2) of onion partitions with closeness 2 from the CKR bounds on anchor point distances and our choice of T (as the nearest neighbors of nodes in K), and they inherit property (3) of onion partitions with band thickness 6 from bounds on the diameter of randomized bucketing (and the fact that taking a subset of a cluster never increases its diameter). They also achieve property (1) with parameter $O(\log k)$ by Lemmas 24 and 26, our choice of T, and a union bound. Thus, we get Theorem 12.

5 Merge functions for HSTs and ℓ_1

In this section, we give explicit perfect merge functions for ℓ_1 and HSTs, and we prove Theorem 3.

Algorithm 3 `MergeL1`(α_1, α_2) Algorithm for merging two ℓ_1 embeddings.

Input: Embeddings $\alpha_1 : Z_1 \to \ell_1$, $\alpha_2 : Z_2 \to \ell_1$, where $|Z_1 \cap Z_2| = 1$.
Output: An embedding of $Z_1 \cup Z_2$ into ℓ_1

1: Let $u = Z_1 \cap Z_2$
2: For all $x \in Z_1$, let $\alpha(x) \leftarrow \alpha_1(x) \circ \alpha_2(u)$, where $\circ$ is concatenation
3: For all $x \in Z_2$, let $\alpha(x) \leftarrow \alpha_1(u) \circ \alpha_2(x)$
4: Return α

5.1 Merge for ℓ_1

We begin by presenting a perfect merge function for ℓ_1 embeddings. This is captured in Algorithm 3, which essentially just extends each input embedding to the entire set $Z_1 \cup Z_2$ by mapping everything not already in the domain to the same point as $u \in Z_1 \cap Z_2$ and concatenating the resulting embeddings together. This is easily seen to meet the requirements of a perfect merge function with contraction bound 1 and thus we get Lemma 18 which states that ℓ_1 has perfect merge functions. We now move on to proving Theorem 3.

▶ **Theorem 3** (ℓ_1 has bi-Lipschitz extensions)**.** *For any subset $S \subseteq X$ with an embedding $\alpha_S : S \xhookrightarrow{c_S, c'_S} \ell_1$, there exists*

1. *a deterministic* weak *bi-Lipschitz extension of α_S, $\alpha : X \hookrightarrow \ell_1$ with extension factors $L_{exp} = O(\log k)$ and $L_{contr} = 1$, where $k := |X \setminus S|$ and*
2. *a deterministic strong bi-Lipschitz extension with $L_{exp} = O(\log k)$ and $L_{contr} = 7$.*

Proof. Part (1) of the proof is almost immediate from Theorems 12, 14, and 19, the fact that ℓ_1 has a perfect merge function with contraction factor 1, and the fact that Algorithm 1 outputs an embedding in which pairs of nodes in S have the same distance as under α_S. The problem is that Theorem 19 implies the existence of a *probabilistic* extension of the original embedding. However, for any finite support probabilistic embedding $\mathcal{D}$ of X into ℓ_1, there exists a deterministic embedding $\alpha : X \to \ell_1$ with $d_\alpha(x, y) = \mathbb{E}_{\alpha' \sim \mathcal{D}}[d_{\alpha'}(x, y)]$. (Note that the dimension of α may be much larger than those of the embeddings in the support of $\mathcal{D}$, and it is not clear how to efficiently find α when $\mathcal{D}$ has a large support.) See e.g. [26] for more. Proving the second part of this theorem requires a bit more of a sophisticated de-randomization method, and we defer it to the full version of the paper. ◀

5.2 Merge for HSTs

We now present a perfect merge function for HSTs to prove Lemma 17, which states that HSTs have perfect merge functions with contraction factor 2.

Algorithm 4 takes in two β-HSTs for a fixed ratio β and returns another β-HST. This algorithm works by "combining" the ancestors of the common point u at each level of the two HSTs. That is, the new HST consists of u and all its ancestors, and each ancestor of u has as its child subtrees the child subtrees of the ancestor of u at its same level in each HST. Figure 3 gives a visual depiction of this algorithm. We defer the correctness proof for Lemma 17 to the full version of the paper and briefly outline the idea here.

For any $x, y \in Z_1$ or $x, y \in Z_2$, the least common ancestor of x and y must be at the same level in the new tree as it was in the original embedding of Z_1 or Z_2, so their distance is unchanged from the original embedding. To bound the distance between $x \in Z_1, y \in Z_2$, consider the least common ancestor of x and y in the HST produced at the algorithm. The first ancestor of y that has descendants from Z_1 is the least common ancestor of y and u.

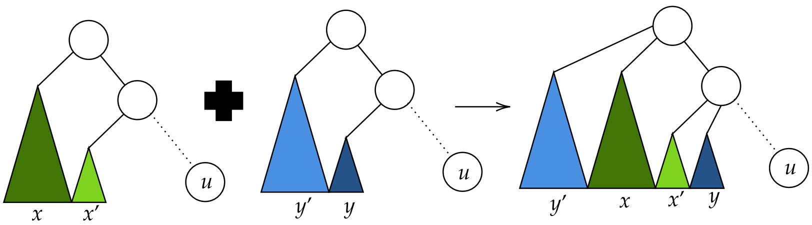

Figure 3 This figure depicts what happens in the algorithm `MergeHST`. In particular, u is the node common to both HSTs on the left side of the figure. The triangles represent subtrees, and the letters below the triangles indicate the presence of the corresponding node as a leaf of that subtree. Note that after merging trees, the least common ancestor of x and y is that of x and u from the first tree, as this was higher than that of y and u. This is the foundational idea behind why $d_\alpha(x, y) \geq \max\{ d_{\alpha_1}(u, x), d_{\alpha_2}(u, y) \}$ holds when α is the output embedding for `MergeHST` on inputs α_1 and α_2.

Thus, the least common ancestor of x and y cannot be lower than the least common ancestor of y and u, so the distance between x and y is at least that between y and u under the original embedding on Z_2. Likewise, the distance between x and y is at least the distance between x and u under the original embedding on Z_1, so we get contraction at most 2.

Algorithm 4 `MergeHST`(α_1, α_2) Algorithm for merging two β-HSTs.

Input: Embeddings $\alpha_1 : Z_1 \to T_1$, $\alpha_2 : Z_2 \to T_2$ for β-HSTs T_1, T_2, where $|Z_1 \cap Z_2| = 1$. We assume without loss of generality that the height of T_1 is at least as large as the height of T_2.
Output: An embedding of $Z_1 \cup Z_2$ into a single β-HST

```
Let u = Z_1 ∩ Z_2
Create a new HST T that is initially a copy of T_1
α(x) ← α_1(x) for all x ∈ Z_1
for i = 0 to the height of T_1 do
    for the subtree of each child of α_2(u)'s ancestor at height i in T_2 do
        if the subtree does not contain α_2(u) then ▷ i.e. if the subtree is not yet merged into T_1
            Copy the subtree and make it a child of α_1(u)'s ancestor at level i in T
            for any x ∈ Z_2 \ { u } such that α_2(x) is in the subtree at this step do
                Set α(x) to be the copy of α_2(x) created in T at this step
Return α
```

6 Outlier embeddings for HSTs

Now, given the existence of bi-Lipschitz extensions for HSTs, we turn to approximating outlier embeddings into HSTs. In [26], this type of approximation is obtained for ℓ_2 by introducing new "outlier variables" into a semidefinite program for a minimum distortion ℓ_2 embedding and rounding. We will employ a similar technique here using the linear program and associated rounding algorithm of [53] to approximate the optimal distortion for probabilistic HST embeddings.

Let Δ be the diameter of (X, d) and let $M = \{ 2^0, 2, 2^2, \ldots, 2^{\lceil \log \Delta \rceil} \}$. Let ζ be the constant factor from Theorem 2 and let $B_i(r) := \{ j \in X : d(i, j) \leq r \}$. We define (HST LP), which is parameterized by inputs $((X, d), k, c)$. The changes from [53] are highlighted in red.

Table 1 We imagine assigning a "representative" at each node of the HST such that the representative of a leaf node is the leaf node itself, and the representative of any other node is the same as one of its children's representatives. This table describes the intuitive meaning of each variable in the LP (which is largely the same as in [53]).

variable	intuitive meaning
δ_i	indicator for whether i is an outlier
$z^r_{ijj'}$	the probability that i is the representative of j and j' at level r
$\gamma^r_{jj'}$	the probability that j and j' are separated at level r
x^r_{ij}	the probability that i is the representative of j at level r

Table 1 gives an intuitive description of the variables in the LP. Note that in our LP c is a parameter, but in [53] it is the variable they are minimizing. The biggest change from [53] is that we have introduced new variables δ_i for each $i \in X$ that we use as an indicator for whether i is an outlier in the embedding.

$$\min \sum_{i=1}^{n} \delta_i \tag{HST LP}$$

$$\forall j, j' \sum_{r \in M} r \cdot \gamma^r_{jj'} \leq (4 + \zeta \cdot (\delta_j + \delta_{j'}) \cdot \log k) \cdot c \cdot d(j, j') \tag{7}$$

$$\forall i, r, (j, j' \in B_i(r)) \qquad \min(x^r_{i,j}, x^r_{ij'}) \geq z^r_{ijj'} \tag{8}$$

$$\forall j, j', i, r \sum_{i: j, j' \in B_i(r)} z^r_{ijj'} \geq 1 - \gamma^r_{jj'} \tag{9}$$

$$\forall j, r \sum_{i: j \in B_i(r)} x^r_{ij} = 1 \tag{10}$$

$$\forall j, j', (r < d(j, j')) \qquad \gamma^r_{jj'} = 1 \tag{11}$$

$$\forall j, j', r \qquad \gamma^r_{jj'} \leq 1 \tag{12}$$

$$\forall i, j, j', r \qquad \gamma^r_{jj'} x^r_{ij}, z^r_{ijj'}, \delta_i \geq 0; \delta_i \leq 1 \tag{13}$$

Given an optimal solution to this linear program, we need to define an outlier embedding with low distortion and few outliers. To do so, we will first use the rounding algorithm of [53] to find an embedding on the entire metric, and then we take as outliers all vertices i such that $\delta_i \geq \delta^*$ for some threshold δ^*. In the full version of the paper we formally prove that this rounding algorithm works and gives us Theorem 1 as a result. In a similar manner to ℓ_2 outlier embeddings in [26], we use the existence of nested embeddings to bound the optimal LP solution and thus obtain our final result.

Further, if we modify (HST LP) by changing the objective function to $\sum_i w_i \cdot \delta_i$ for some set of weights w_i, we can use the same technique to obtain a bicriteria approximation for the weighted variant of the outlier embedding problem, where the cost of making i an outlier is w_i. This result is expressed as Corollary 10, and we prove it in the full version of the paper.

7 Applications

In the Minimum Cost Communication Tree (MCCT) problem, we are given a finite metric $\mathcal{X} = (X, d)$ and weights r_{ij} for all pairs of nodes i, j and we are asked for a non-contracting embedding α of X into a tree with vertex set X such that $\sum_{i,j \in X} r_{ij} \cdot d_\alpha(i, j)$ is minimized. This problem is motivated by evolutionary biology, in which a scientist may attempt to reconstruct a tree of relationships between multiple organisms, and she comes to the problem with some known minimum distances between them, derived through a method such as DNA sequence analysis. In such a setting, the r_{ij} are an additional tunable parameter that can allow the scientist to indicate pairs she considers "more likely" to be related, by giving them a bigger r_{ij} value. As noted by Wu et al. [60], this problem can be approximated in expectation to within a factor of $O(\log n)$ by simply embedding randomly into an HST and converting it to a tree with vertex set X. Directly applying the algorithm of Munagala et al. [53] with the observations of Wu et al. [60], we have a randomized algorithm that gets an $O(c(\mathcal{X}))$-approximation for the MCCT problem, where $c(\mathcal{X})$ is the minimum distortion needed to probabilistically embed $\mathcal{X}$ into HSTs.

It is quite natural to consider an outlier version of the MCCT problem. In the evolutionary biology setting, we could consider the case that we are given some noisy data points, or points that don't really belong in the set being analyzed. One approach to this type of outlier problem may be to look for a k outlier tree embedding of minimum distortion, but the set of outliers that minimizes the embedding distortion does not necessarily also minimize the MCCT solution cost. We may expect that the outlier MCCT problem is easier to solve on tree metrics than on general metrics, so we could instead try the following approach: embed the input metric into a tree, and *then* try to find the best outliers for the resulting tree metric. In fact, if no outliers are allowed, MCCT is trivial on tree metrics. Further, in Subsection 7.1 we give a bicriteria approximation to the k-outlier MCCT problem when the input is a tree metric. In particular, if the optimal cost of a k outlier solution for a given instance of a tree metric is OPT_k, then we give an algorithm that finds a solution with at most $3k$ outliers and at most $24OPT_k$ cost. Let $c_k(\mathcal{X})$ be the minimum distortion required for probabilistically embedding all but at most k points in $\mathcal{X}$ into HSTs. We can use the outlier embedding algorithm from this paper to embed any input metric into a tree with outliers and then apply the approximation algorithm for outlier MCCT on trees, obtaining a solution with at most $O(\frac{k}{\epsilon}\log k)$ outliers and expected cost $O((1+\epsilon)\cdot c_k(\mathcal{X})\cdot OPT_k)$. By contrast, directly using the embedding algorithm of [53] with the tree approximation algorithm from this section instead gives a solution with at most $3k$ outliers and $O((1+\epsilon)\cdot c_0(\mathcal{X})\cdot OPT_k)$ expected cost, so we are unable to save on the cost of the optimal solution if there is a large subset that embeds well into HSTs but the entire metric does not. We explain this further in the full version of the paper.

7.1 MCCT with outliers

Let (T, d) be a tree metric, and let k be an integer. Then we define variables x_{ij} for each pair of nodes ij and variables δ_i for each variable i. We can use them to make the following linear program.

$$\begin{aligned} &\min \sum_{i,j\in T} x_{ij}\cdot r_{ij}\cdot d(i,j) \quad s.t. && \text{(MCCT Outlier LP)}\\ \forall i,j\in T:\quad &\delta_i+\delta_j+x_{ij}\geq 1 &&\\ &\sum_{i\in T}\delta_i\leq k &&\\ \forall i,j\in T:\quad &\delta_i, x_{ij}\in[0,1] && \end{aligned}$$

▷ Claim 27. Let (T, d) be a tree metric such that there exists a subset $K\subseteq T$, $|K|\leq k$, and $\sum_{i,j\in T\setminus K} r_{ij}\cdot d(i,j)\leq D$.

Then the following procedure finds a subset $K' \subseteq T$ with $|K'| \leq 3k$ and an embedding α of $T \setminus K'$ into a tree with the same vertex set and $\sum_{i,j \in T \setminus K'} d_\alpha(i,j) \leq 24D$.

1. Solve MCCT Outlier LP and let x_{ij}, δ_j be the resulting values
2. Let $K' = \{\, i : \delta_i \geq 1/3 \,\}$
3. Embed $T \setminus K'$ into a tree T' with distortion at most 8, using the algorithm of Gupta [43]

Proof. First, MCCT Outlier LP has a solution of value at most D. In particular, let K be as in the claim statement and consider setting $\delta_i = \begin{cases} 1 & i \in K \\ 0 & else \end{cases}$ and $x_{ij} = \begin{cases} 1 & \delta_i = \delta_j = 0 \\ 0 & else \end{cases}$. As there are only k outliers, the second constraint is clearly satisfied, and we have defined x_{ij} such that the other constraints are also satisfied. Further, x_{ij} is 0 if either i or j is an outlier in K, so the objective function value is just $\sum_{i,j \in T \setminus K} r_{ij} d(i,j)$, which by the claim statement is bounded by D.

Let K' be as in the claim. As the δ_i sum to at most k, we know $|K'| \leq 3k$. Consider any x_{ij} with $i, j \notin K'$ We have $x_{ij} \geq 1 - \delta_i - \delta_j \geq 1/3$ by how we chose K'. The cost associated with a given pair ij of non-outliers in our solution is $r_{ij} d(i,j)$, but that same pair contributed at least $r_{ij} d(i,j)/3$ to the optimal objective function value, so our final cost is at most 3 times the objective, which is at most $3D$. At this point, we have an isometric (distortion 1) embedding of the metric induced by $T \setminus K'$ into the original metric T itself. Thus, we can view K' as a set of Steiner points in our original metric.

Finally we consider the loss incurred applying the algorithm of Gupta [43]. This creates a tree embedding that removes Steiner points and incurs distortion at most 8, so the cost of this embedding on $T \setminus K'$ is at most $24D$, as desired. This step is what prevents us from applying this algorithm directly to general metrics. ◁

8 Open questions

There are still many open questions we can ask about topics in this space, such as:

- Is it NP-hard to find the optimal outlier set size when probabilistically embedding a given metric into HSTs with a small fixed distortion? This question is known to be hard for deterministic embeddings into some spaces [59, 26].
- Even if finding the optimal outlier set size for probabilistically embedding into HSTs is NP-hard, can the approximation factors in this paper be improved?
- Which other metric spaces admit perfect merge functions (and thus bi-Lipschitz extensions), and can we approximate optimal outlier embeddings in such spaces?

References

1 I. Abraham, Y. Bartal, J. Kleinberg, T.-H.H. Chan, O. Neiman, Kedar Dhamdhere, A. Slivkins, and Anupam Gupta. Metric embeddings with relaxed guarantees. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 83–100, 2005. doi:10.1109/SFCS.2005.51.

2 Ittai Abraham, Yair Bartal, and Ofer Neiman. On low dimensional local embeddings. In *Proceedings of the 2009 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 875–884, 2009. doi:10.1137/1.9781611973068.95.

3 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.

4 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.

5 Ittai Abraham, Shiri Chechik, Michael Elkin, Arnold Filtser, and Ofer Neiman. Ramsey spanning trees and their applications. *ACM Trans. Algorithms*, 16(2):19:1–19:21, 2020. preliminary version published in SODA 2018. `doi:10.1145/3371039`.

6 Noga Alon, Mihai Badoiu, Erik D. Demaine, Martin Farach-Colton, Mohammad Taghi Hajiaghayi, and Anastasios Sidiropoulos. Ordinal embeddings of minimum relaxation: General properties, trees, and ultrametrics. *ACM Trans. Algorithms*, 4(4):46:1–46:21, 2008. `doi:10.1145/1383369.1383377`.

7 Noga Alon, Richard M Karp, David Peleg, and Douglas West. A graph-theoretic game and its application to the k-server problem. *SIAM Journal on Computing*, 24(1):78–100, 1995. `doi:10.1137/S0097539792224474`.

8 Sanjeev Arora, László Lovász, Ilan Newman, Yuval Rabani, Yuri Rabinovich, and Santosh Vempala. Local versus global properties of metric spaces. *SIAM Journal on Computing*, 41(1):250–271, 2012. `doi:10.1137/090780304`.

9 Mihai Badoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Low-distortion embeddings of general metrics into the line. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 225–233. ACM, 2005. `doi:10.1145/1060590.1060624`.

10 Mihai Badoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Embedding ultrametrics into low-dimensional spaces. In Nina Amenta and Otfried Cheong, editors, *Proceedings of the 22nd ACM Symposium on Computational Geometry, Sedona, Arizona, USA, June 5-7, 2006*, pages 187–196. ACM, 2006. `doi:10.1145/1137856.1137886`.

11 Mihai Badoiu, Piotr Indyk, and Anastasios Sidiropoulos. Approximation algorithms for embedding general metrics into trees. In Nikhil Bansal, Kirk Pruhs, and Clifford Stein, editors, *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2007, New Orleans, Louisiana, USA, January 7-9, 2007*, pages 512–521. SIAM, 2007. URL: `http://dl.acm.org/citation.cfm?id=1283383.1283438`.

12 Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 184–193, 1996. `doi:10.1109/SFCS.1996.548477`.

13 Yair Bartal. On approximating arbitrary metrices by tree metrics. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, STOC '98, pages 161–168, New York, NY, USA, 1998. Association for Computing Machinery. `doi:10.1145/276698.276725`.

14 Yair Bartal. Graph decomposition lemmas and their role in metric embedding methods. In *Algorithms - ESA 2004, 12th Annual European Symposium, Bergen, Norway, September 14-17, 2004, Proceedings*, pages 89–97, 2004. `doi:10.1007/978-3-540-30140-0_10`.

15 Yair Bartal. Advances in metric ramsey theory and its applications. *CoRR*, abs/2104.03484, 2021. `arXiv:2104.03484`.

16 Yair Bartal, Arnold Filtser, and Ofer Neiman. On notions of distortion and an almost minimum spanning tree with constant average distortion. *J. Comput. Syst. Sci.*, 105:116–129, 2019. preliminary version published in SODA 2016. `doi:10.1016/j.jcss.2019.04.006`.

17 Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric ramsey-type phenomena. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 463–472, New York, NY, USA, 2003. Association for Computing Machinery. `doi:10.1145/780542.780610`.

18 Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. On metric Ramsey-type dichotomies. *Journal of the London Mathematical Society*, 71(2):289–303, 2005. `doi:10.1112/S0024610704006155`.

19 Yair Bartal, Nathan Linial, Manor Mendel, and Assaf Naor. Some low distortion metric ramsey problems. *Discret. Comput. Geom.*, 33(1):27–41, 2005. `doi:10.1007/s00454-004-1100-z`.

20 Guy E. Blelloch, Yan Gu, and Yihan Sun. A new efficient construction on probabilistic tree embeddings. *CoRR*, abs/1605.04651, 2016. `arXiv:1605.04651`.

21 J. Bourgain, T. Figiel, and V. Milman. On Hilbertian subsets of finite metric spaces. *Israel J. Math.*, 55(2):147–152, 1986. doi:10.1007/BF02801990.

22 Gruia Calinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. *SIAM Journal on Computing*, 34(2):358–372, 2005. doi:10.1137/S0097539701395978.

23 T. H. Hubert Chan, Michael Dinitz, and Anupam Gupta. Spanners with slack. In Yossi Azar and Thomas Erlebach, editors, *Algorithms – ESA 2006*, pages 196–207, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. doi:10.1007/11841036_20.

24 Moses Charikar, Chandra Chekuri, Ashish Goel, Sudipto Guha, and Serge Plotkin. Approximating a finite metric by a small number of tree metrics. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS '98, page 379, USA, 1998. IEEE Computer Society.

25 Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Local global tradeoffs in metric embeddings. *SIAM Journal on Computing*, 39(6):2487–2512, 2010. doi:10.1137/070712080.

26 Shuchi Chawla and Kristin Sheridan. Composition of nested embeddings with an application to outlier removal. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1641–1668, 2024. doi:10.1137/1.9781611977912.66.

27 Victor Chepoi, Feodor F. Dragan, Ilan Newman, Yuri Rabinovich, and Yann Vaxès. Constant approximation algorithms for embedding graph metrics into trees and outerplanar graphs. *Discret. Comput. Geom.*, 47(1):187–214, 2012. doi:10.1007/s00454-011-9386-0.

28 Yeshwanth Cherapanamjeri and Jelani Nelson. Terminal embeddings in sublinear time. *TheoretiCS*, 3, 2024. doi:10.46298/THEORETICS.24.6.

29 Karine Chubarian and Anastasios Sidiropoulos. Computing bi-lipschitz outlier embeddings into the line. In *International Workshop and International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 36:1–36:21, 2020. doi:10.4230/LIPIcs.APPROX/RANDOM.2020.36.

30 Mark de Berg, Krzysztof Onak, and Anastasios Sidiropoulos. Fat polygonal partitions with applications to visualization and embeddings. *J. Comput. Geom.*, 4:212–239, 2010.

31 Jeff Edmonds, Anastasios Sidiropoulos, and Anastasios Zouzias. Inapproximability for planar embedding problems. In Moses Charikar, editor, *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 222–235. SIAM, 2010. doi:10.1137/1.9781611973075.20.

32 M. Elkin, A. Filtser, and O. Neiman. Terminal embeddings. *Theoretical Computer Science*, 697:1–36, 2017. doi:10.1016/j.tcs.2017.06.021.

33 Michael Elkin, Arnold Filtser, and Ofer Neiman. Prioritized metric structures and embedding. *SIAM J. Comput.*, 47(3):829–858, 2018. doi:10.1137/17M1118749.

34 Michael Elkin and Ofer Neiman. Lossless prioritized embeddings. *SIAM J. Discret. Math.*, 36(3):1529–1550, 2022. doi:10.1137/21M1436221.

35 Matthias Englert, Anupam Gupta, Robert Krauthgamer, Harald Räcke, Inbal Talgam-Cohen, and Kunal Talwar. Vertex sparsifiers: New results from old techniques. *SIAM J. Comput.*, 43(4):1239–1262, 2014. doi:10.1137/130908440.

36 Jittat Fakcharoenphol, Chris Harrelson, Satish Rao, and Kunal Talwar. An improved approximation algorithm for the 0-extension problem. In *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 257–265, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics. URL: http://dl.acm.org/citation.cfm?id=644108.644153.

37 Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 448–455, 2003. doi:10.1145/780542.780608.

38 Michael R. Fellows, Fedor V. Fomin, Daniel Lokshtanov, Elena Losievskaja, Frances A. Rosamond, and Saket Saurabh. Distortion is fixed parameter tractable. *ACM Trans. Comput. Theory*, 5(4):16:1–16:20, 2013. doi:10.1145/2489789.

39 Arnold Filtser, Yuval Gitlitz, and Ofer Neiman. Light, reliable spanners. In Wolfgang Mulzer and Jeff M. Phillips, editors, *40th International Symposium on Computational Geometry, SoCG 2024, June 11-14, 2024, Athens, Greece*, volume 293 of *LIPIcs*, pages 56:1–56:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.SOCG.2024.56.

40 Arnold Filtser, Robert Krauthgamer, and Ohad Trabelsi. Relaxed voronoi: A simple framework for terminal-clustering problems. In Jeremy T. Fineman and Michael Mitzenmacher, editors, *2nd Symposium on Simplicity in Algorithms, SOSA 2019, January 8-9, 2019, San Diego, CA, USA*, volume 69 of *OASIcs*, pages 10:1–10:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/OASIcs.SOSA.2019.10.

41 Arnold Filtser and Hung Le. Clan embeddings into trees, and low treewidth graphs. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 342–355. ACM, 2021. doi:10.1145/3406325.3451043.

42 Arnold Filtser and Hung Le. Low treewidth embeddings of planar and minor-free metrics. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 1081–1092. IEEE, 2022. doi:10.1109/FOCS54457.2022.00105.

43 Anupam Gupta. Steiner points in tree metrics don't (really) help. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 220–227, 2001. URL: http://dl.acm.org/citation.cfm?id=365411.365448.

44 William B. Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into hilbert space. *Contemporary mathematics*, 26:189–206, 1984. URL: https://api.semanticscholar.org/CorpusID:117819162.

45 Jon Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. *J. ACM*, 56(6):32:1–32:37, September 2009. doi:10.1145/1568318.1568322.

46 Rasmus Kyng, Maximilian Probst Gutenberg, and Tim Rieder. Random-shift revisited: Tight approximations for tree embeddings and ℓ_1-oblivious routings. In *66th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2025, Sydney, Australia, December 14-17, 2025*, pages 2132–2142. IEEE, 2025. doi:10.1109/FOCS63196.2025.00113.

47 Christiane Lammersen, Anastasios Sidiropoulos, and Christian Sohler. Streaming embeddings with slack. In Frank Dehne, Marina Gavrilova, Jörg-Rüdiger Sack, and Csaba D. Tóth, editors, *Algorithms and Data Structures*, pages 483–494, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. doi:10.1007/978-3-642-03367-4_42.

48 J. R. Lee and A. Naor. Extending lipschitz functions via random metric partitions. *Inventiones Mathematicae*, 160(1):59–95, 2005. doi:10.1007/s00222-004-0400-5.

49 Sepideh Mahabadi, Konstantin Makarychev, Yury Makarychev, and Ilya P. Razenshteyn. Nonlinear dimension reduction via outer bi-lipschitz extensions. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 1088–1101. ACM, 2018. doi:10.1145/3188745.3188828.

50 Jiří Matoušek. On embedding expanders into ℓ_p spaces. *Israel Journal of Mathematics*, 102(1):189–197, 1997.

51 Jiří Matoušek and Anastasios Sidiropoulos. Inapproximability for metric embeddings into R^d. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 405–413. IEEE Computer Society, 2008. doi:10.1109/FOCS.2008.21.

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

53 Kamesh Munagala, Govind S. Sankar, and Erin Taylor. Probabilistic Metric Embedding via Metric Labeling. In Nicole Megow and Adam Smith, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023)*, volume 275 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:10, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.APPROX/RANDOM.2023.2.

54 Assaf Naor and Yuval Rabani. On lipschitz extension from finite subsets. *Israel Journal of Mathematics*, 219, June 2015. doi:10.1007/s11856-017-1475-1.

55 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.

56 Shyam Narayanan and Jelani Nelson. Optimal terminal dimensionality reduction in euclidean space. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1064–1069. ACM, 2019. doi:10.1145/3313276.3316307.

57 Amir Nayyeri and Benjamin Raichel. Reality distortion: Exact and approximate algorithms for embedding into the line. In Venkatesan Guruswami, editor, *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 729–747. IEEE Computer Society, 2015. doi:10.1109/FOCS.2015.50.

58 Anastasios Sidiropoulos, Mihai Badoiu, Kedar Dhamdhere, Anupam Gupta, Piotr Indyk, Yuri Rabinovich, Harald Räcke, and R. Ravi. Approximation algorithms for low-distortion embeddings into low-dimensional spaces. *SIAM J. Discret. Math.*, 33(1):454–473, 2019. doi:10.1137/17M1113527.

59 Anastasios Sidiropoulos, Dingkang Wang, and Yusu Wang. Metric embeddings with outliers. In *ACM-SIAM Symposium on Discrete Algorithms*, 2015.

60 Bang Ye Wu, Giuseppe Lancia, Vineet Bafna, Kun-Mao Chao, R. Ravi, and Chuan Yi Tang. A polynomial-time approximation scheme for minimum routing cost spanning trees. *SIAM Journal on Computing*, 29(3):761–778, 2000. doi:10.1137/S009753979732253X.